PAPERS and PROCEEDINGS

the 13th

FEDERAL FORECASTERS CONFERENCE

2003

Sponsored by

Bureau of Health Professions
Bureau of Labor Statistics
Bureau of Transportation Statistics
Department of Veterans Affairs
Economic Research Service
Internal Revenue Service
International Trade Administration
National Center for Education Statistics
U.S. Census Bureau
U.S. Geological Survey

Announcement

The 14th
Federal Forecasters Conference
(FFC/2005)

will be held

on

April 21, 2005

in

Washington, DC

More information will be available in the coming months.

Federal Forecasters Conference - 2003

Papers and Proceedings

Edited by

Debra E. Gerald
National Center for Education Statistics
Institute of Education Sciences
U.S. Department of Education
Washington, DC

Norman Saunders
Bureau of Labor Statistics
U.S. Department of Labor
Washington, DC

Edited and Prepared by

Marybeth Matthews
Beth Babick
Healthcare Analysis and Information Group
Veterans Health Administration
Office of the Assistant Deputy Under Secretary for Health
Department of Veterans Affairs
Milwaukee, Wisconsin

Donald Stockford
Policy Analysis Group
Veterans Health Administration
Office of the Assistant Deputy Under Secretary for Health
Department of Veterans Affairs
Washington, DC

Federal Forecasters Consortium Governing Board

Stuart Bernstein
Bureau of Health Professions
U. S. Department of Health and Human Services

Jeff Busse
U.S. Geological Survey
U.S. Department of the Interior

Russell Geiman
Internal Revenue Service
U.S. Department of the Treasury

Debra E. Gerald, Chairperson
National Center for Education Statistics
U.S. Department of Education

Karen S. Hamrick
Economic Research Service
U.S. Department of Agriculture

Frederick L. Joutz
The George Washington University

Elliot Levy
International Trade Administration
U.S. Department of Commerce

Stephen A. MacDonald
Economic Research Service
U.S. Department of Agriculture

Norman Saunders
Bureau of Labor Statistics
U.S. Department of Labor

Brian Sloboda
Bureau of Transportation Statistics
U.S. Department of Transportation

Kathleen Sorensen
Office of Assistant Secretary for Policy and Planning
U.S. Department of Veterans Affairs

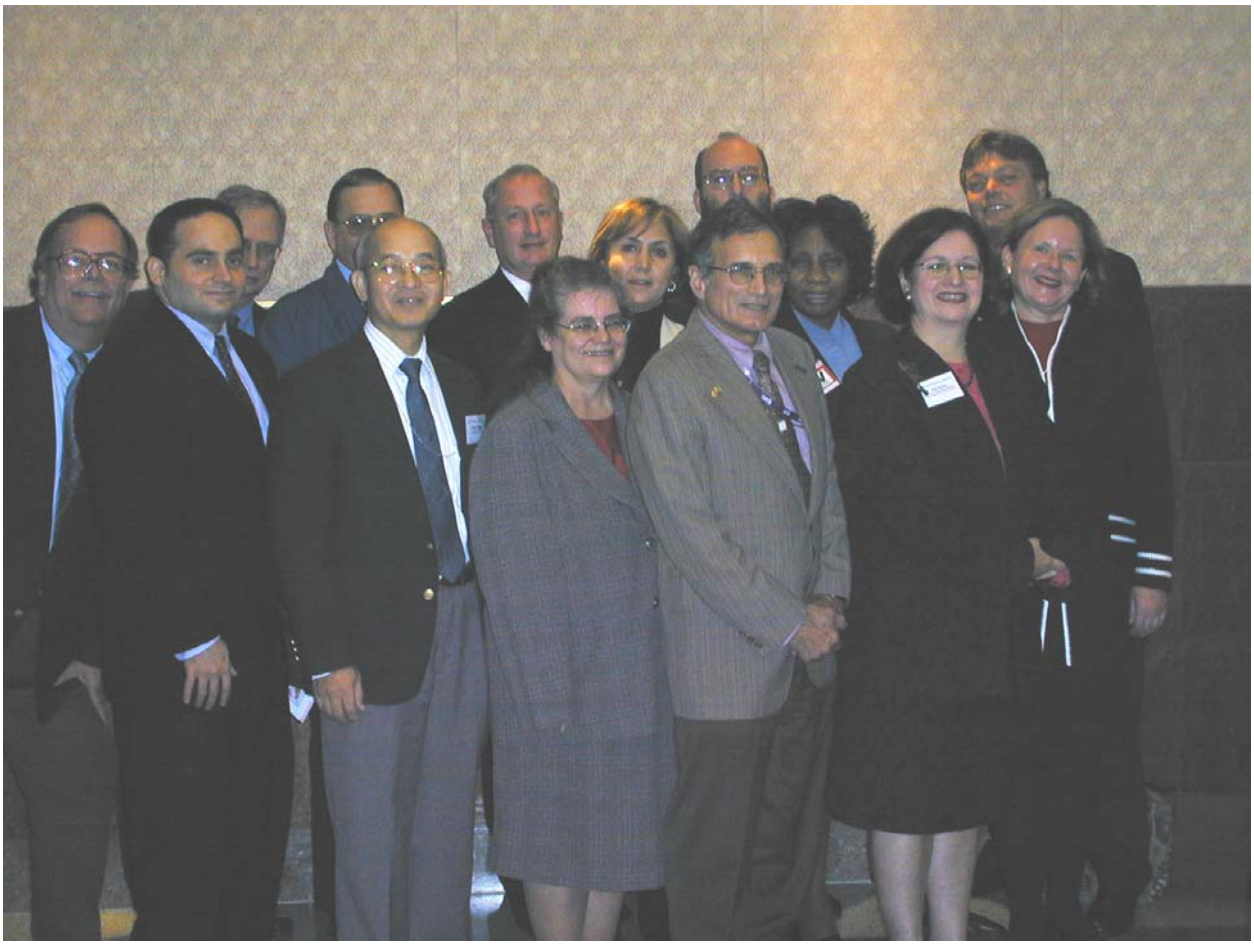
Donald Stockford
Veterans Health Administration
Office of the Assistant Deputy Under Secretary for Health
U.S. Department of Veterans Affairs

Mitra Toossi
Bureau of Labor Statistics
U.S. Department of Labor

Ching-li Wang
U.S. Census Bureau
U.S. Department of Commerce

Peg Young
Bureau of Transportation Statistics
U.S. Department of Transportation

Federal Forecasters Consortium Governing Board



(Front Row) **Brian Sloboda**, Bureau of Transportation Statistics; **Ching-li Wang**, U.S. Census Bureau; **Kathleen Sorensen**, Department of Veterans Affairs; **Elliot Levy**, International Trade Administration; **Peg Young**, Bureau of Transportation Statistics; and **Karen S. Hamrick**, Economic Research Service.

(Back Row) **Norman C. Saunders**, Bureau of Labor Statistics; **Howard N Fullerton, Jr.** (Retired), Bureau of Labor Statistics; **Jeffrey Osmint** (Retired), U.S. Geological Survey; **Donald Stockford**, Department of Veterans Affairs; **Mitra Toossi**, Bureau of Labor Statistics; **Jeff Busse**, U.S. Geological Survey; **Debra E. Gerald**, National Center for Education Statistics; and **Russell Geiman**, Internal Revenue Service. (Not Pictured) **Stuart Bernstein**, Bureau of Health Professions; **Stephen A. MacDonald**, Economic Research Service; and **Frederick L. Joutz**, The George Washington University.

Foreword

Following the cancellation of the 13th Federal Forecasters Conference (FFC/2003) on September 18, 2003 due to Hurricane Isabel, FFC/2003 was reconvened on October 27, 2003, in Washington, DC. FFC/2003 provided a forum where forecasters from different federal agencies and other organizations could meet and discuss various aspects of forecasting in the United States. The theme was "The Right Data: Measurement, Methodology, and Policy."

Nearly 200 forecasters attended the daylong conference. The program included opening remarks by Mitra Toossi and welcoming remarks from Norman Saunders of the Bureau of Labor Statistics. Following the remarks, a panel made up of James R. Caplan, Chief of Survey Technology Branch, Defense Manpower Data Center; Mark Mazur, Director of Research, Analysis, and Statistics, Internal Revenue Service; J. Gregory Robinson, Special Assistant for Demographic Analysis, Population Division, U.S. Census Bureau; and Art Klein, Acting Assistant Deputy Under Secretary for Health, U.S. Department of Veterans Affairs, was moderated by Kathleen Sorensen. Brian Sloboda and Jeff Busse presented awards from the 2003 Federal Forecasting Contest and Fred Joutz of George Washington University presented awards for Best Papers from FFC/2002.

In the afternoon, 12 concurrent sessions in two time slots were held featuring 41 papers presented by forecasters from the Federal Government, private sector, and academia. Varieties of papers were presented dealing with areas related to agriculture, the economy, forecasting techniques, health, labor, population, taxpayers, transportation, and veterans. These papers or abstracts are included in these proceedings. Another product of the FFC/2003 is the *2003 Federal Forecasters Directory*.

Acknowledgments

Many individuals contributed to the success of the 13th Federal Forecasters Conference (FFC/2003). First and foremost, without the support of the cosponsoring agencies and dedication of the Federal Forecasters Consortium Governing Board, FFC/2003 would not have been possible. Russell Geiman of Internal Revenue Service (IRS) prepared the announcement. Mitra Toossi of Bureau of Labor Statistics (BLS) chaired the morning program. Norman Saunders (BLS) gave the welcoming remarks. Frederick L. Joutz of The George Washington University and Brian Sloboda of the Bureau of Transportation Statistics (BTS) presented certificates for the 2003 Forecasting Contest and Best Papers 2002. Kathleen Sorensen of the U.S. Department of Veterans Affairs (VA) and Brian Sloboda (BTS) organized the morning panel. Debra Gerald of the National Center for Education Statistics prepared the program and FFC 2003 directory. Jeff Busse of U.S. Geological Survey, Russell Geiman (IRS), Stephen M. MacDonald of Economic Research Service (ERS), Donald Stockford (VA), Ching-li Wang (U.S. Census Bureau), and Peg Young (BTS) provided various conference materials. Karen S. Hamrick (ERS) served as program chair and organized the two afternoon concurrent sessions. Brian Sloboda (BTS) organized and conducted the 2003 Federal Forecasting Contest. Donald Stockford (VA) coordinated the agency's work on the production of the conference program and directory. Jeff Busse (USGS) prepared certificates for the forecasting contest and best papers. Norman C. Saunders (BLS) secured conference facilities, handled logistics, and was the photographer for the conference. Stuart Bernstein of Bureau of Health Professions and Elliot Levy of International Trade Administration provided support for the conference.

Special thanks go to Edward N. Gamber of Lafayette College and Frederick L. Joutz and Julian Silk of The George Washington University for reviewing the papers presented at the 12th Federal Forecasters Conference and selecting awards for the FFC/2002 Best Conference Papers.

Special thanks go to Linda D. Felton, Patricia Cleveland, and Alma Young of ERS for directing the organization of materials into conference packets and staffing the registration desk.

Special thanks go to Marybeth Matthews and Beth Babick of VA for producing the conference program, directory of federal forecasters, papers and proceedings, name badges, and certificates of attendance.

Last, special thanks go to all presenters, discussants, and attendees whose participation made FFC/2003 another successful conference.

2003
Federal Forecasters Conference
Forecasting Contest

WINNER

Patrick Flanagan
Bureau of Transportation Statistics

First Runner Up

Tom Snyder
National Center for Education Statistics

Second Runner Up

Mirko Novakovic
Bureau of Labor Statistics

Third Runner Up

Peggy Podolak
U.S. Department of Energy

Honorable Mention

Terry Schau
Betty Su
Bureau of Labor Statistics

2002
Best Conference Paper

WINNER

"Contingent Forecasting of Bulges in the Left and Right Tails of the Nonmetro Wage and Salary Income Distribution"

John Angle
Economic Research Service

Honorable Mention

"Accounts Receivable Resolution and the Impact of Lien Filing Policy on Sole Proprietor Businesses"

Alex Turk, Ph.D. and Terry Ashley, Ph.D.
Internal Revenue Service

"The Labor Force Over the Next 50 Years"

Mitra Toossi
Bureau of Labor Statistics

Contents

TIME SERIES MODELING AND SEASONAL ADJUSTMENT

LONG-RANGE FORECASTING

SOCIAL PROGRAMS AND FORECASTING MEASURES OF WELL-BEING

TRANSPORTATION AND ENERGY FORECASTING

DATA ISSUES AND STRATEGIES IN POPULATION PROJECTIONS

BUSINESS CYCLES AND GLOBAL FACTORS IN SHORT-RANGE FORECASTING

FORECASTING TECHNIQUES

MACROECONOMIC ISSUES

Panel Discussion

The Right Data: Measurement, Methodology, and Policy

Moderator: Kathleen Sorensen, U.S. Department of Veterans Affairs

Forecasts can be no better than the data they rely upon for input. But what data are available, and what products and activities can be measured, are not always in sync with the specific forecasts needed for public policy. The data, in turn, are the fundamental determinates in proper model identification and specification. And once forecasts are presented, their proper interpretation and level of accuracy can only be understood in the context of the data used as inputs.

The panelists seek to explore the efforts made to ensure the quality of data that go into forecasts; the techniques used to arrive at the best model among alternative forecasting methodologies; and the actions taken to ensure the proper interpretation of forecasts.

Understanding Nonresponse in Employee Populations: Who, What and Why?

James R. Caplan, Defense Manpower Data Center, Department of Defense

The Defense Manpower Data Center, Department of Defense's central repository for human resources information, handles over 300 requests per day for analysis and personnel data and is also the home of the Department 's personnel survey arm, Survey and Program Evaluation Division. DMDC surveys the various Defense communities, such as active-duty, reservists, their family members, and civilian employees of the Department on a variety of organizational attitudes and opinions such as morale, job satisfaction, and retention intention. All personnel surveys are now either distributed via a Web application or with a combination of paper and Web. DMDC has moved from one large paper and pencil survey that took 2 years between development and results to 10 or more surveys per year, often issuing reports in 60 days or less. When the transition from paper to Web was approved, major questions arose about mode effects, differences in access between postal and Web surveys, and subsequent effects of nonresponse on continuity and validity. To investigate these issues, DMDC undertook a series of research and administrative steps to understand and reduce nonresponse.

How Accurate Are Census Benchmarks Used in Population Forecasts?

J. Gregory Robinson, Population Division, U.S. Census Bureau

Decennial census data serve as the benchmarks or starting points for many population forecasts. These forecasts may vary in the degree of demographic detail, but they all make implicit assumptions about the accuracy of census data in terms of completeness of coverage of the population across demographic groups. In this presentation, Greg Robinson systematically reviews what is known about census coverage up to and including the Census 2000 results, how net coverage differs across demographic groups, and how coverage errors have varied from census to census. This review of levels and patterns of coverage errors will help forecasters in assessing the assumptions made in future forecasts.

Forecasting Demand to Assist in Development of Health Care Policy Alternatives to Close Resource Gaps

Art Klein, Acting Assistant Deputy Under Secretary for Health, Veterans Health Administration, Department of Veterans Affairs

VA health policy decisions makers require timely and accurate information on the determinants of demand and supply of VA health care. They must also test how these determinants affect health care enrollment, utilization, and costs. This is critical for the development of policies that ensure provision of equitable access and quality health care to all veterans at reasonable cost. In his presentation, Art discusses scenario-based approaches VA uses that weigh supply against demand to identify service or resource gaps, and how policies are developed to fill those gaps. While strategic planning focuses on long-term issues impacting veterans health care, decisions are required today in order to ensure VA meets those demands tomorrow.

The National Research Program – Measuring Tax Compliance in a Less Burdensome Fashion

Mark Mazur, Research, Analysis and Statistics, Internal Revenue Service

The fairness of the federal tax system depends in large part upon voluntary compliance. The National Research Program (NRP) is the Internal Revenue Service's comprehensive approach to measuring compliance with U.S. tax law. NRP is designed to capture strategic measures of filing compliance, payment compliance and reporting compliance, and their overall relationship to an estimated $280 billion gross tax gap. At the same time, NRP also reflects a new innovative program which will be far less intrusive and burdensome on taxpayers compared to previous IRS programs designed to measure compliance. This presentation provides an overview of this major research initiative in federal tax administration.

Forecasting Demand to Assist in Development of Health Care Policy Alternatives to Close Resource Gaps

Art Klein, Acting Assistant Deputy Under Secretary for Health
Veterans Health Administration,
Department of Veterans Affairs

Abstract

VA health policy decisions makers require timely and accurate information on the determinants of demand and supply of VA health care. They must also test how these determinants affect health care enrollment, utilization, and costs. This is critical for the development of policies that ensure provision of equitable access and quality health care to all veterans at reasonable cost. In his presentation, Art discusses scenario-based approaches VA uses that weigh supply against demand to identify service or resource gaps, and how policies are developed to fill those gaps. While strategic planning focuses on long-term issues impacting veterans health care, decisions are required today in order to ensure VA meets those demands tomorrow.

Slide 1

Forecasting

Demand to Assist in Development of Health Care Policy Alternatives to Close Resource Gaps

9:30 AM Panel Discussion
The 13th Federal Forecasters Conference
October 27, 2003

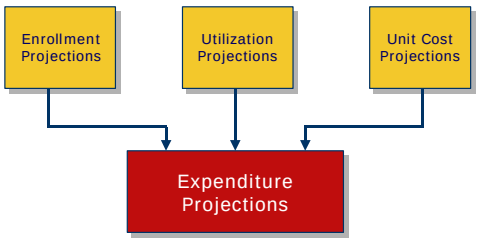
Art Klein, Acting Assistant Deputy Under Secretary for Health
Veterans Health Administration
U.S. Department of Veterans Affairs

Slide 2

VHA Health Care Demand Model

Slide 3

Demand Model Overview



Slide 4

Demand Model Has Evolved Since 1999

Predictive performance has been refined with:

- Detailed data on enrollee reliance on VA, income, insurance coverage from surveys and VA/Medicare data match
- Improved data inputs from VA (pharmacy data)
- Internal and external review
- Planned annual enhancements to methodology

Slide 5

FY 2002 Projected Versus Actual

- Average enrollees +0.09%
- Live-end-of-year enrollees - 0.34%
- Unique patients* + 0.19%
- Obligations* +2.60%

*Adjusted for new enrollees on primary care wait list as of September 2002

Slide 6

Demand Model Now Supports

- Secretary's annual enrollment decision
- Budget formulation
- VA+Choice
- CARES
- Strategic planning
- Future:
  - Real-time interaction
  - Tool for efficiency performance screening
- Model is a powerful tool, we need to be sure we use it appropriately

Slide 7

Turn Information to Insight

Recognize Challenges

Formulate Policies in Support of Strategic Goals

Effect Change

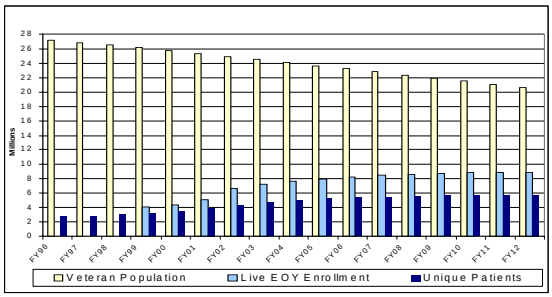
Slide 8

Challenges
of the Last Few Years

- Growing enrollment has exhausted VA's marginal capacity to provide care
- Growing wait lists and longer waiting times
- Aging veteran population
- Rising cost of health care
- Securing resources for a discretionary program in a tight Federal budget climate is difficult

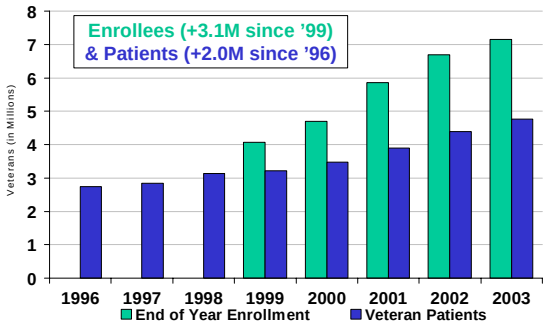
Slide 9

FY 1996-2012 Veterans, Enrollees and Patients



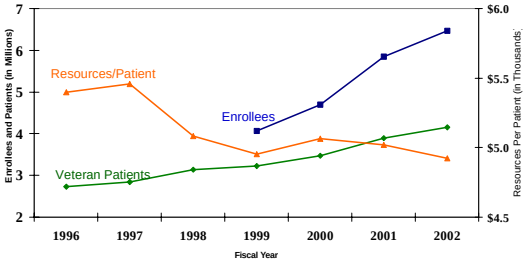
Slide 10

Unprecedented Growth



Slide 11

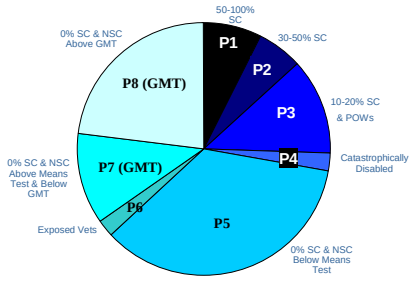
VHA CHALLENGES
Enrollees, Patients & Capacity* 1996-2002



* Capacity defined as resources available per patient

Slide 12

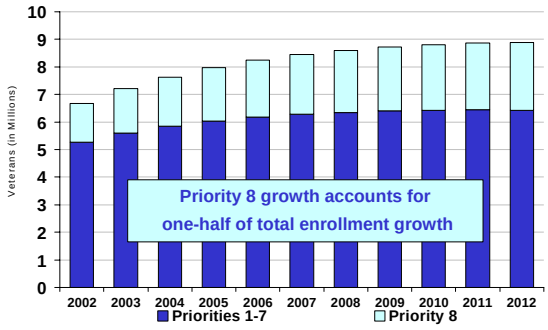
Patient Priorities
Proportioned to Total Enrollment



Slide 13

Enrollment Projections

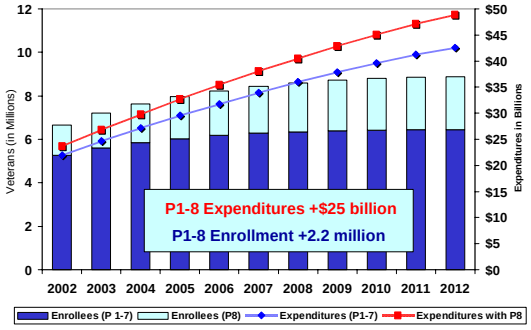
FY 2002-2012



Slide 14

Projections for Priorities 1-8

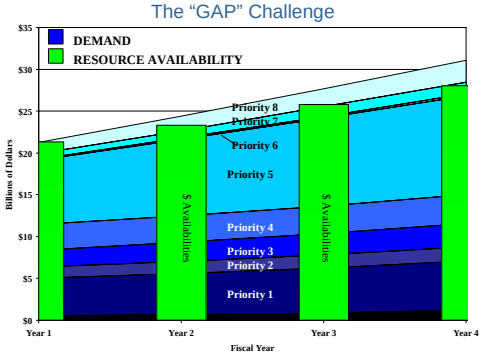
FY 2002-2012



Slide 15

Demand and Resource Availability

The "GAP" Challenge



Slide 16

Strategic Policies to Close Gap

Between Demand and Resource Availability (Supply)

- Enrollment Actions
- Services Provided
- Cost Sharing Proposals
- Efficiencies
- Resources

Slide 17

Enrollment Actions

Impact Demand (Closing the Gap)

- Disenroll
- Stop New Enrollment
- Create Sub-categories Within Priorities

Slide 18

Services Provided

Impact Demand (Closing the Gap)

- Limit Services Available
  - By priority?
  - Medical benefits or discretionary?
  - Regulation or legislation?

Slide 19

Cost Sharing Actions

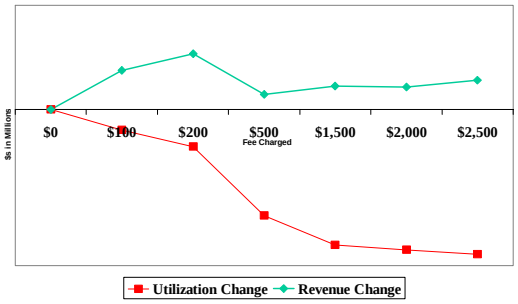
Impact Demand and Supply (Closing the Gap)

- Enrollment Fee
- First-use Fee
- Deductible
- Co-payments
  - Inpatient
  - Outpatient (preventative/primary/specialty care)
  - Pharmacy

Slide 20

Enrollment Fee

Impact at Various Fee Levels



Slide 21

Efficiencies

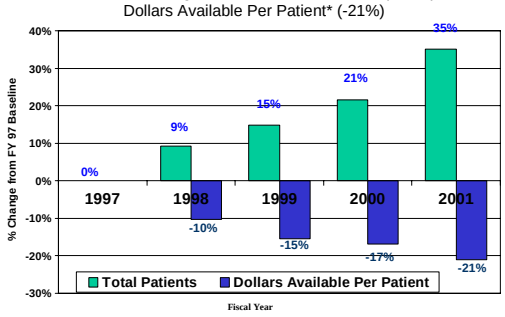
Impact Supply and Resources (Closing the Gap)

- Clinical
- Administrative/Operational

Slide 22

Medical Care Resource Efficiencies

Cumulative % Change from '97 Base - Patients (+35%) and Dollars Available Per Patient* (-21%)

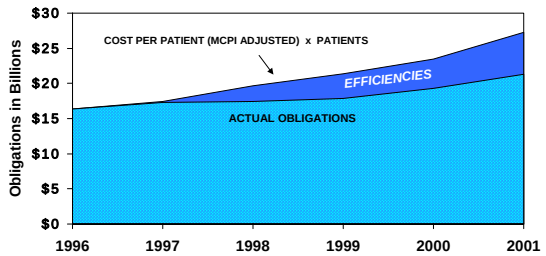


* In constant FY'97 dollar terms, adjusted for Medical Consumer Price Index (CPI)

Slide 23

Medical Care Resource Efficiencies

Clinical and Administrative, FY 1996-2001



Slide 24

Other Balancing Considerations

Impact Demand and Supply (Closing the Gap)

- Access to Care
  - Inaction could result in longer waits for care
- Capacity Issues
  - Primary care staffing
  - Infrastructure
- Budget
  - Supplemental
  - Adjust budget requests

Slide 25

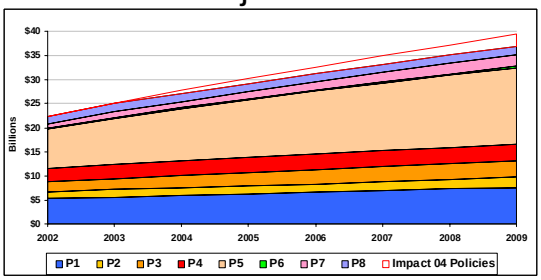
FY 2004 President's Budget Policies

Needed to Close the Gap

- Suspend Priority 8 Enrollment (January 17, 2003)
- $250 Enrollment Fee on P7c-P8 (Oct. 1, 2003)
- Increase Copays on P7c-P8 (Oct. 1, 2003)
  - Increase Primary care copay from $15 to $20
  - Increase Rx Copay from $7 to $15
- Efficiencies
- Record breaking resource request

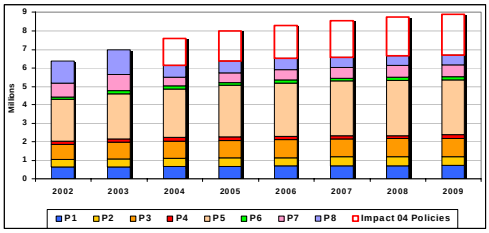
Slide 26

FY 2002-2009 Resource Projections



Slide 27

FY 2002-2009 Enrollment Projections



Slide 28

Turn Information to Insight

Recognize Challenges

Formulate Policies in Support of Strategic Goals

Effect Change

Slide 29

Art Klein,
Acting Assistant Deputy Under Secretary for Health
Veterans Health Administration
U.S. Department of Veterans Affairs

202.273.8934
art.klein@hq.med.va.gov

The National Research Program–Measuring Tax Compliance in a Less Burdensome Fashion

Mark Mazur, Research, Analysis, and Statistics, Internal Revenue Service

Abstract

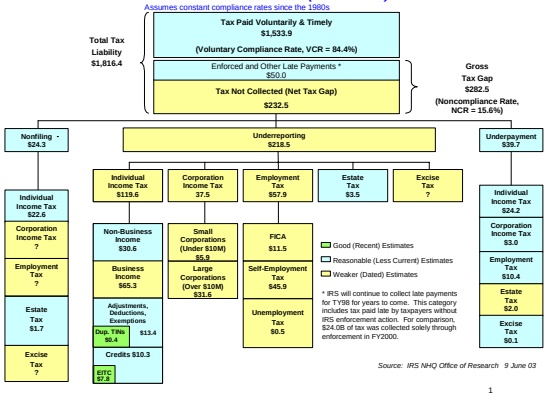
The fairness of the federal tax system depends in large part upon voluntary compliance. The National Research Program (NRP) is the Internal Revenue Service's comprehensive approach to measuring compliance with U.S. tax law. NRP is designed to capture strategic measures of filing compliance, payment compliance and reporting compliance, and their overall relationship to an estimated $280 billion gross tax gap. At the same time, NRP also reflects a new innovative program which will be far less intrusive and burdensome on taxpayers compared to previous IRS programs designed to measure compliance. This presentation provides an overview of this major research initiative in federal tax administration.

Slide 1

Slide 2

TAX GAP MAP FOR TAX YEAR 1998 *($ in Billions)*

Assumes constant compliance rates since the 1980s



1

Slide 3

3 MEASURES OF TAX COMPLIANCE

- There are three mutually exclusive measures of compliance that, taken together, cover all types of tax compliance.
- These are:
  - Filing Compliance
  - Reporting Compliance
  - Payment Compliance

2

Slide 4

NRP IS CURRENTLY ENGAGED IN A STUDY TO MEASURE THE REPORTING COMPLIANCE BEHAVIOR OF INDIVIDUAL TAXPAYERS.

- The study will allow NRP to develop the **Voluntary Reporting Rate** (VRR), which is the percentage of the total tax required to be reported that taxpayers voluntarily reported on their timely-filed returns.

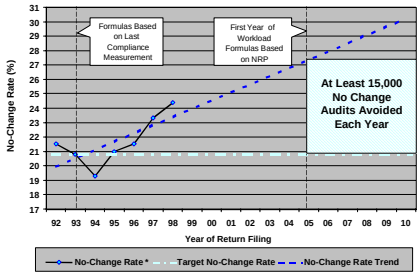
$$VRR = \frac{\text{Total Tax Reported on Timely Filed Returns}}{\text{Total Tax Reported + Estimate of Tax Misreported}} \times 100$$

- NRP's objectives with regard to reporting compliance include:
  - measuring an overall reporting compliance rate for the IRS, as well as for the SB/SE and W&I Operating Divisions;
  - update workload selection formulas for audits;
  - provide IRS Operating Divisions with results pertaining to their taxpayers so they can establish strategies, implementation plans and performance measures; and
  - improve the IRS's ability to detect noncompliance and develop appropriate cost-effective treatments for prevention and early intervention.

3

Slide 5

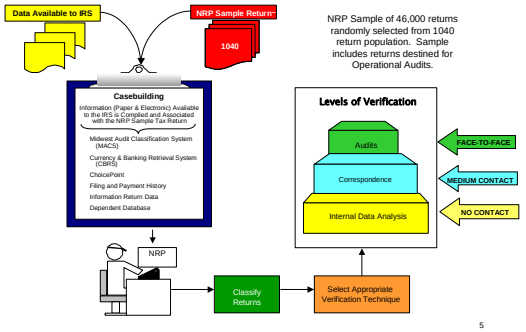
WITH NRP, THE IRS WILL AUDIT FEWER COMPLIANT TAXPAYERS.



4

Slide 6

IN A DEPARTURE FROM EARLIER REPORTING COMPLIANCE STUDIES, NRP WILL RELY HEAVILY ON CASEBUILDING AND CLASSIFICATION PROCEDURES PRIOR TO CONTACTING TAXPAYERS.



5

Slide 7

THE REPORTING COMPLIANCE STUDY IS UNDERWAY.

- NRP has designed a sample of 46,000 randomly selected Form 1040 returns, which will represent most Form 1040 filers (except international filers), including:
  - 95 million served by the W&I Operating Division; and
  - 35 million served by the SB/SE Operating Division.
- NRP processing of cases is well underway. Through September 2003, we have
  - Selected all the individual returns in the sample;
  - Determined type of IRS contact required – accept as is, correspondence, or face-to-face
  - Identified specific issues on those returns that need to be examined;
  - mailed letters to nearly 35,000 taxpayers requiring examination;
  - completed examinations of nearly 15,000 cases; and
  - perfected data on over 11,300 cases and entered them onto final NRP database.
- Examinations began in the Fall 2002 and will continue through FY 2004.
- Final data from NRP reporting compliance study expected by the end of 2004, with more thorough analyses and use of the data occurring in 2005.
- New audit selection formulas likely to be implemented in early 2006.

6

CONCURRENT SESSIONS I

Measurement and Projection Issues in Federal Tax Administration

Chair: Russell Geiman, Internal Revenue Service, U.S. Department of the Treasury
Discussant: Bruce Colton, Bureau of Transportation Statistics, U.S. Department of Transportation

Measuring Compliance on Individual Tax Returns: The National Research Program Sample Design

Karen Masken, Internal Revenue Service, U.S. Department of the Treasury

The fairness of the U.S. Individual Income Tax system depends upon individuals to accurately self-report their income, credits, and deductions on their tax returns. The National Research Program (NRP) reflects a new approach by the IRS to systematically measure the extent to which U.S. taxpayers meet their voluntary reporting requirement. This paper looks at the sample design behind this major research effort. It discusses the variables used for sample stratification and their tie to the survey objectives. It also examines the role of projections in calculating sample sizes and allocating the sample across strata and how the actual sample has fallen out.

Emerging Trends That Can Distort Research Results Secured by Surveys

Henry Sloan, Internal Revenue Service, U.S. Department of the Treasury

The media have reported that many of the projections made during the 2002 political campaign were inaccurate, due to the failure of pollsters to take into account emerging technological, social, and demographic trends. This paper looks at three current trends that may have been ignored by political surveys, but are of interest to survey researchers: The increasing use of cell phones; the continuing increase in diversity among the foreign-born within the United States; and the increasing complexity of age groups in America. Findings suggest accounting for emerging and persistent patterns of behavior that, if ignored, could distort survey findings.

Partnerships Naughty and Nice - Forecasting Partnership Filings Received by the Internal Revenue Service

Daniel Killingsworth, Internal Revenue Service, U.S. Department of the Treasury

The news media have recently highlighted the use of partnerships as a means of sheltering income, some to the point of abuse. However, partnerships, in general, are often a logical economic business formation and can even serve as a legitimate tax-saving device. And the ups and downs in the popularity of partnerships actually comprise an interesting story spanning several decades. This paper takes a look at the historical trend in the filings of partnership tax returns and some of the factors affecting that trend. It also considers some of the alternative forecasting models used by IRS staff to prepare projections for this return series.

Forecasting Employment Levels of IRS Compliance Staff with a Micro Model of Exits and Job Changes

Thomas Mielke and Alex Turk, Internal Revenue Service, U.S. Department of the Treasury

In five years, over 50 percent of the IRS's current workforce will be eligible for either full or early retirement. Complicating matters is the fact that only limited hiring has occurred since the early 1990s. A micro model of quits and job changes is used to forecast attrition rates for three specific occupations within the IRS. Employment data for 1997-2002 is used to develop the model. This period includes a major re-organization of the IRS which resulted in a large number job changes and separations. The impact of the reorganization is thus accounted for in the data development, model specification and the subsequent forecasts.

Measuring Compliance on Individual Tax Returns: The National Research Program Sample Design

Karen Masken, Internal Revenue Service

The fairness of the U.S. Individual Income Tax system depends in large part upon individuals to accurately self-report their income, credits and deductions on their tax returns. The National Research Program (NRP) reflects a new approach by the Internal Revenue Service (IRS) to systematically measure the extent to which U.S. taxpayers meet their voluntary reporting requirement. This paper takes a close look at the sample design behind this major research effort to measure tax compliance. It discusses the variables used for stratification in the sample and their tie to the major objectives in doing this survey. It also examines the role of projections in calculating sample sizes and allocating the sample across strata and how the actual sample has fallen out.

Background

The IRS uses the results of its compliance studies to aid in the development of strategic planning as well as to develop workload selection formulas. In terms of strategic measures, the IRS considers compliance on three core dimensions: compliance with **filing** required tax returns; compliance with **paying** taxes on timely filed returns; and compliance with properly **reporting** income and deductions on their returns. The last comprehensive **reporting** compliance study on individual income tax was conducted on tax year 1988 returns filed in 1989. While several efforts were made in the 1990's to conduct a new study, none were ultimately implemented. The current NRP study examines individual returns for tax year 2001 that were filed in 2002.

Sample Design

The sample design is a stratified random sample, meant to address two requirements. The IRS splits individual income tax filers between two of its operating divisions: Small Business/Self Employed (SBSE) or Wage and Investment (W&I). Generally speaking, Form 1040 returns with a Schedule C (Profit or Loss from Business), Schedule F (Profit or Loss from Farming), Schedule E (Supplemental Income or Loss), or Form 2106 (Employee Business Expenses) attached are designated as SBSE, the remainder are W&I. The first requirement of the NRP sample is to be able to detect a 0.5% difference in the voluntary reporting rate (tax reported/ (tax reported + change in tax liability after audit)) for each of the operating divisions from this study to the next. The second requirement is to update the workload selection models used to identify the returns most in need of IRS examination. This is primarily for SBSE cases, but affects several of the W&I strata as well.

Sampling Population

The sampling population is all Form 1040 series returns – U.S. Individual Income Tax Return filers for tax year 2001 who file in processing year 2002, with the exception of duplicate or amended returns, or international filers. There are about thirty-five million taxpayers serviced by SBSE and ninety-five million serviced by W&I for a total of approximately 130 million individual taxpayers.

Strata

The NRP sample is a stratified random sample. The returns are initially divided into their respective operating division (SBSE or W&I). Then, based on research conducted over the years by several outside consultants, the next level of stratification corresponds to the current definition of IRS' "examination class". Examination classes are unique and mutually exclusive groupings of taxpayer returns by common sets of characteristics that help the IRS determine which returns are most in need of audit for compliance purposes. For non-business returns, the examination class is generally based on Total Positive Income (TPI), while for business (Schedule C) and farm (Schedule F) returns it is based on Total Gross Receipts (TGR). In some instances, returns with relatively small amounts of Schedule C or Schedule F income are classified as non-business returns. These examination classifications are then further split into subgroups typically based on dollar ranges for Adjusted Gross Income (AGI). For NRP, there are twenty strata within SBSE and ten within W&I (see Appendix for specific definitions of each stratum).

Allocation

The first step in the sample design process was to project the population for each stratum. These projections were based on the examination and operating division projections published by IRS' Projections and Forecasting Group (PFG) in the Fall 2001 updates of IRS Document 6187 and Document 6186. Some additional calculations had to be made

to prorate the projected populations into the subgroups within the examination classes.

Once the strata populations were projected, the sample was allocated across the strata in an iterative process. The first iteration was to allocate the sample so that the sample size within each stratum was proportional to its respective square-root of the mean voluntary reporting rate (VRR). The VRR's within each stratum were estimated from the prior compliance study of tax year 1988 returns. Due to the age of the data, we did not want to rely too heavily on the variance estimates of the VRR, which is why the more conventional method of Neyman allocation was not used.

In order to model new workload selection formulas, a minimum sample size for each stratum was calculated next based on their "profitability to audit" factor (an IRS metric that considers the likely change in tax liability from an audit to the IRS resource cost to conduct the examination). The effect of the second iteration was to review each stratum and increase the sample size, where necessary, to the minimum required for the workload selection modeling.

Finally, coefficients of variation (CV) were calculated for each stratum for the voluntary reporting rate to ensure that it was less then ten percent. In cases where the CV was greater than ten percent, the sample size was increased within that stratum to meet this requirement. The 8^{th} column in the appendix presents the final design sample size expected by stratum.

Sampling Technique

Returns were sampled as soon as they were processed and posted to the IRS electronic master file system. The primary Social Security Number (SSN) was transformed into a permanent random number, and that random number was used to determine if the return fell into the sample or not. One issue with this technique is that the IRS Statistics of Income (SOI) Division uses the same method to draw their sample of individual filers for their unique statistical purposes (which are independent of NRP). The starting point for their sampling intervals is always zero. Therefore, in an effort to reduce the overlap between the two samples, the starting point for each of the NRP sampling intervals was chosen randomly. Once a return was selected for the NRP sample, it was retrieved from IRS files, photocopied and sent to the case-building group within IRS.

Case-building

One of the obstacles in implementing a new study was that the conventional compliance study audits were excessively burdensome on the taxpayer. Traditionally, the compliance study required examiners to conduct line-by-line audits where the taxpayer had to provide documentation to support each line item entry on the return above a nominal amount. To address this issue, IRS developed a new approach with NRP that is less burdensome to the taxpayer. Instead of obliging the taxpayer to substantiate the entire return, more extensive IRS and third-party databases were used to try and verify as much information on the return as possible before contacting the taxpayer. Under this "case-building" phase, IRS staff compiled all of the available third party data and then the cases were sent to the "classification" group.

Classification

One of the most significant changes implemented in the current study was the introduction of a classification step. In the past, all sampled returns were audited in a face-to-face meeting with the taxpayer. In the current study, examiners review the tax return, along with the supporting documentation put together by the case-building group, and make a determination about which line items on the return are significant enough to require further investigation, and which can be verified without contacting the taxpayer. The examiner can then classify the return into one of three audit categories. They may decide that all line items can be verified without any further documentation and therefore the return is accepted as filed. In these cases, the taxpayer will never know that they were part of the compliance study.

Alternatively, the examiner may find only one or two issues and determine that the audit can be conducted in writing or over the phone. In these cases, the return is classified as "correspondence examination". It reduces the burden on the taxpayer in that they do not have to set up a meeting with the auditor, and they are being asked about only a few line items.

Finally, the examiner may determine that the return requires a "face-to-face examination", in which case the taxpayer is contacted and an appointment is made to meet face-to-face. This is still less burdensome than in previous studies because the examiner will ask for supporting documentation only for line items that seem suspect, not all of them. It is, in effect, an examination of "limited scope".

Calibration

There was concern that by departing from the traditional method of line-by-line audits, the new NRP method might bias the results of the compliance study from an historical measurement perspective. In an effort to address this concern, it was decided that a small "calibration" sample would be drawn for each of the three classification categories. Cases in the calibration sample will be examined in a traditional line-by-line, face-to-face manner. The results will then be compared within the three classification groupings to see if there are any significant differences between the NRP results and those from the calibration samples.

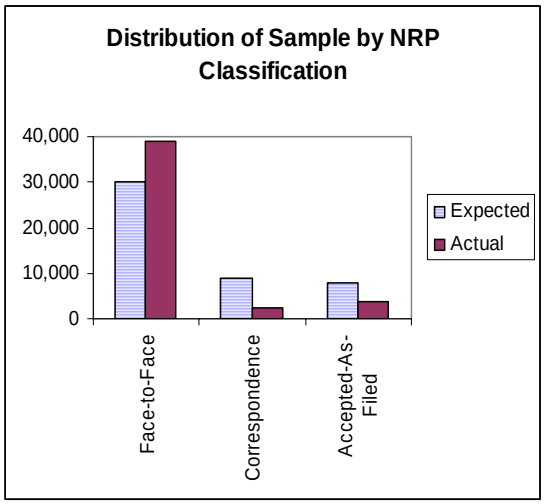
During the design stage, some estimates were made on the proportion of returns that would fall into each classification category within each stratum. These estimates were then used to determine how large a sample would be needed to get enough of each type, recognizing that two of the categories would have to be oversampled in order to get enough returns in the smallest expected category (i.e., correspondence). Returns in the calibration sample were classified, and then subsamples of the accepted-as-filed and face-to-face categories were drawn. To ensure that the calibration sample was a representative sample, the subsamples were designed so that the proportion of sampled returns in each stratum was equal to the population proportion.

Issues

The biggest issue in designing the sample was that the compliance measurement data on which the design was based were thirteen years old. It was acknowledged that the characteristics of the population and the distribution of income had changed considerably since the last study, but there were no good alternatives. There were some adjustments made to age the data and account for the change in income distribution. But for the most part, the sample design had to rely upon old data.

Another issue was that there were some significant differences in what was projected for the strata populations versus the final actual volumes. The major cause of this was the sluggish economic recovery. The overall return volume for Tax Year 2001 slowed noticeably from the previous year. While this was not totally unexpected, the magnitude of the slow down was a development not fully captured in IRS projections. In addition, for the returns that were filed, there was a shift from the high income strata into the lower income ones. As a result, the sampling rates for the high income strata had to be increased during the year. Fortunately, it was possible to go back retrospectively and pick up all returns that would have been selected under the new rate.

As of this writing, virtually all sampled returns have been classified and the majority of audits have been initiated. One outcome is that the distribution by classification of returns did not fall out as expected. A much larger proportion of returns were classified as face-to-face than expected, which has created some workload issues for IRS examiners. The following chart presents a comparison between the expected classifications and the actual results for the face-to-face, correspondence and accepted-as-filed groupings. This outcome was due primarily to not having any solid data to make the initial estimate.



Conclusion

Results from the current NRP are expected in late 2004. The current schedules call for preliminary data by June 2004 and the final NRP database available by December 2004. These results will provide the IRS with up-to-date reporting compliance measures for individual returns that are integral to the strategic planning and budgeting process. The method used to collect the data is less burdensome on the taxpayers than in previous studies, and will assist the IRS in meeting its mission to apply the tax law with fairness to all.

Acknowledgments

The views expressed in this article represent the opinions and conclusions of the author, they do not represent the opinion of the Internal Revenue Service. This article would not have been possible without the help of Ron Bartyczak, who is largely responsible for the sample design descried in this paper. Special thanks also go to Russell Geiman for his encouragement in writing this article.

References

Cochran, William G., *Sampling Techniques,* Third Edition, John Wiley & Sons, Inc. New York, NY, 1977.

Internal Revenue Service, National Research Program, *Presentation Kit,* July 2002.

General Accounting Office, *IRS is Implementing the NationalResearch Program as Planned,* June 2003.

EMERGING TRENDS THAT CAN DISTORT RESEARCH RESULTS SECURED BY SURVEYS

Henry Sloan, Department of Treasury, Internal Revenue Service

INTRODUCTION

The Wall Street Journal (WSJ) reports that many projections and forecasts made by political pollsters during the 2002 elections were inaccurate and misleading.[1] The WSJ provides a laundry list of possible causes, including such likely contributors as the failure to take into account statistical precision/margin of error when interpreting results; the growing wariness on the part of the populace to take surveys; the use of Caller ID to screen out researchers; a greater participation of women in the work force, which leaves fewer people at home to answer survey questions; a reluctance on the part of researchers to pursue "expensive" call-backs; a greater willingness on the part of researchers to use convenience samples; and a growing interest in Internet surveys that could be non-representative of the general population.

The WSJ also suggests that at least some inaccuracies might result from the failure of political survey researchers to take into account emerging and on-going technological, demographic, and social trends. However, the need to consider such trends clearly applies not only to researchers in the political arena but also to researchers, and survey consumers, in general.

With this in mind, this paper examines three important trends that might have an adverse impact on survey results, if ignored by researchers:

1. The increasing use of cell phones.
2. Diversity in immigration patterns.
3. Differences among age groups.

INCREASING USE OF CELL PHONES

Although the number of cell phones in use varies by source, more and more people appear to be using them – and fewer and fewer seem to be using other means of communication. According to EPM Communication, U. S. mobile phone subscribers increased from 7.6 million in 1991 to 109 million in 2000.[2] CNN.com/technology suggests approximately 7.5 million wireless subscribers have abandoned their home phone to go strictly cellular.[3] The number of U. S. landline phones has dropped more than 5 million, or nearly 3 percent, since 2000. Currently, cell phones comprise about 43 percent of all U.S. phones. Some predict that by 2010, 25 percent of all voice calls will be made using wireless technology.[4] The switch to cell phones is particularly noticeable among younger age groups, with college students increasingly using cell phones while at the same time abandoning traditional long distance services.[5] (See Chart 1 for projected growth of cell phone adoption by U. S. households.)

Impact on Research:

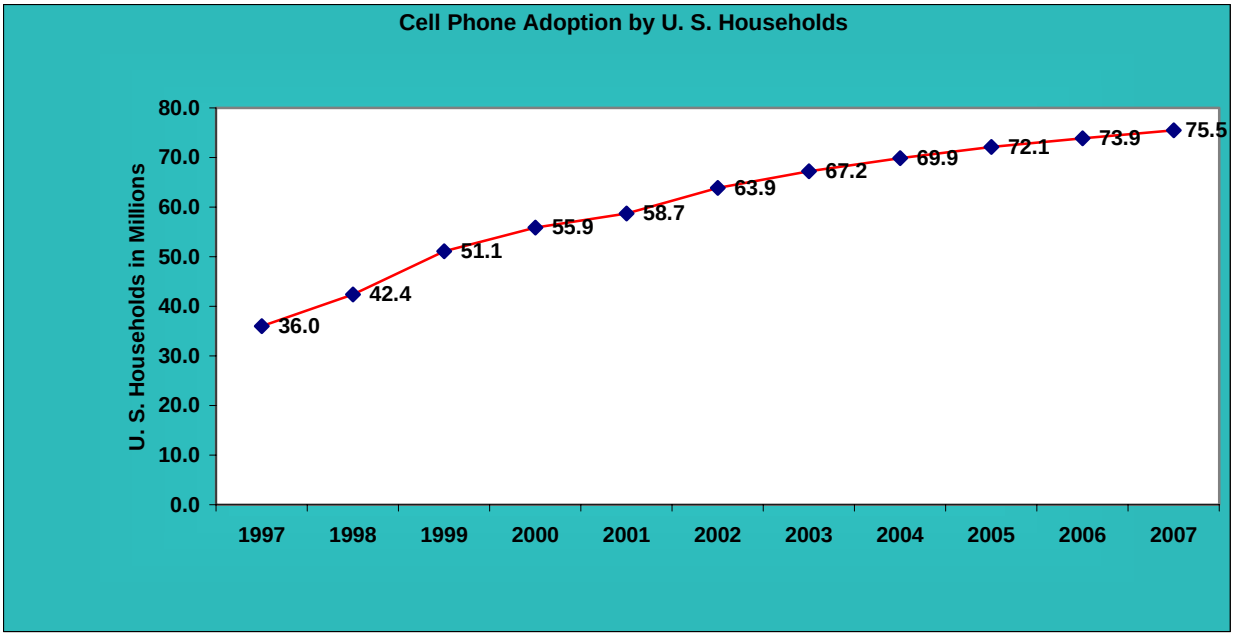
From a behavioral perspective, it is fair to suggest that technology interacts with lifestyle choices to produce new modes of behavior and new population subgroups. If so, cell phone use provides an example of that phenomenon. Researchers who are not sensitive to such trends run the risk of inadvertently excluding, or poorly representing, developing subgroups in their studies, especially if they are depending on traditional landline telephone surveys to collect data. Despite the use of random digit dialing, some conventional telephone surveys might still rely to some extent on standard telephone listings as sources for sampling frames. Consequently, researchers, in general, should be aware that:

1. Cell phone users might not be included in the usual listings (i.e., phone books).
2. Adequate sampling frames for cell phone users might be difficult to find.
3. Because individuals might have multiple phones and multiple numbers for cell phone and landline services, potential sample members could be counted in the population more than once, thereby affecting sample selection probabilities.

An example:

The following example illustrates the potential for possibly under-representing a survey subgroup. While analyzing differences between adopters and non-adopters of Internal Revenue Service (IRS) electronic tax services, IRS researchers compared the results of a survey of taxpayers with a population database roughly equivalent to the survey population. Overall, the survey

Chart 1



SOURCE: Forrester.com

age estimations matched the database population figures fairly closely. However, there appeared to be substantial differences between sample and population percentages of taxpayers 18-24 (about 8 percent in the sample and 20 percent in the population database). One might wonder whether these differences reflect systematic under-representation of younger cell phone users in the survey. Further analysis, of course, is necessary to determine the merit of this argument, but the issue will be addressed when the next IRS tax survey of its kind enters the design stage.

DIVERSITY IN IMMIGRATION PATTERNS

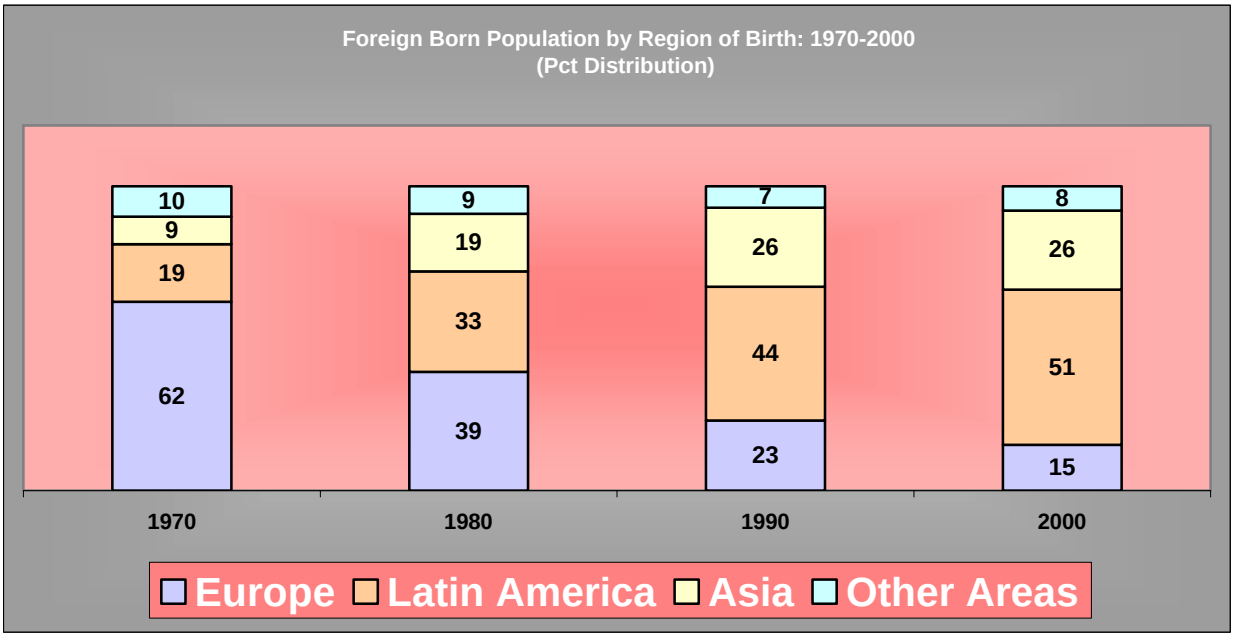
The presence of the foreign-born is increasing in the United States. The Census Bureau projects the number of foreign-born to grow from 28 million in 2000 to about 34 million in 2010 – a 20 percent increase. This converts to about 10 percent of the U.S. population in 2000 and 11 percent in 2010.

Asians and Latin Americans are the two largest foreign-born groups in the U.S. - Asians increased from 9 percent of the foreign-born population in the U.S. in 1970 to 26 percent in 2000. Latin Americans, as a group, grew from 19 percent in 1970 to 51 percent in 2000. In contrast, the European percentage of the foreign-born decreased from 62 percent in 1970 to 15 percent in 2000. (See chart 2 for the percentage distribution of foreign-born populations by region of birth, 1970 to 2000.)

Researchers usually consider demographic and socioeconomic differences between foreign-born groups when developing research designs and sampling plans. But differences *within* foreign-born groups should be investigated as well. For example, Hispanics are sometimes viewed as one broad group with little variation among constituents. However, the Hispanic foreign-born do show some differences, especially with regard to age distribution, educational attainment, and income level. (See Chart 3 for a breakdown of the Hispanic population by sub-category.)

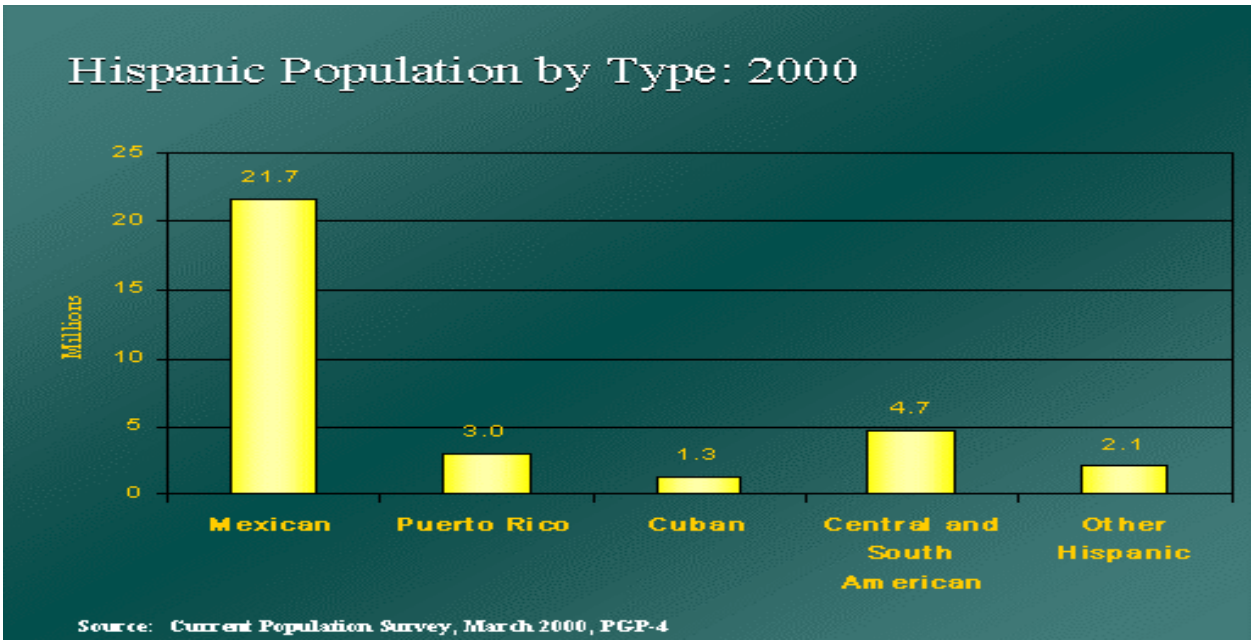
Indeed, Census data indicates that 38 percent of Mexicans and 19 percent of Cubans are under age 18, while 4 percent of Mexicans and 21 percent of Cubans are over age 65. Educationally, 32 percent of Mexicans have stopped attending school before reaching the 9th grade, as compared to 18 percent of both Cubans and Puerto Ricans. Seven percent of Mexicans have a Bachelor's degree or higher while 23 percent of Cubans have a Bachelor's degree or better. Seventy-two percent of Mexicans had full time year round earnings of less than $30 thousand in 1999. Fifty-six percent of Cubans had earnings less than $30 thousand in 1999 – but 18 percent of Cubans made more than $50 thousand in 1999. These differences are important because they influence, and are, in turn, influenced by life experiences that permeate subgroup identification, motivation, cultural orientation, fears, and

Chart 2



SOURCE: U. S. Census Bureau

Chart 3



behavioral practices.

Further, the problematic nature of these differences are compounded by the fact that many immigrants are limited in their ability to speak and understand English. In fact, many have self-reported to the U. S. Census Bureau that they speak English "not well" or "not at all." Many are not literate in their own language, and 23 percent are not literate in any language at any level.[6]

Moreover, even after decades in the U. S., many immigrants are geographically and linguistically isolated from mainstream American society. They have not, or will not, learn English. The reasons vary, of course: Some immigrants do not have the time or aptitude for learning languages; ties to native countries ameliorates motivation for some; and living and working amidst "language islands" isolates others from the need to speak English.[7] As a result, the obstacles researchers might face in obtaining sample respondents, communicating effectively with them, and fairly representing their views, opinions, and concerns are increased.

If, as the Census Bureau suggests, the United States will become a nation of minorities by 2005, researchers must consider the importance of the following issues:

1. It will continue to be important for researchers to obtain an accurate statistical sampling of immigrants.

2. Researchers will face the challenge of how best to weight various ethnic groups and subgroups to capture needed sample estimates; an assessment of one set of weights as opposed to another could change the outcome of one's estimates.

3. In the case of surveys and other behavioral studies, researchers will have to become even more sensitive to cultural differences among participants. This will have an effect on survey construction, where questions often are already subject to distortion and misunderstanding. Differences in interpretation represent a confounding factor that has to be carefully considered and addressed.

4. Response rates among immigrants may be low, due to difficulties with language, or the inability of researchers to locate and successfully engage them verbally.

5. While always necessary, the need for adequate follow-up might increase in urgency. Non-respondents tend to be "different" on particular survey items from those who do respond. Cultural and social diversity may very well broaden and intensify those differences.

DIFFERENCES AMONG AGE GROUPS

Researchers are increasingly aware that age plays an important role beyond the simple passage of time; that the social and psychological environment in which a person "comes of age" marks that person for life.

Generational Cohorts:

Age cohorts are categorized by external and internal events common to the life of cohort members. However, age boundaries limiting the span of particular generations are not always rigid. For example, Baby Boomers are variously defined as being born between 1943-1960 or 1946-1964[8]. Despite such imprecision, cohort characteristics remain more or less consistent.

Examples:

1. The Silent Generation: Their coming of age occurred somewhere between 1930 and 1939. The members of this group tend to be conservative, conformist and oriented toward the past. Their defining moment was the Great Depression.

2. Leading-Edge Boomers: With their coming of age, between 1963 and 1972, cohort members tend to be concerned with convenience and social justice. They also tend to be big spenders and interested in pursuing second careers. Their defining moments include the assassination of JFK, Vietnam, and the landing on the moon.

3. Generation Y: This generation's coming of age began in 1995 and should run through about 2013. This cohort tends to respect authority, but members are easily bored and easily distracted. Their defining moments include the Internet, the Columbine School shootings, Clinton's impeachment and September 11th.

Cohort identification provides some substance and color to typical age range category analysis. Why do people in certain age ranges behave as they do? How are they really different from individuals in other age ranges?

Of course, as some cohorts leave the stage, others take their place. Over time, the values and practices of younger cohorts replace those of older ones. Until then, they co-exist in a complex mixture of values, attitudes, and behavior. Thus, individual age range groups may be occupied by several cohorts at once – meaning not

all people in a specific age range will have the same perspectives, concerns and loyalties. As such, they should not necessarily be treated uniformly.

It is expected that between 2001 and 2010 cohort migration will impact age ranges in the following manner.[9] (See Chart 4 for projected age group population changes between 2001 and 2010.)

1. The Silent Generation will increase the 75-plus age category by approximately 10 percent.

2. Leading-Edge Boomers will increase the 55-64 age category by 44 percent.

3. Trailing-Edge Boomers will increase the 45-54 age group by 14 percent.

4. Generation X will replace Boomers in the 35-44 age group, a loss of 11 percent.

5. Generation Y will move into their 20s and 30s, thereby increasing the 18-34 age group by 8 percent.

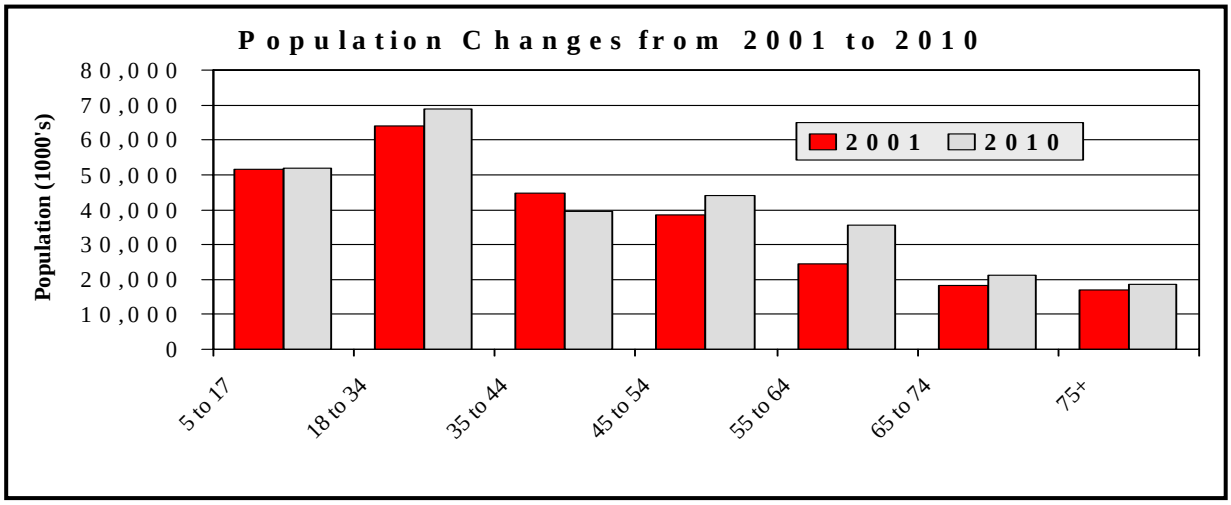
cohorts, however, might work for or against good sampling and survey results:

1. Trailing-Edge Boomers (ages 36-46), for example, value privacy and convenience. It is possible that any survey of this cohort will encounter a high level of non-response. In the case of telephone surveys, members of this cohort might have their numbers unlisted or use Caller ID or some other blocking device. And, since many do not want to be inconvenienced, Trailing-edge Boomer respondents might be more likely to prematurely end the survey or simply hang-up the telephone.

2. Generations X (ages 25-35) and Y (ages 24 and under) also might be difficult to pin down. Being technologically oriented, their commitment to cell phones could make them harder to find.

3. Members of the Silent Generation are conservative in nature and have strong feelings with regard to issues of moral obligation, trust in government, authority, and social conformity. Members of this cohort might be at home more often than others,

Chart 4



SOURCE: U. S. Census Bureau

Impact of Cohort Worldviews on Data Collection:

Researchers might want to consider the impact of worldviews on behavior. This will require an increasing sensitivity to differences among cohorts. At the same time, sampling techniques for studies concerned with age and behavior must ensure various cohorts are fairly represented. This will require updating the age mix in target populations in order to retain their cohort identity. The very nature of some and, in an effort to be good citizens, might be more likely to make a concerted effort to comply with requests to answer non-personal, non-threatening survey questions.

A Practical Application of Cohort Analysis in the Work-a-Day World:

It is important to be aware of cohort differences where applicable in the survey process. However, the

usefulness of this awareness can be found in practical situations, as well. Research finds cohort identity analysis to be beneficial when it comes to hiring, training, and retaining employees. In recent years, organizations have become flatter, with employees retiring later, seeking less stressful jobs, and facing rapid technological change. As a result, more cross-generational cohorts are working together. Since there are differences in values and philosophies among different cohorts, understanding where each cohort is "coming from" could be beneficial to recruiters, trainers, and employers who want to find the right employees for particular jobs, train them to perform successfully, and encourage them to stay within the organization as long as it benefits all concerned. As such, when it comes to motivating different generations, different messages work better for one group than another. Older cohorts want to feel respected for past experience in the workforce; Boomers want to feel they are important and that their contributions are meaningful; members of Generation-X favor an environment with few rules; and Generation-Y wants to work with bright and creative people.

CONCLUSIONS AND IMPLICATIONS

Society is not a monolithic entity. It comprises a population of diverse and complex individuals characterized by a multitude of attitudes, motivations and behaviors. As such, it is advisable that when appropriate, researchers and non-researchers keep in mind the following when analyzing any survey data.[10]

1. There are emerging trends and persistent patterns of behavior that impact on research and analysis. Ignoring their influence could negatively affect findings, interpretations, and conclusions.

2. It is helpful to understand social-psychological, cultural, and technological traits that might influence respondent answers to survey questions.

3. It is necessary to understand how these traits interact with available data collection techniques and analytical methods to produce information that may be valuable to researchers, or casual users, of survey data.

END NOTES

[1]Harwood, John and Shirley Leung. "Hang-Ups: Why Some Pollsters Got It So Wrong This Election Day. Cell phones and Caller ID Make It Harder to Get A Good Sample of Voters. Surprises in Georgia, Colorado." The Wall Street Journal, November 8, 2002.

[2]EPM Communication. Analysis of Cellular Telecommunications and Internet Association data. No date.

[3]CNN.com/Technology. "Bye, Bye Landline Phones." August 4, 2003.

[4]Trend Letter. "Callers Cut the Cord on Landline Telephones." Vol. 21. No. 22. October 28, 2002.

[5]NYTimes.com. January 20, 2003.

[6]Olson, Roy. "Multilingualism in the United States." July 2001.

[7]Olson, Roy." How do People Learn English?" March 2003.

[8]Anderson, Kay. "Taxpayer of the Future." January 2003.

[9]Larsen, Eric J. "Cohort Marketing May Assist in Understanding and Catering to Taxpayers In The Future." May 2002.

[10]Although the author works for the Internal Revenue Service, the views expressed in this paper represent only the opinions and conclusions of the author. They do not necessarily reflect the opinions of the IRS.

PARTNERSHIPS NAUGHTY AND NICE - FORECASTING PARTNERSHIP FILINGS RECEIVED BY THE INTERNAL REVENUE SERVICE

Daniel Killingsworth, Internal Revenue Service, U.S. Department of the Treasury

Introduction

In the last year or so, there have been several media stories on accounting scandals. Tales of accounting irregularities have caused some publicly traded firms to plummet in share price, causing worry for investors and future retirees. This paper will examine forecasting the number of partnerships, a type of business organization, at the IRS in terms of the number of Form 1065 (U.S. Income Tax for Partnerships) tax returns filed annually. Forecasting this series is quite interesting, partly because partnerships were at the center of some of these accounting scandals. The following paper will examine a rationale for partnership organizations, a history of partnership return filings over the last 40 years, as well as ways these business entities can be manipulated for tax evasion. The current method that the IRS uses to forecast the number of partnerships will then be noted, including implications for future partnership forecasts in the current environment.

Background

One of the better known media stories about the accounting scandals features a case of the use of naughty partnerships. While this may not seem to be the customary word used to describe such a situation, it seems to catch the flavor of what occurred quite well. The word "naughty" seems to express the idea of misbehavior as well as a recalcitrant hope that what you are doing isn't all that wrong and you won't be caught, and even if you are, it won't be that bad. Participants in these partnerships seem to be finding that their hope is misguided. When the naughtiness is uncovered, it may mean more than having to stand in an unpleasant corner for a long time. Certain behavior violates the law and it creates significant challenges for the administration of the income tax system.

The Internal Revenue Service (IRS) takes great care in forecasting the number of tax forms that come in every year. Among other uses, the IRS uses these forecasts to estimate and allocate processing and examination resources. Processing resources include the number of people that it takes to remove returns from millions of envelopes and transcribe information on those returns, the computers that are needed to tabulate the amounts and maintain the relevant account information, leases for buildings so people and machines will have a place to do their work, to name just a few of the major processing component requirements. It is also necessary to verify information on the returns to help ensure compliance with tax laws including some examinations by IRS auditors and other costly professionals. The number of highly skilled professionals required for examination depends in part upon the number of returns coming in to the IRS by type of form. So forecasts of return filing volumes are a major first step in many IRS resource planning processes.

Attempting to forecast the partnership series can be "messy" at times, having turning points that are not easily foreseen, if at all predictable. In the case of partnerships, the volume of Forms 1065 has been affected by changes in legislative rules by which partnerships are administered, and then taxed. The change in behavior by most taxpayers may or may not be clearly understood by the forecaster. There may also be illegal behavior; i.e. abusive tax shelters that are created, that affect the number of partnership tax returns filed in a given year. This kind of behavior, if significant, is also difficult to forecast. Yet if the resultant forecasts are poor, then the estimation of resources required for tax administration will be misaligned, which can adversely affect IRS processing or examination programs.

Why Form a Partnership

Before we examine how to forecast the number of partnerships, it will help to discuss the environment of business organization. Why would someone want to be partner in a partnership? There are other ways to organize a

business. Broadly speaking, the list of possible business organizations includes sole proprietorships, partnerships, and corporations. In general, businesses organize themselves based on pre-determined business situations and the preferences of the people running the business. More specifically, owners of a business have tax and non-tax consequences to consider. Tax considerations are fairly straightforward. In general, a sole proprietor will pay taxes on any profits he makes on his business each year on his Form 1040 return, and he will pay taxes based on his marginal rate under individual income tax law.

At the other end of the complexity spectrum and again speaking in general terms, a corporation will first pay taxes on the amount of profits the company makes and then, once dividends are paid out to individuals, the individual will pay taxes on those dividends as reported on their Form 1040 return. This is commonly referred to as the double taxation of corporate profits. First the corporation pays and then the individual pays. So far, the sole proprietorship sounds far better. After all, you only have to pay taxes once for the profits that are made.

However there are non-tax considerations to be examined. First, it is unlikely that a sole proprietorship could amass the capital needed for anything but a small business. Second, there are liability issues. The corporation can help in both of these concerns. The sole proprietor is responsible for all debts of his business. If he cheats or harms someone, he will not only lose his business; his creditors can also pursue him for his personal wealth. In a corporation, the shareholders are not liable for anything beyond the value of their shares. And, very importantly, to amass capital, a corporation can sell more shares. But, as stated, the shareholders must submit to the double taxation of their profits. Now we'll turn to partnerships, where there are some benefits of both the sole proprietorship and the corporation to be found.

Briefly, a partnership is a business entity where two or more individuals join to carry on a business that is not a corporation. It is more complicated than that, but we can use that definition. So far so good, but one more exiting fact about a partnership - a partnership cannot be taxed! So far, they sound like such a great idea that you may wish to run out right now and form a partnership. However, while the partnership does not pay any tax directly, any income that the partnership makes "flows through" to the individuals that make up the partnership. Each individual pays a tax using their individual tax rate on their 1040 Form much as they do in the sole proprietorship situation. But unlike the sole proprietorship, a partner in a partnership is not necessarily liable for anything beyond the money he or she invested into the partnership.

Lets assume that a group of people decides to start a partnership. The partnership, as it is not taxed, must allocate gains (and/or losses) to the individual members of the partnership and these individuals (partners) must pay tax on the gains, even if the gains are retained by the partnership for growth purposes. On the other hand, a corporation may retain some of their earnings, but the shareholders do not pay tax directly on those earnings. A shareholder in the corporation will be liable for any capital gains if they sell their shares in the now (presumably) more valuable corporation.

There are several kinds of partnerships. They are: general partnerships, limited liability partnerships, limited partnerships, and limited liability companies. Do not let the appellation "companies" on limited liability companies confuse. The limited liability company is indeed treated as a partnership for tax purposes by the IRS.

A general partnership means, most importantly, that the assets of the partnership, as well as the assets of the individuals in the partnership are fair game to creditors for the debts of the partnership. If one of the general partners is a poor manager or fritters away the assets of the general partnership by some means, the other general partners are obligated by the debt. If you are a general partner in a partnership, you are taking considerable responsibility for the actions of others. A limited liability partnership, or LLP, is often used by a legal, medical, or accounting practice. In this organization, malpractice by one partner will not obligate the other partners, but in other aspects it is similar to a general partnership. A limited partnership can be useful for obtaining working capital. In a limited partnership, there are one or more general partners and often, many more limited partners. Limited real estate partnerships are very common. In the last few years, another form of business organization has become popular. A limited liability company, or LLC, has the

benefit of limited liability for the partners, as well as avoiding the double taxation of the corporate income tax. The LLC is now recognized by all of the states and the District of Columbia as a business organization whose members are not liable for the poor decisions of other members, save for the amount of assets they bring to the table.

The Time Series Data on Partnership Return Filings

Now that we have some background on the reasons for partnership formation, we can address how to forecast the number of partnership forms that come into the IRS each year. Let's turn to an examination of the partnership series in Chart 1. Note the general rise in the volume of forms until 1987. This is the year in which The Tax Reform Act of 1986 (TRA86) directly affected the taxation of partnerships through the changes in passive loss rules. After 1987, the number of partnerships gradually dropped until 1995, when they began to rise again.

In treating passive losses differently from active losses, Congress disallowed many tax shelters with the passage of TRA86. So what are passive losses? Lets use an example to illustrate. Take a hypothetical situation where a group of owners formed a partnership to start a hissing beetle ranch. (Some people enjoy keeping hissing beetles as pets.) Now assume that ten new partners of the partnership invested $10,000 each into the hissing beetle ranch. There were, of course, quite a few losses in the early days of the business, as there usually is with most new businesses. The partnership borrowed heavily to purchase the land upon which the hissing beetles could roam and procreate. Losses to the partnership mounted, including depreciation, interest, and other miscellaneous costs. At the end of the year, the partnership discovered that they had $200,000 in losses. Remember that the losses flow through to the individual partners. Each partner now had $20,000 in losses ($200,000 divided by 10) to report to the IRS. We can assume that the partners borrowed heavily to purchase the land for the hissing beetle farm. (The land itself was probably used as security for the loan so that the partners themselves were not liable for any non-payment of loans.) So, now each partner could deduct $20,000 from salary, interest and dividends that they obtained from sources other than the hissing beetle farm. In this way, each partner, with an investment of $10,000 was able to deduct $20,000 of losses. It can be said that the partners "sheltered" their income with the hissing beetle farm.

Faced with situations analogous to this illustration, Congress responded with passive activity loss rules in the 1986 Act. In the previous example, the partners in the hissing beetle concern did not actually work down on the farm. They were passive owners. The new rules allow only passive losses to be deducted from passive income. Our partners in the hissing beetle farm can not now deduct their losses against salary, interest, or dividends for other income sources as they once could.

To return to Chart 1, again, examine the behavior of the series after 1987 up until 1995. After the implementation of the passive loss rules, the number of partnerships filing each year dropped more or less consistently for the next several years, until 1995. In that year the number of partnerships began to rise again. Refer to Table 1. (This table was compiled by the IRS's Statistics of Income (SOI) division. This data will not correspond exactly to the data in Chart 1 because SOI data are based on samples and cover a somewhat different yearly position in time. However, SOI data contains more detailed information in partnership.) Partnerships in this table are divided into 3 groupings, active (total) partnerships, limited partnerships, and limited liability companies. The number of limited partnerships and of limited liability companies are probably understated here because some businesses failed to answer the question on their tax form about type of partnership. One *could* obtain an estimate of the number of general and limited liability partnerships by subtraction, but it is not necessary for our discussion. Of all the categories of partnerships, only the limited liability company partnerships are experiencing significant growth. While data is not available for all years and the last year available is 2000, it is evident that the rise in volume has been rapid. It also appears that growth has slowed down to about 22% for 2000, the most recent year available. Further, it appears that almost the entire rise in the number of partnerships as a whole since 1995 is due to the rise of the LLCs.

The Rise in Limited Liability Companies

The rise in partnerships after 1995 seems to be consistent with the rise of the limited liability companies and their treatment as partnerships by the IRS. An understanding of recent tax law will make this clear. In 1988 the IRS agreed to treat LLCs as partnerships for tax purposes. As has been stated, this allows the business to limit the liability of the partners and avoids the problem of double taxation. All in all this is quite appealing to the owners of a new business.

Further, for the 1997-tax year, the IRS implemented "check the box regulations". This allows a group to choose their tax status without regard to other characteristics of the organization. Here, any group with more than one member that qualifies can choose to be treated as a partnership or a corporation for federal tax purposes. If no election is made, then the organization is treated as a partnership. However, it is unclear whether this would significantly raise the number of partnerships if the "check the box" option were not available. The choice to be treated as a partnership or corporation requires a detailed knowledge of the tax laws and what is most beneficial to the owners of the business. Even with the default choice as a partnership, it is unlikely that the "check the box regulation" will materially affect the number of partnerships that are formed.

Evidence here suggests that prior to 1986 many partnerships were being used to shelter income. The drop in partnership volume from 1987 to 1994 (after the passive loss rules were implemented) points strongly to this interpretation. But that decline did bottom out in 1994 and the volume has since continued to climb up to the current (2003) estimates. The recent economic recession and current sluggish recovery have not seemed to dampen the growth in Form 1065 filings, so while it is unclear, some partnerships may still be used to shelter income.

Current Use of Partnerships to Shelter Income

How this is being done may be understood by a few more examples. First, with a single partnership, it is possible to misreport income or deductions to income to evade individual income taxes. It seems likely that the possibility to generate these accounting untruths increases with many interwoven partnerships. Since it is possible to create a business entity consisting of many partnerships. In this situation, a partnership does not cause income/loss to "flow through" to an individual. The partnership instead distributes the income/loss to another partnership! And that partnership may distribute its income/loss to yet another partnership. And so on and so on, until the income/loss drop down to an individual. This is known as "tiering". It must be clearly stated that there may be justifiable reasons for this type of partnership. For example, it may be necessary to segment the operations of a business that are the responsibility of many different people at different points in time. Just because a business has a tiered business structure does not mean they are involved in any untoward behavior. However, this sort of structure can create a complicated web to mask inappropriate shelter activity in some instances.

Here's how. The income from a partnership is the income that remains after all expenses are accounted for in the partnership. For every partnership, there are costs that will decrease the amount of profit (income) that the partnership makes. Now, let's assume in a tiered partnership structure, we have layers of cost layered on layers of cost. By the time we get down to the lower tier of the partnership structure and the gain (or most probably, loss) falls through to the individual, there is little or no income to declare. In this way the individual(s) who created the tiered structure have succeeded in their task. Either they can declare large amounts of loss to write off gains from other sources of income, or they can declare a small amount of having it reduced by the mostly frivolous costs that were incurred in the many layers of the tiered partnership structure. In the latter case, they have hidden income upon which they do not pay tax. In the former case they are "sheltering" income that should properly be reported.

Such sheltering of income may not be uncommon. In fact, one of the largest failures in United States business history was tied to questionable partnership accounting. In that case, they did not use one or two partnerships. They used, at last estimate, almost 900 partnerships. Partnerships were used to hide debt from the real balance sheets of the firm to make the otherwise debt-laden firm appear to be doing quite well.

We've previously examined how it is possible to hide income and not pay tax on the hidden income. In the 900-partnership firm, a publicly traded company, the idea was to take debt "off the books" so that the firm appeared to be doing well. In essence, they "hid debt" instead of hiding income. When debt is reported as a part of the disclosure process of public companies, the company with less debt is seen as more valuable. Their stock can go up in value. As a result of this accounting, shareholders were lulled into a sense of well being as they were steered toward insolvency.

More: just as a musician can create a tune with only one instrument, so can a less-than-truthful individual create an abusive tax shelter with only one partnership. But, as it takes several musicians to create chamber music, it may take partnerships, trusts, S corporations and regular corporations to create a tax shelter. With these fundamental business entities, the dishonest individual can manipulate gains (interest, dividends, and royalties) and other accounting quantities (ordinary income, business income, rental income, rental activity income, and passive income) to create a structure that allows them to pay less tax than their fair share.

It is unclear how many of these multientity businesses are participating in wrongful tax shelters. As a result, it is difficult to quantify any untoward behavior in forecasting the number of partnerships that come in each year. For the last several years, (or at least since 1995), the formation of partnerships have been reasonably well behaved in a forecasting context. It appears that the limited liability companies are enjoying a rise due to their limited liability as well as due to the avoidance of double taxation on corporations. They may also be rising due to an increase in abusive tax shelters, or they may not. The jury is still out. But there does seem to be enough anecdotal evidence to justify further research.

Forecasting the Partnership Series

From the previous discussion, it appears that the partnership forms volume is sensitive to changes in legislation as well as IRS administrative rules. The 1986 Tax Act and the "check the box" regulations in the 1990s significantly altered the numbers of partnerships formed in the U.S. One could, in theory, forecast the partnership series using dummy variables for different regimes, say for the long rise in partnership volume from 1959 to 1986, another regime from 1987 to 1995 to reflect the decline due to the restrictions of passive loss due to TRA86. Finally, a final regime from 1996 until the current year could also be used reflecting the rise of limited liability companies. A proxy for general business activity should also be considered. One would expect that better economic times would produce more partnerships, and a resultant increase in the series.

While theoretically correct, such a model will not work as a forms volume forecast. The partnership forms volume is by nature a continuous series, *when legislative and economic factors are held constant*. By continuous, it is assumed that the first forecasted value must be reasonable in context with the last value in the series. Planners, or other users of the forecast who see a discontinuous value from one time period to the next will properly question the forecast and the methodology. An acceptable forecast begins with two simplifying assumptions. First, the legislative environment must be constant, i. e. no change will occur in the near future that will affect the forms volume. Second, the current economic environment will continue in the near future. This may be the most uncomfortable assumption that is made in forecasting the partnership series. Obviously, the economic environment is fluid and change does occur. (The economy is currently experiencing a sluggishness in activity due to a variety of factors.) However, as we have seen in the history of the partnership series, it is unclear how the levels of economic activity affect the series. The fact that the volume fell in the late 1980s was most probably due to the Tax Reform Act of 1986 and not necessarily the recession that occurred shortly thereafter. Also, the recent slowdown that began in 2000 did not reduce the number of partnerships filed. One could make the argument that except for the increase of LLCs, the number of partnerships would have declined or stagnated, but reliable data is not available to test the proposition.

Therefore, in forecasting the partnership series, the assumptions are that 1) a constant legislative regime has existed from 1996 until 2002 (the last year for which data is available), and 2) the economic climate has little influence on the volume of partnership tax returns filed. Perhaps this last assumption is a bit much to swallow.

However in terms of the mechanics of partnership formation, this may not be an unreasonable assumption. Due to the limited dataset, there will be no holdout data. The source for data used to forecast partnerships is obtained from the IRS business master file. The master file is a tax administration computer file that contains information encoded from virtually all received tax forms.

It is common in forecasting IRS forms volume that a limited number of observations are available. Legislative environments, economic conditions, tax administration considerations, data availability, and data quality all must be considered when a model selection is made. To forecast Form 1065 volume, Box-Jenkins, or autoregressive integrated moving average (ARIMA) models have some appealing features. They are widely used, easily understood, and software exists to make inexpensive, rapid computations.
These models generate forecasts that are easily computed, and easy to explain to the users of the forecast.

To forecast partnerships, an ARIMA model was developed. The process is not complicated. The partnership series obviously needs to be detrended, so single differencing was used. An examination of the ACF and the PACF show that ARIMA models from ARIMA(1,1,0) to ARIMA(5,1,5) may be promising. Therefore, a grid of ARIMA(p, d, q) choices were run for p=0 , q=0 to p=5, q=5. The Schwartz Baysian Information Criterion (SBIC) was used to rank these generated models. An ARIMA(1,1,0) model had the lowest SBIC. The model produces reasonable forecasts, but the trend may be too damped. On reflection, this is acceptable due to the slowing of the LLC component of the partnership series as a whole. After generating forecasts, the residual errors show a slight tendency to overpredict in the most recent time periods, but it is not unacceptable.

Below are the estimated parameters of the chosen model.

Autoregressive factor: $1 - .8342\ (W_{t-1})$ with a t-value of 11.2.

Refer to Chart 1 for the forecasts. Refer to Table 2 for the numeric forecasts. The forecasts call for a continued rise in the number of partnership returns filed over the forecast period. The projected growth for the immediate years ahead is in the 3 percent range, then gradually slowing to below 2 percent by 2010.

Conclusion

This paper has explored the rationale for the use of partnerships in organizing business activity as well as why the number of these partnerships hold interest for the IRS. The use of partnerships has a history several decades old. Their use is often correct and appropriate, given the current laws of the land. Over time, the legislative and administrative environment can cause individual economic actors to find them more or less useful in organizing the way business is structured. The use of partnerships for untoward purposes, on the other hand, has a colorful anecdotal history. It is clear that some partnerships have been used to illegally shelter income.

This article also presented a method to forecast the total number of partnership returns received each year at the IRS. This method made use of an ARIMA(1,1,0) model, which generated reasonable forecasts given the somewhat limited historical data. The forecast might be improved with additional data. Form count volumes are not yet available for each category of the partnerships (general partnerships, limited liability partnerships, limited partnerships, and limited liability companies). However, forecasting the counts of each of these four partnership types and then summing each of the forecasts to a total partnership count will capture the differing change in each sub-series. Thus, forecasting by parts may provide a better method of projecting Form 1065 volumes in the future.

Note:

The views expressed in this article represent the opinions and conclusions of the author. They do not necessarily represent the opinions of the Internal Revenue Service.

References

Eugene Willis (Editor) et al, West Federal Taxation 2003, 26th edition: Comprehensive Volume,
Howard W Sams & Co; (2002).

Makridakis, et al, Forecasting: Methods and Applications, Third Edition, John Wiley and Sons, Inc. New York, NY, 1998.

Daily Telegraph, "The Whistle Was Blown, But Enron Never heard It.", WWW.TELEGRAPH.CO.UK/MONEY, Filed Jan 19, 2002.

Buckel Julie, "Flow-throughs and Tiering; Using Schedule K-1 Data to Study Tax Compliance", The IRS Research Bulletin, Proceedings of the 2003 IRS Research Conference, IRS, Publication 1500

Table 1.--Partnership Returns: Income Years, 1980-2000

[All figures are estimates based on samples]

Item	1980	1985	1990	1995	1998	1999	2000
Total number of active partnerships	1,379,654	1,713,603	1,553,529	1,580,900	1,855,348	1,936,919	2,057,500
Number of limited partnerships	170,336	279,878	285,769	295,304	342,726	354,295	348,973
Number of limited liability companies	n.a.	n.a.	n.a.	118,559	470,657	589,403	718,704
Pct change in limited liability companies					297%	25%	22%

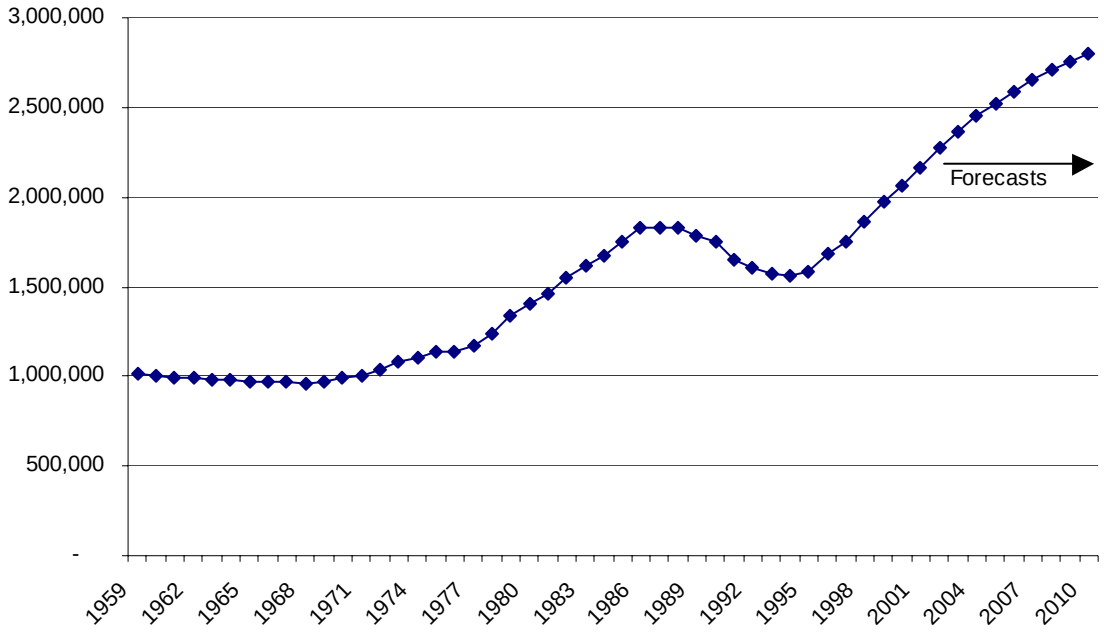
Source: IRS Statistics of Income Division

Table 2. Form 1065 (Partnership) Volumes: Historical and Projected

	F1065 Volume		Year	F1065 Volume		Year	F1065 Volume
Historical:	1,010,998	**Historical:**	**1980**	1,401,639	**Historical:**	**2001**	2,165,011
	1,006,107		**1981**	1,457,974		**2002**	2,271,755
	994,730		**1982**	1,552,735	**Projected:**	**2003**	2,365,661
	992,920		**1983**	1,613,493		**2004**	2,448,860
	983,307		**1984**	1,675,605		**2005**	2,523,124
	979,224		**1985**	1,755,339		**2006**	2,589,937
	972,748		**1986**	1,831,600		**2007**	2,650,534
	968,253		**1987**	1,824,166		**2008**	2,705,945
	969,027		**1988**	1,825,865		**2009**	2,757,029
	959,344		**1989**	1,779,617		**2010**	2,804,505
	966,498		**1990**	1,750,921			
	991,904		**1991**	1,652,276			
	1,005,365		**1992**	1,608,727			
	1,035,986		**1993**	1,567,150			
	1,076,869		**1994**	1,558,404			
	1,106,429		**1995**	1,580,292			
	1,132,839		**1996**	1,678,786			
	1,138,770		**1997**	1,755,403			
	1,165,904		**1998**	1,861,009			
	1,235,388		**1999**	1,974,667			
	1,341,863		**2000**	2,066,796			

Source: IRS Document 6186 and Author's Forecasts

Chart 1. Partnership (Form 1065) Return Volumes Through 2010 (Actuals 1959 - 2002)



FORECASTING EMPLOYMENT LEVELS OF IRS COMPLIANCE STAFF WITH A MICRO MODEL OF EXITS AND JOB CHANGES

Thomas Mielke, Internal Revenue Service
Alex Turk, Internal Revenue Service

Introduction

The Federal Government's workforce is rapidly aging (Government Accounting Office (GAO), 2001). Within the Internal Revenue Service (IRS) this trend is even more pronounced for two of the IRS's mission critical jobs, Revenue Agents (RA) and Revenue Officers (RO).

Revenue Agents, Revenue Officers and Tax Compliance Officers (TCOs) make up a large proportion of the IRS compliance workforce. Revenue Officers generally work with taxpayers that are delinquent in paying their tax liability. Revenue Agents and TCOs conduct audits of previously filed tax returns to determine if tax liability was correctly reported. RA and TCO positions, while similar, differ in the complexity of work assigned to them. TCOs were examined in our original study, but they will not be discussed in this paper. During 1997 to 2002, the TCO position endured a significant amount of change and currently the future of the position is in doubt.

Many Revenue Agents and Revenue Officers are near retirement age. In just under five years, 37% of the full-time RAs and ROs will be eligible for retirement. An additional 25% will be eligible for early retirement. With 62% of the employees eligible for full or early retirement, the IRS will be (or is) facing a human capital crisis in the near future.

In this paper, we develop a micro model of attrition for both IRS Revenue Agents and IRS Revenue Officers. We use this model to develop forecasts of the number of RAs and ROs that change jobs or leave the IRS under two different scenarios. The first scenario assumes no new employees are hired. The second scenario assumes hiring levels of RAs and ROs that maintain a constant staffing level.

Background

A significant amount of research has focused on employee turnover[1]. Previous research has explored the relationship of wages, human capital, and demographics to the length of employee tenure in a job or organization. The model developed in this paper is consistent with the body of previous research but does not add significantly to the understanding of worker tenure decisions. Instead, it focuses on using the model of individual tenure decisions to provide aggregate attrition forecasts of the IRS compliance workforce. Developing the forecasts in this manner provides the ability to predict attrition under almost any hiring plan.

Model and Forecast Methodology

Empirical Model

Assume that workers choose at time t to remain employed in their current job, change jobs internally or leave the IRS altogether. For the model used here, we do not distinguish between internal job transfers and leaving the service. Thus, we assume that employees compare the net benefit between the two employment opportunities based on a set of exogenous factors x_{it-1} and a stochastic shock ε_{it}. Let $e_t=0$ represent the employee choice of remaining in their current job at time t and let $e_t=1$ represent exiting their current job for employment elsewhere. An individual will choose to leave their current job if

$$E_t^* = U(e_t=1, x_{it-1}, \varepsilon_{it}) - U(e_t=0, x_{it-1}, \varepsilon_{it}) > 0.$$

Unfortunately, the value of E_t* is not revealed to us. Only the sign of E_t* is revealed by observing if the individual retains their job at time t. Assume that the net benefit from changing jobs can be represented as

$$E_t^* = x_{it-1}\alpha + \varepsilon_{it}\,.$$

Assuming that ε_{it} is distributed normally, the decision to exit their current job can then be represented as

$$P(E_t^* > 0) = \int_{-\infty}^{x_{t-1}\alpha} \phi(z)dz = \Phi(x_{t-1}\alpha)\,,$$

where ϕ is the normal density function and Φ is the normal distribution function.

The standard Probit model discussed above generates a probability that a given worker will leave their current job within the next year conditional on being in the job in the current year. We use the one-year transition probabilities to generate aggregate predictions of attrition over the next five years in both RA and RO occupations.

Forecast Methodology

The current year forecast of attrition rates is derived by aggregating the predicted probabilities of each employee leaving before time *t*, denoted as $P_{it.}$ For t = 2003, expected attrition is

$A_t = \sum_{\forall i} P_{it}$ for all employees in the respective job at time t-1.

2003 expected attrition is based on the observed characteristics of the employees in 2002. However, to predict attrition between 2003 and 2004, we need to know the characteristics of the employees that will be in the labor pool in 2003. To accomplish this, we "aged" the current employees and recomputed all the variables derived from age and tenure. The expected number of employees exiting at time t+1 is then

$A_{t+1} = \sum_{\forall i} P_{it+1} = \sum_{\forall i} (1 - P_{it}) P_{it+1}$ for all employees in the respective job at time t-1. At time t+2 the forecasted attrition is

$$A_{t+2} = \sum_{\forall i} P_{it+2} = \sum_{\forall i} (1 - P_{it})(1 - P_{it+1}) P_{it+2}.$$

In general, the *K* period ahead forecast of attrition can be expressed as

$$A_{t+K} = \sum_{\forall i} \left(\left(\prod_{k=0}^{K-1} (1 - P_{it+k}) \right) P_{it+K} \right).$$

Attrition forecasts are generated for two different scenarios. In the first, no additional employees are hired to replace those who leave. Thus, the forecast formula above is applied to the existing employees in 2002.

The second scenario consists of hiring sufficient numbers to maintain the number of employees in a given occupation at the 2002 level. To account for new employees entering the IRS labor force, we identified all new hires during the sample period. We use these individuals as a pseudo pool of potential applicants in the subsequent years. We then randomly "clone" individuals out of this pool to be the new hires in each forecast year. In this scenario, the forecast formulas are applied to the existing workforce and the "clones" that represent the new hires. One problem with this scenario is that the RA and RO occupations have had only limited hiring during the sample period. However, most of the hiring occurred in the more recent years. Thus, we feel that the past hires should be very similar to the qualified applicants that would be in future applicant pools.

Data

Our data comes from IRS payroll data. We obtained annual data from the 20th bi-weekly pay periods of each calendar year during 1997-2002. The payroll data contained an abundance of employment information. During this period, the IRS underwent a substantial reorganization that resulted in many RAs and ROs changing jobs. Some RAs and ROs left the IRS and then were rehired a few years later (2001 and 2002).

In each year of our data, the total number of RAs and ROs has declined. As Table 1 shows, from 1997 to 2002, the total number of RAs has declined by 15.4% and ROs by 21.1%. Without significant hiring, this downward trend is not likely to end. As Figure 1 depicts, 17% of all RAs and ROs will be retirement eligible by the end of 2003, and another 41.5% will become eligible within the next 10 years. In 10 years, 58.5% of all currently employed RAs and ROs will have retired or be eligible for retirement. In addition, there is a large cohort of employees (33% of all ROs and RAs) that has 14 to 16 years of tenure. For the most part, these employees will be eligible within the next 15 years.

Table 1. Employee Changes During the Year, 1997-2002

Job	Year	Total Number Employed	Attrition	New Hires	Hires within the IRS
Revenue Agent	1997	15,028	910	19	86
	1998	14,223	679	35	129
	1999	13,708	688	24	145
	2000*	13,189	1,123	460	223
	2001	12,730	712	532	162
	2002	12,712	NA	NA	NA
Revenue Officer	1997	7,454	432	6	40
	1998	7,068	428	6	72
	1999	6,718	414	6	50
	2000*	6,360	636	240	305
	2001	6,269	450	3	56
	2002	5,878	NA	NA	NA

* Primary IRS Reorganization Year

Since the early 1990s, years of tight budget conditions have limited IRS hiring. This has resulted in a void of workers at the lower end of the tenure distribution

(Figure 2). In Fiscal Year (FY) 2001 and 2002, the IRS

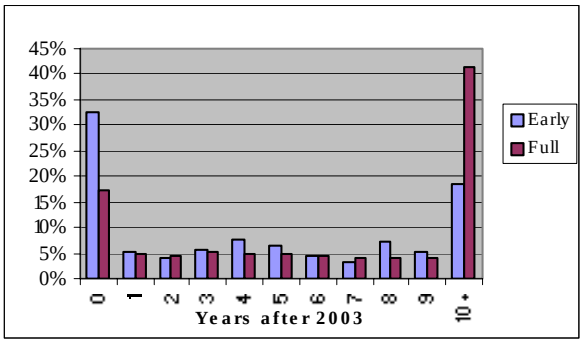


Figure 1-Distribution of the RA/RO Workforce by the Number of Years to Retirement Eligibility

hired 992 RAs and 243 ROs (Table 1, calendar years 2000 and 2001). However, the average age of these hires was 38 (Small Business and Self-Employed (SB/SE) Internal Scan) and some hires were individuals who left the service and subsequently returned. The average age suggests the IRS is not hiring recent college graduates but rather employees with significant

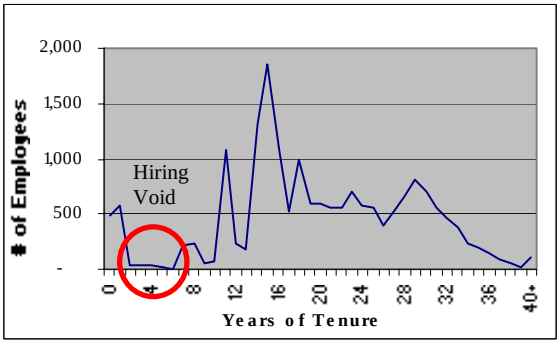


Figure 2- IRS Tenure of RAs and ROs, 2002

labor market experience. A potential disadvantage of this practice is some new hires could be underemployed and may be primed for separation when the economy improves. For FY 2004, the IRS plans to hire 950 new RAs and ROs (IRS Office of Strategic Human Resources, 2003).

External hiring is a challenge because the skills required for Revenue Agents are among the most competitive in the external job market (Hall, 2003). The heightened competition makes it difficult for the IRS to recruit Revenue Agents (SB/SE Internal Scan). In 1997-1998, 41,170 bachelor's degrees and 6,725 Master's Degrees in accounting were awarded and this was a continuation of a downward trend. Over 28,750 accepted jobs in the private sector, and another 5,750 became self-employed. Thus, the public sector faces intense competition for employees with an accounting background (IRS Office of Strategic Human Resources, 2003). Applicants for Revenue Officer positions can come from various academic backgrounds, so the pool of potential applicants is much larger.

Excluding retirement eligible years, employee turnover in private and public sector jobs is the highest in the first years of tenure (new employees). The IRS experience has been no different. Figure 3 displays the exit rates for RAs and ROs from 1997 to 2001. In addition to the retention problems of new hires, IRS Strategic Human Resources has identified factors that are expected to complicate the retaining and replacing of experienced employees. The retention and replacement of employees will be affected by 1) a portable retirement system, 2) a growing pay gap between the public and private sectors, 3) high external competition for candidates and 4) an emerging pattern of frequent job changes during an employee's life span. The effects of these factors will likely be in remission until private sector jobs become plentiful again.

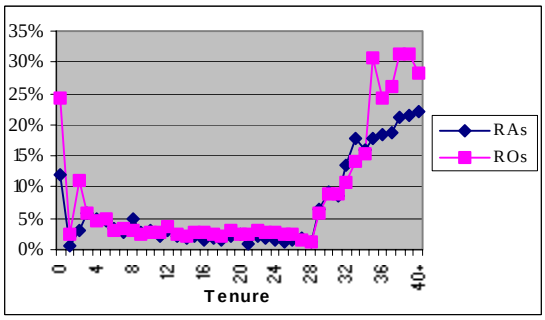


Figure 3 – RA and RO Quit Rates by Tenure, 1997-2001

The majority of federal employees are under one of two retirement systems, the Civil Service Retirement System (CSRS) and the Federal Employees Retirement System (FERS). CSRS is a traditional pension plan and FERS is comparable to a 401K plan where employee and employer contribute. Since 1987, every new federal employee is covered in the FERS retirement system. For the most part, employees hired before 1987 are covered under the CSRS program. CSRS employees that leave before retirement eligibility stand to lose a sizable amount of their retirement savings. Given a more portable retirement system, FERS employees incur lower exit costs and thus may be more inclined to exit federal service when the opportunity arises. The difference in retirement plans is utilized in

creating certain retirement variables in our model. We include dummy variables for FERS employees who are eligible for early retirement, CSRS employees who are retirement eligible and at the top of their pay scale, and finally any FERS employees who have reached the top of their pay scale. In addition, we use many variables that change with age and tenure. These include retirement eligibility, being retirement eligible for the 3rd year, becoming eligible for early retirement, and also having low tenure. Each of these variables has been "aged" when we develop forecasts.

Figure 4 displays the observed exit rates by the number of years since reaching retirement eligibility. The exit rate for retirement eligible employees is relatively more stable for the RA position. Exit rates for the RO position show more variation overall peaking after 5 years of eligibility. The exit rate appears to decline for both RA and RO employees who have been eligible for 3 years. We attempt to control for these differences in the model.

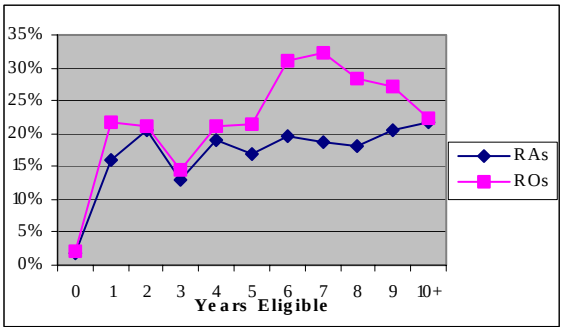


Figure 4- Exit Rates by Years of Being Retirement Eligible 1998-2001

Using the yearly changes in an employee's sick and annual leave balances enables us to control for an individuals use of leave. We control for those workers who begin hoarding annual leave (accruing annual leave in a year without using any of it) as many employees who are planning to retire exhibit this behavior. A second dummy variable was included for those workers who had been advanced a substantial amount of sick leave (more than 200 hours). This variable could indicate a serious medical illness-perhaps leading to disability retirement. In addition, we included a dummy variable for those who tend to use a substantial amount of sick leave (more than 96% of accrued sick leave).

Many factors that may affect an employee's decision to leave their job are not measured with the available data. For example, we do not have an indicator of financial standing. Wealth makes retirement more feasible, and may make workers more mobile. Generally, wealth tends to increase with age and pay, so parameter estimates associated with these variables may also include a wealth effect. Another factor is the number of dependents. Having dependents may make retirement less financially feasible, and makes workers less mobile. We include a dummy variable for family heath care coverage as a proxy of family status. Unemployment rates by region could also have an effect on turnover. If unemployment is low, obtaining another job is not as difficult, so turnover should increase (and vice versa). We considered including a regional unemployment measure in the model but felt that we needed a longer sample period to obtain a defensible estimate of the effect of local labor market conditions. The IRS's reorganization would further confound our ability to measure local labor market conditions. Instead, we used annual and regional dummy variables to control for these effects.

An issue with using a micro model to develop forecasts is that it is not known how many of the individual factors may change in the future. For example, we don't know how characteristics like the hoarding of annual leave, sick leave balances, and an employee's performance evaluation may change over time. For each of these factors we used the 2002 values for the forecasted years.

Model Estimates

The probit model parameter estimates for the RA and RO models are reported in the Appendix. For the most part, these estimates are consistent with previous research. In addition, the results of the RA model are similar to the estimates for the RO model.

An interesting finding is that the overall retirement plan dummy variable (FERS) was negative and insignificant. This suggests that there is no difference in quit rates between FERS and CSRS employees who are not retirement eligible. However, the model does indicate that not using annual leave is a good indicator that employees are going to quit. Workers who receive poor performance evaluations and are not receiving awards for performance are also more likely to quit.

Forecast Scenario 1- *Attrition with No Hiring*
We first examined the extreme case where no new employees are hired. Both the RA and RO forecasts reported in Table 2 suggest an increase in the attrition rate over time. However, the number of employees leaving each year is actually declining because we assume there is no hiring and therefore the labor force

is shrinking. Between 2002 and 2003, the estimated attrition rate for RAs is 6.4% and for ROs, 7.7%. Our estimates modestly decline for 2003/2004 but then increase until the 2006/2007 year when our estimated attrition figures are 6.6% for RAs and 8.3% for ROs. If this occurs, we expect to see the number of RAs decline by 22% to 9,810 employees and the number of ROs decline by 27.5% to 4,258 employees. The forecasted attrition rates over the next five years are higher than the observed rate has been over the past years, with the exception of 2000.

Table 2. Attrition Estimates with No New Hires

Year	Revenue Agents		Revenue Officers	
	Count	Attrition	Count	Attrition
1997	15,028	6.06%	7,454	5.80%
1998	14,223	4.77%	7,068	6.06%
1999	13,708	5.02%	6,718	6.16%
2000*	13,189	8.51%	6,360	10.00%
2001	12,730	5.59%	6,269	7.18%
		Est.		Est.
2002	12,712	6.37%	5,878	7.74%
2003	11,902	6.17%	5,423	7.57%
2004	11,168	6.16%	5,013	7.71%
2005	10,480	6.40%	4,626	7.96%
2006	9,810	6.58%	4,258	8.34%

* Primary IRS Reorganization Year

Forecast Scenario 2- *Maintaining the Status Quo*
In this scenario, every employee who leaves is back-filled with a new hire from the external labor market. Thus, we keep the number of ROs and RAs at the 2002 levels. Those workers that were hired externally between 1997 and 2002 are used to proxy the pool of potential applicants. We randomly selected, with replacement, employees to back-fill.

Table 3 displays our forecast results for both RAs and ROs. Both RA and RO attrition is forecasted to initially rise as the new hires are introduced and then eventually decline. Recall that both models include dummy variables for employees with less than two years of tenure. In the first few years of the simulated hiring, new employees account for a larger percentage of the workforce than they do in later years. As the new employees age beyond the initial two years, attrition rates start to fall.

While both models include dummy variables for low tenured workers, the magnitudes of the increase are different. Both are positively related with quits, but the RA estimate is very small and not statistically

Table 3. Attrition Estimates with Hiring to Maintain a Constant Workforce

Year	Revenue Agents		Revenue Officers	
	Hires	Attrition	Hires	Attrition
Base	12,712	Est.	5,878	Est.
2002	810	6.37%	455	7.74%
2003	822	6.46%	473	8.05%
2004	815	6.41%	518	8.81%
2005	820	6.45%	516	8.77%
2006	798	6.28%	514	8.75%

significant. The RO estimate is much larger and is significant at any reasonable level. One possible explanation is that ROs come from a broader background in terms of academic and labor market experience. Prior academic and labor market experience may be a much better screening device for RAs applicants than for RO applicants. Thus, the RA hiring process may be more likely to produce applicants who are a good match with the job duties.

Comparing the Two Scenarios
Figure 5 and Figure 6 depict the difference in the estimates of the two scenarios. The RA estimates don't differ dramatically for the two scenarios in the initial years of the hiring. The RO model, depicted in Figure 6, shows a larger deviation, but it appears to be converging in the later years of the estimate. This convergence can be seen with the difference in the two estimates peaking at 1.1% in 2004, and then dropping to 0.4% in 2006. Both scenarios show that for ROs, increasing hires will increase employee turnover in the short-term. The difference is smaller for the RA hires. In both cases, the scenario with hiring to match attrition tends to moderate attrition rates in the long run. The moderation occurs because the hiring eventually re-populates the segments of the tenure distribution that have low quit rates.

Conclusions and Direction for Further Research

The model developed here not only provides a tool to forecast staffing levels, it provides some insight into the tenure decision of workers within the IRS.

First, we don't see a mass exodus once employees become eligible for retirement. Rather, only a fraction of retirement eligible RAs and ROs leave the IRS each year. In addition, Revenue Agents appear to have more

incentive to continue working. It would be interesting to explore to what degree this decision is

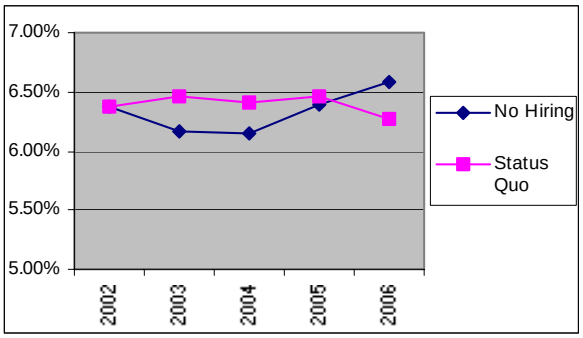


Figure 5 - Forecasted RA Attrition Rates for the "No Hiring" and "Status Quo" Scenarios

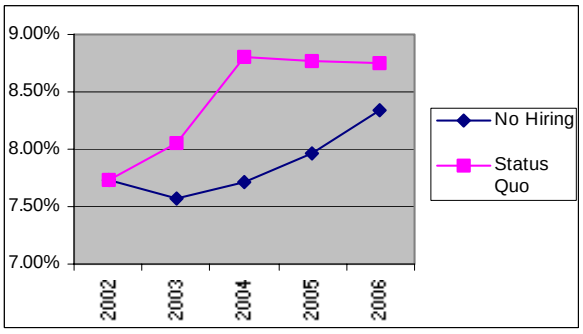


Figure 6 - Forecasted RO Attrition Rates for the "No Hiring" and "Status Quo" Scenarios

being driven by financial issues versus job satisfaction issues.

Second, the difference in retirement plans does affect tenure decisions, but only to a point. The results suggest that there is not a significant difference in attrition between CSRS and FERS employees who are not eligible for any form of retirement. One interpretation of this result is that portable retirement funds are not making IRS compliance staff more mobile. However, the model did show FERS employees are more likely to leave when they are eligible for early retirement or if they have reached the top of their salary scale. In addition, CSRS retirement eligible employees are more likely to leave when they have reached the top of their salary scale. Since CSRS pension payments are related to the highest 3 years of pay, retirement eligible employees who can receive a pay increase have more incentive to delay retirement.

Revenue Agent and Revenue Officer attrition rates are forecasted to be higher in the next five years than they have been in the past five years, excluding calendar year 2000. We forecast that by 2007, 22.8% of the current RA staff and 27.5% of the current RO staff will no longer be employed as a RA or a RO. We also found that ROs are more likely to leave their job than RAs, especially in the first years of employment. Thus, as the IRS increases hiring to replace ROs, there will be noticeable increases in attrition rates. For new hires, attrition for RAs is more evenly spread out in the initial years of employment.

This research could be expanded in several ways. Differentiating employees who make internal job changes from those who leave the service could provide forecasts that are more useful. Some variables, like performance evaluations, may have qualitatively different impacts on internal promotion than on quits. Also, including measures of wages would improve the forecast and would provide the ability to forecast attrition with various proposed pay raises. However, more data would be needed to estimate the wage effects with any degree of confidence. Additional years of data, especially with new hires, would also give more confidence about attrition in the early years of employment.

References

Ehrenberg, Ronald and Robert Smith, 200, *Modern Labor Economics*, 7th Edition, Addison-Wesley.

Government Accounting Office (GAO), *High-Risk Update (GAO-01-263)*, January 17, 2001.

Hall, Lee, 2003 *Bright Job Market for Accountants*, Atlanta Business Chronicle, June 16, 2003.

Ingram, Tonya, 2002, *The State of Human Capital in SB/SE,* SB/SE Internal Scan, IRS Small Business and Self-Employed Operating Division, pp. 30-35.

IRS Office of Strategic Human Resources, *FY 2004 Strategic Assessment of Human Capital,* March 25, 2002.

IRS Office of Strategic Human Resources, 2001, *Strategic Assessment: Our People.*

IRS Office of Strategic Human Resources, 2003 *Workforce Analysis Report and Multiyear Planning*, April.

Koch, Paul D. and James F. Ragan Jr., 1986, "Investigating The Causal Relationship Between Quits and Wages: An Exercise in Comparative Dynamics", *Economic Inquiry*, Vol. 23, January 1986, pp. 61-83.

Light, Audrey and Manuelita Ureta, 1992, "Panel Estimates of Male and Female Job Turnover Behavior: Can Female Nonquitters Be Identified?", *Journal of Labor Economics*, Vol. 10 No. 2, April 1992, pp. 156-181.

Rees, Daniel, 1991, "Grievance Procedure Strength and Teacher Quits", *Industrial and Labor Relations Review*, Vol. 45 No. 1, October 1991, pp. 31-43.

Topel, Robert, 1991, "Specific Capital, Mobility, and Wages: Wages Rise With Job Seniority," *Journal of Political Economy*, Vol. 99, No 1, pp. 145-176.

Turk, Alex H., 2000, Public Sector Unions and the Free-Rider Problem, Ph.D. Dissertation, Iowa State University.

Wolpin, Kenneth I., 1992, "The Determinants of Black-White Differences in Early Employment Careers: Search, Layoffs, Quits, and Endogenous Wage Growth", *Journal of Political Economy*, Vol. 100, No. 3, 1992, pp. 535-560.

Note

The views expressed in this article represent the opinions and conclusions of the authors. They do not necessarily represent the opinion of the Internal Revenue Service.

Endnotes

[1] For examples see Topel (1991), Koch and Ragan (1986), Light and Ureta, (1992), Rees (1991), Wolpin (1992) and Turk (2000).

Appendix:

Probit Model Estimates

Parameter	Revenue Agents			Revenue Officers		
	Estimate	Standard Error		Estimate	Standard Error	
Intercept	-0.4332	0.0905	*	-0.3989	0.1303	*
Southern US	0.0931	0.0292	*	0.0954	0.0378	*
Western US	0.0115	0.0267		-0.0654	0.0351	
North Central US	-0.0639	0.0274	*	-0.1162	0.0400	*
Eastern US	-0.1610	0.0269	*	-0.1674	0.0376	*
Washington DC Area	0.1887	0.0368	*	0.2732	0.0489	*
Bargaining Unit	-0.3669	0.0254	*	-0.4591	0.0316	*
Part-Time	0.0793	0.0501		0.2807	0.1002	*
Under 2 Years of Tenure	0.0246	0.0917		0.5024	0.1242	*
Tenure	-0.0416	0.0051	*	-0.0381	0.0085	*
Tenure Sq.	0.0008	0.0001	*	0.0008	0.0002	*
Age (minus 21)	-0.0324	0.0048	*	-0.0302	0.0075	*
Age (minus 21) Sq.	0.0004	0.0001	*	0.0004	0.0001	*
Veteran	0.0002	0.0194		0.0414	0.0241	
Perf. Eval. Outstanding	-0.0150	0.0227		-0.0350	0.0278	
Perf. Eval. Poor	0.7871	0.0698	*	0.7954	0.0712	*
Bachelor's Degree	-0.1388	0.0226	*	-0.0076	0.0244	
Advanced Degree	-0.0691	0.0350	*	0.0321	0.0471	
Year 1997	0.1453	0.0300	*	-0.0044	0.0424	
Year 1998	-0.0412	0.0282		-0.0384	0.0375	
Year 1999	-0.0167	0.0280		-0.0211	0.0374	
Year 2000	0.2836	0.0261	*	0.2614	0.0352	*
Received No Cash Award	0.0811	0.0233	*	0.2234	0.0335	*
Manager (Eligible)	-0.2091	0.0708	*	-0.2038	0.1221	
Health Plan (Enrolled)	-0.0907	0.0245	*	-0.1171	0.0313	*
Medical Disability	0.0849	0.0484		0.1221	0.0578	*
FERS Early Eligible	0.2064	0.0496	*	0.1193	0.0675	
FERS Step 10 (not Eligible)	0.2226	0.0651	*	0.1718	0.0877	*
CSRS at Step 10 (Eligible)	0.1082	0.0378	*	-0.0467	0.0591	
Hoarding Annual Leave	1.4929	0.0759	*	1.2665	0.1116	*
Sick Leave User	0.1525	0.0244	*	0.1616	0.0317	*
Sick Leave <(-200hrs)	0.6896	0.0958	*	0.5770	0.1129	*
No Sick Leave	0.6898	0.0629	*	0.6265	0.0808	*
Family (Part. Proxy)	-0.0210	0.0189		-0.0670	0.0245	*
Early Eligible in 97'	0.1013	0.0437	*	0.1683	0.0606	*
Retire Eligible	1.0211	0.0369	*	1.0399	0.0557	*
3rd Year of Ret. Eligible	-0.1709	0.0506	*	-0.2894	0.0841	*
Under FERS	-0.0125	0.0304		-0.0188	0.0405	
Race	-0.0396	0.0203		-0.0645	0.0251	*
N	68.878			33.869		
Log Likelihood	-13723.71			-7555.9		

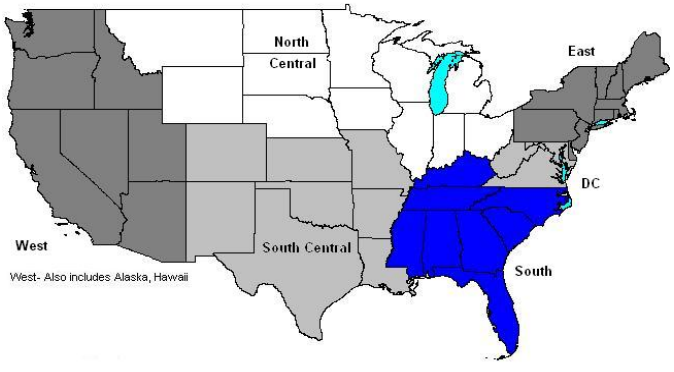
* Denotes significance at the 5% level

Note: The Intercept is South Central US, CSRS employee, without a college degree

Descriptive Statistics

Variable	Revenue Agents		Revenue Officers	
	Mean	Std. Deviation	Mean	Std. Deviation
Southern US	0.129	0.336	0.155	0.362
Western US	0.205	0.404	0.248	0.432
North Central US	0.203	0.402	0.161	0.367
Eastern US	0.239	0.427	0.213	0.410
Washington DC Area	0.054	0.225	0.053	0.224
Bargaining Unit	0.886	0.318	0.874	0.332
Part-Time	0.028	0.166	0.010	0.101
Under 2 Years of Tenure	0.007	0.081	0.006	0.079
Tenure	18.496	8.346	18.425	7.797
Tenure Sq.	411.777	336.759	400.284	309.274
Age (Minus 21)	25.5	8.387	25.322	8.013
Age (Minus 21) Sq.	720.599	437.844	705.397	409.029
Veteran	0.212	0.454	0.223	0.477
Perf. Eval. Outstanding	0.178	0.383	0.232	0.422
Perf. Eval. Poor	0.006	0.080	0.013	0.114
Bachelor's Degree	0.768	0.422	0.515	0.500
Advanced Degree	0.083	0.275	0.064	0.244
Year 1997	0.218	0.413	0.220	0.414
Year 1998	0.206	0.405	0.209	0.406
Year 1999	0.199	0.399	0.198	0.399
Year 2000	0.191	0.393	0.188	0.391
Received No Cash Award	0.140	0.346	0.101	0.302
Manager (Eligible)	0.008	0.090	0.004	0.065
Health Plan (Enrolled)	0.862	0.345	0.855	0.352
Medical Disability	0.025	0.158	0.033	0.179
FERS Early Eligible	0.038	0.191	0.039	0.195
FERS Step 10 (not Eligible)	0.014	0.116	0.016	0.125
CSRS at Step 10 (Eligible)	0.045	0.206	0.034	0.182
Hoarding Annual Leave	0.005	0.068	0.004	0.066
Sick Leave User	0.136	0.343	0.158	0.365
Sick Leave <(-200hrs)	0.004	0.061	0.005	0.074
No Sick Leave	0.009	0.094	0.011	0.105
Family (Part. Proxy)	0.287	0.453	0.347	0.476
Early Eligible in 97'	0.057	0.232	0.055	0.229
Retire Eligible	0.118	0.322	0.088	0.283
3rd Year of Ret. Eligible	0.016	0.127	0.011	0.106
Under FERS	0.492	0.500	0.450	0.498
Race	0.238	0.426	0.330	0.470
N	68,878		33,869	

Figure 7- Map of Geographic Variables



Time Series Modeling and Seasonal Adjustment

Chair: Brian C. Monsell, U.S. Census Bureau, U.S. Department of Commerce

Time Series Modeling Using Unobserved Component Models

Rajesh Selukar, SAS Institute

Unobserved Components Models (UCM), also called Structural Models, decompose the response series into components such as trend, seasonals, cycles, and the regression effects due to predictor series. The components in these models are chosen so that they capture the salient features of the series that are useful in explaining and predicting its behavior. Traditionally, ARIMA models have been the main tools in the analysis of time series data. The UCMs capture the versatility of ARIMA models while possessing the interpretability of exponential smoothing models. A new procedure in SAS/ETS, the UCM Procedure, for structural time series modeling will be demonstrated.

Tools for X-12-ARIMA

Catherine C. Hood, Kathleen M. McDonald-Johnson, and Roxanne M. Feldpausch
U. S. Census Bureau, U.S. Department of Commerce

The Census Bureau's X-12-ARIMA program is a free program used worldwide for time series modeling and seasonal adjustment by statistical agencies and central banks. To use the program more easily and effectively, Census Bureau staff has developed a series of SAS and Excel programs to manage X-12-ARIMA input and output files. The programs include X-12-Graph, the companion graphics package to X-12-ARIMA; a SAS interface to X-12-ARIMA; X-12-Write, a SAS program that writes and edits X-12-ARIMA input files; X-12-Rvw, a SAS program and Excel macro that summarizes X-12-ARIMA diagnostics; and X-12-Data, an Excel macro that converts Excel files to X-12-ARIMA data files.

An Implementation of Component Models for Seasonal Adjustment Using the SsfPack Software Module of Ox

John A. D. Aston, U. S. Census Bureau, U.S. Department of Commerce
Siem Jan Koopman, Free University Amsterdam

An alternative to traditional methods of seasonal adjustment is to use component time series models to perform signal extraction, such as the structural models of Andrew Harvey currently implemented in STAMP, or the ARIMA decomposition models of Hillmer and Tiao currently used in SEATS. A flexible implementation allowing easy specification of different models has been developed using the SsfPack software module of the Ox matrix programming language. This allows the incorporation of heavy-tailed distributions into certain components within the model. Examples of robust seasonal adjustments using this method will be shown.

An Implementation of Component Models for Seasonal Adjustment Using the SsfPack Software Module of Ox

John AD Aston *US Census Bureau and NISS*
Siem Jan Koopman *Free University Amsterdam*

October 2003

1 Introduction

Seasonal adjustment is one of the fundamental tasks that many government agencies must deal with when releasing macroeconomic data. Economists and other officials and commentators rely on seasonally adjusted figures to gain insight into the economy without subtle effects being masked by the sometimes large seasonal differences that can occur. Seasonal adjustment and the software used to produce these adjustments is therefore of much interest.

Currently there are many competing methodologies for doing seasonal adjustment which generally fall into two categories. Firstly there is the non-parametric filter based approaches of software such as X-12-ARIMA (Findley, Monsell, Bell, Otto, and Chen 1998). Secondly, there are the model based approaches, where a model is assumed and the data analyzed in the context of this model. Examples include the STAMP (Koopman, Harvey, Doornik, and Shephard 2000) and SEATS (Gómez and Maravall 1997) software packages. This paper will focus exclusively on this latter category of model based seasonal adjustment.

Here, the model based seasonal adjustments will be considered in a unifying framework, where the models implicit in STAMP and SEATS are special cases, through the use of state space modeling techniques. A software package has been produced to allow easy specification of these models and allow analysis of these more general models, in both Gaussian and non-Gaussian settings.

2 Background theory

Model based seasonal adjustment comes in many forms. However, it can be shown that many of the different model types that people routinely consider for seasonal adjustment all are contained within the same framework of unobserved component models. Here, the commonly used models will be briefly introduced and the general model will also be considered.

Model based seasonal adjustment is defined most generally as follows

$$y_t^{\text{sa}} = y_t - S_t \tag{1}$$

where y_t is the data, S_t the seasonal component and y_t^{sa} the seasonally adjusted data. Here, S_t comes from the model under consideration and is an unobserved component. This leads to the fact that the seasonally adjusted data is also unobserved, a fact sometimes forgotten by data users.

2.1 State Space Models

State space modeling is a technique where the model setup comprises of two complimentary systems, an underlying system of states with time related transition structure, and an observed system which relates these underlying states to the observed output. Breakthroughs such as the Kalman Filter (Kalman 1960) allow maximum likelihood estimation of parameters involved in the system through the use of computationally efficient recursive algorithms. These provide both powerful and fast methods for the analysis of data.

The general state space model is given by

$$\begin{aligned} y_t &= Z\alpha_t + u_t \qquad u_t \sim \mathcal{N}(0, \sigma^2 H) && (2)\\ \alpha_{t+1} &= T\alpha_t + Rv_t \qquad v_t \sim \mathcal{N}(0, \sigma^2 Q) && (3)\\ \alpha_1 &\sim \mathcal{N}(a, \sigma^2 P) && (4) \end{aligned}$$

where y_t is the data, α_t the states, Z is relation of the states to the data, T the transitions of the states and u_t, v_t the underlying independent noise processes, with associated variance matrices H, Q. The initial states must be specified and these are given by a and P, where $P = 0$ indicates known initial states. Durbin and Koopman (2001) give a full account of the use of state space methods for time series analysis.

2.2 Structural models

Structural time series models are becoming increasingly more popular for seasonal adjustment through the use of programs such as STAMP (Koopman, Harvey, Doornik, and Shephard 2000). These models explicitly define the components each with their own variance structure as opposed to the overall

model for the time series itself. The addition of these components yields the full model.

The two most recognizable forms of seasonal structural time series model are the ones that are based on a dummy seasonal formulation and on a set of trigonometric functions at seasonal frequencies which are described in (Harvey 1989). The structural time series model with dummy seasonal S_t (period s), trend T_t and irregular I_t components is given by

$$\begin{array}{rcll} y_t & = & T_t + S_t + I_t, & I_t \sim N(0, \sigma_\varepsilon^2), \\ S_t & = & -\sum_{j=1}^{s-1} S_{t-j} + \omega_t, & \omega_t \sim N(0, \sigma_\omega^2), \end{array} \tag{5}$$

where T_t is modelled as

$$\begin{array}{rcccccl} T_t & = & T_{t-1} & + & D_{t-1} & + & \eta_t \\ D_t & = & & & D_{t-1} & + & \zeta_t, \end{array} \tag{6}$$

for $t = 3, \ldots, n$, known as the local linear trend specification and I_t as Gaussian white noise. The trigonometric seasonal model is the same but with the seasonal component

$$\begin{array}{rcl} S_t & = & \sum_{j=1}^{[s/2]} \gamma_{j,t}, \\ \gamma_{j,t+1} & = & \gamma_{j,t} \cos \lambda_j + \gamma_{j,t}^* \sin \lambda_j + \omega_{j,t}, \\ \gamma_{j,t+1}^* & = & \gamma_{j,t} \cos \lambda_j - \gamma_{j,t}^* \sin \lambda_j + \omega_{j,t}^*, \\ \omega_{j,t} & \sim & N(0, \sigma_{\omega_j}^2), \\ \omega_{j,t}^* & \sim & N(0, \sigma_{\omega_j}^2). \end{array} \tag{7}$$

The trigonometric formulation of the seasonal model is general and therefore assumptions are often made in estimating the components. Most variances, if not all, are assumed to be the same for the disturbances associated with the seasonal trigonometric components. The large number of parameters that would need to be estimated for the full model tends to require large datasets for identifiability so restricting the number of parameters is usually not only wanted but necessary for estimation.

2.3 ARIMA models

ARIMA model based seasonal adjustment relies on the principle that the ARIMA model can be decomposed into different unobserved components. The seasonal component is removed from the data to obtain seasonally adjusted data. The seasonal component is not observed and hence has to be inferred from the data itself through the use of a model.

The general autoregressive integrated moving average (ARIMA) model (**?**) is given by

$$\phi(B)D(B)y_t = \theta(B)\xi_t, \quad \xi_t \sim \mathcal{N}(0, \sigma^2), \quad t = 1, \ldots, n, \tag{8}$$

where $\phi(B)$ and $\theta(B)$ are polynomial functions in the lag operator B with coefficients that ensure $\phi(B)$ has all its zeros outside the unit circle and $\theta(B)$ has all its zeros on or outside the unit circle. The stationary autoregressive moving average (ARMA) model is (8) with $D(B) = 1$. Details of how this can be specified in state space form are given in the appendix.

The canonical decomposition (Hillmer and Tiao 1982; Burman 1980) can be used to take an ARIMA model and decompose it into component form. A general decomposition for general ARIMA model can be of the form

$$y_t = S_t + T_t + C_t + I_t \tag{9}$$

where each of the components, S_t seasonal, T_t trend, C_t cycle and I_t irregular, are ARIMA models of their own. Assumptions need to be made in order to define unique decompositions, and the canonical assumption is that all the white noise is removed from the components and placed into the white noise irregular. Details of how to calculate the decomposition are given by Hillmer and Tiao (1982) and by Burman (1980). These routines are the underpinning of the seasonal adjustment software SEATS (Gómez and Maravall 1997) which is widely used to extract the seasonal component using ARIMA model based decomposition.

2.4 RegComponent Models

Bell (2003) provides the first extensive formal discussion of time series models with a regression mean function and an error process that is sum of independent component time series, each being a known scalar multiple (e.g. 1.0) of a time series that follows an individual ARIMA model.

$$y_t = x_t'\beta + \sum_i h_{it} z_{it} \tag{10}$$

where x_t is a $k \times 1$ vector of fixed explanatory variables and β is the $k \times 1$ vector of coefficients. The ith component constitutes the scalar multiple h_{it}, which is fixed and known for all i and t, and ARIMA process z_{it}. A detailed discussion of the model is given by Bell (2003) and this reference also describes a software program developed at the Census Bureau named regCMPNT for estimating the unknown parameters of these RegComponent models, as they are called. Such models obviously include the structural models of Kitagawa and Gersch (1984) and Harvey (1989), but the paper describes three other kinds of examples, including model with time varying trading day regression coefficients. Once all parameters are specified, after parameter estimation or by inputting the parameter values of the canonical seasonal decomposition models of an estimated regARIMA model following Hillmer and Tiao (1982), this program can calculate the optimal linear estimates of the unobserved components using a state space smoothing algorithm. (RegCMNT does not calculate the parameter values of the component models of the canonical decomposition itself, so these must be obtained externally.)

RegComponent models have been used to allow modelling of effects that can be determined outside the time series, but will have an impact on any seasonal adjustment procedure. A common example of this is the effect of sampling error on the time series. Separate estimates of sampling error are

sometimes available when information is known about survey used to collect the data, and a component can be included in the model to account for this (Bell and Pugh 1990; Bell and Otto 1992).

The RegComponent model framework contains STAMP and ARIMA models as special cases. Thus this framework will be considered in the rest of the paper as the framework of reference, as any method that can be applied generally will also be applicable in these special cases.

2.5 Gaussian vs. Non-Gaussian

So far all the methods considered are based on assuming Gaussian disturbances for the components. However, there are instances where departures from this assumption would be preferred. One major application of non-Gaussian models is to help in the modeling of outliers. Seasonal adjustment can be affected by outliers in the data. Whilst additive outliers are essentially associated with the irregular component, they can lead to changes in other components. As most series are constantly being updated with further data, if Gaussian outlier detection methods are used, based on a cutoff threshold for deciding whether an outlier has occurred, outliers can come into and out of a model of a time series as time passes and observations are added to the series. This can lead to instabilities in all components. If, however, a non-Gaussian heavy tailed distribution, e.g. a t-distribution, is used, no threshold is needed and each datum is weighted according to its probability. A heavy tailed distribution has higher probability of more extreme observations, while outlying observations bias the Gaussian standard deviation estimate and need to be removed before calculation. The use of a heavy tailed distribution allows a more continuous approach to the modeling of outliers in time series, and hence more robust seasonal adjustments.

However, adding non-Gaussian components increases the computational burden as the likelihood now needs to be computationally assessed as opposed to there being explicit analytical form. However, recent advances in state space modeling with non-Gaussian disturbances through the use of importance sampling (Durbin and Koopman 2000) has helped to make this task easier.

3 Implementation

A program implementing models with non-Gaussian components has been designed in the object-oriented matrix programming environment of `Ox`, see Doornik (2001). Extensive use is made of the state space functions in the `SsfPack` library, see Koopman, Shephard, and Doornik (1999). For Gaussian models, once the model is in state space form, the Kalman filter can be used to construct the exact likelihood function. By contrast, non-Gaussian estimation of the model

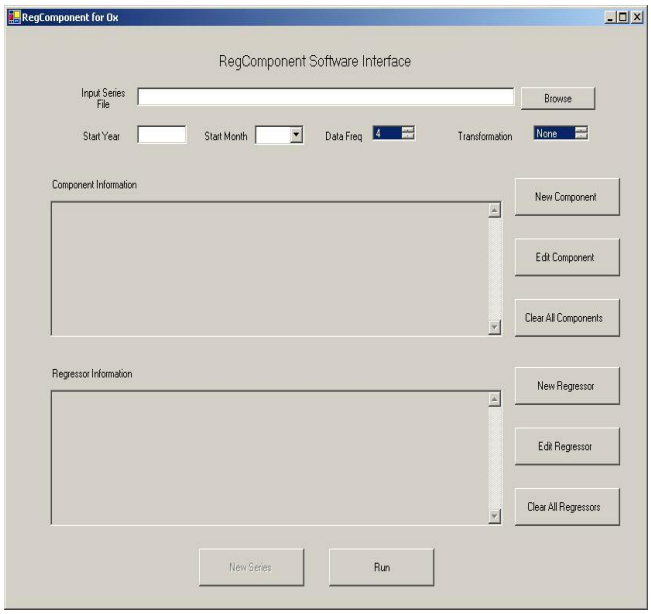


Figure 1: Initial Screen of the User Interface

is carried out using the importance sampling and simulation methods described in detail in Durbin and Koopman (2000) and Durbin and Koopman (2001).

The fundamental idea behind the program is to incorporate different parts of the above ideas into a single program. First it allows the model flexibility of RegComponent to handle many types of component models rather than choosing one of either SEATS or STAMP. Also, it is able to calculate the ARIMA component model decompositions used in SEATS and thus apply non-Gaussian error terms to components, as well as performing integrated seasonal adjustment. A user interface has been designed in Microsoft Visual Basic to aid the user with running the software (see Figures 1, 3, 6 and 9). This flexibility allows an integrated approach to seasonal time series modeling.

4 Examples

Here three examples will be considered. Each looks at a different type of model but all calculations are carried out in the general state space framework. The first example concerns an unobserved component STAMP type model, whilst the second and third contain different ARIMA models, one with a t-distributed component and the other with a sampling error component.

It should be noted that the models are already preselected in terms of their type, and only the parameters are estimated. Here, the problem of selecting the model type will not be considered.

4.1 Example 1 - STAMP: Unobserved Components Model

Data about the annual flow at the Aswan dam on the Nile river (see Figure 2) was analyzed. It is analyzed using a simple

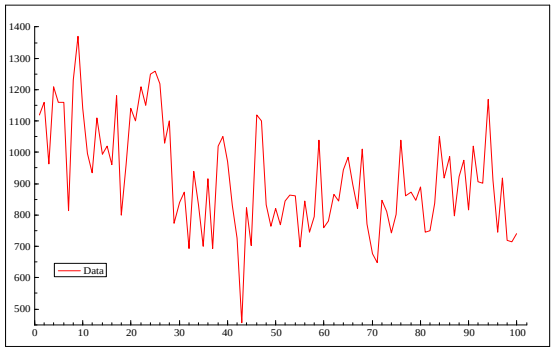


Figure 2: Nile Data. 100 Years of Annual Flow of the river at the Aswan Dam.

STAMP type model containing only a trend and an irregular component.

$$y_t = T_t + I_t \tag{11}$$

The trend (T_t) component is assumed to be a local linear trend model where there is an estimated disturbance term for the level and the slope is assumed fixed. The irregular term (I_t) is assumed to be Gaussian white noise with variance parameter to be estimated. These can be specified in the STAMP model interface portion of the new software (Figure 3). As can be

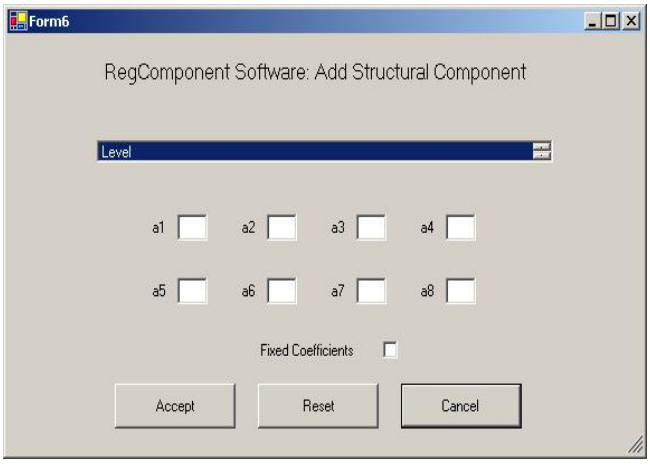


Figure 3: STAMP Model input Interface

seen in Figure 4, it is not only possible to estimate T_t but also possible to estimate confidence intervals around T_t from the data. These can be found from the estimated standard errors

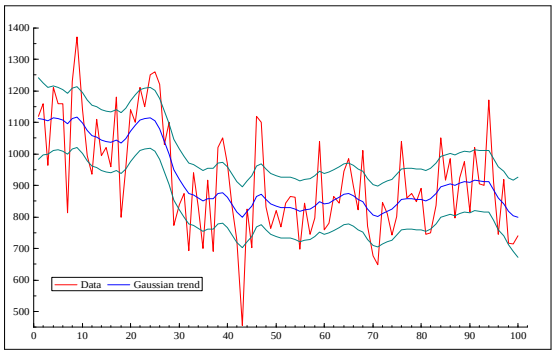


Figure 4: Nile series y_t and trend T_t; also included are confidence intervals for the trend.

on the components within the state space framework, which require very little extra calculation to compute.

4.2 Example 2 - ARIMA: t-distributed Airline Model

The US Census Bureau series, Retail Sales of Automobiles January 1967 to March 1988, was examined. It was determined that there were outliers present using the full data set. However when a shortened subset of the data was taken (Jan 1977-Mar 1988, Figure 5), it was found that the outliers, assumed to be in the data, depended on the threshold that was chosen for outlier detection. If a relatively small threshold

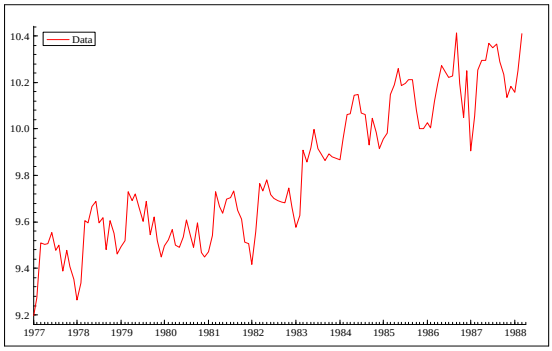


Figure 5: US Automobile Retail Series 1977-1988. Source: US Census Bureau

was used (that is 3.0) then just two of the three outliers determined were found, whilst if this number was larger (that is 4.5) then none of the outliers were detected. Typical values of thresholds used in seasonal adjustment tend to be high due to the large number of tests being performed. Thus a t-distributed model, which does not have the inherent problems

of threshold specification, was used for the analysis.

$$y_t = S_t + T_t + \tilde{I}_t, \quad \tilde{I}_t \sim t(\nu, \sigma^2_{\tilde{I}}), \quad t = 1, \ldots, n. \tag{12}$$

where S_t is the seasonal component, T_t the trend, and $\tilde{I}_t$ the t-distributed white noise irregular. The seasonal and trend components result from the canonical decomposition of a standard three parameter airline model

$$(1 - B)(1 - B^{12})y_t = (1 - \theta B)(1 - \Theta B^{12})\epsilon_t \tag{13}$$

where two of the parameters θ and Θ are the MA parameters associated with the non-seasonal and seasonal lags respectively and the third is the variance of the white noise process ϵ_t. The airline portion of the model can be specified as in Figure 6. Two additional parameters control the t-distribution

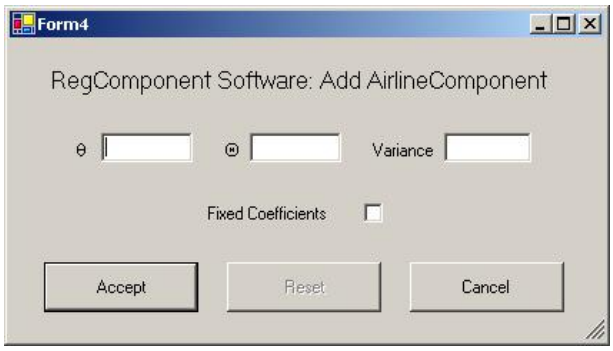


Figure 6: Airline Model Interface

on the irregular component. However, as the irregular term is modeled using a t-distribution, the decomposition is done within each maximum likelihood estimation step, as opposed to the usual method of estimating the final ARIMA model and then performing the decomposition on the final model. The decomposition is constrained to be admissible throughout the maximum likelihood process. The components resulting from fitting the model with the t-distribution to account for outliers can be seen in Figure 7.

4.3 Example 3 - RegComponent Model: ARIMA + Sampling Error

The US Bureau of Labor Statistics series on Teenage Unemployment from 1972 to 1984 (Figure 8) was analyzed. There is a sampling error component in this series and a model was selected which incorporated this component:

$$\begin{gathered} \mathbf{y_t} = \mathbf{S_t} + \mathbf{T_t} + \mathbf{I_t} + \mathbf{\Sigma_t} \\ S_t + T_t + I_t = \text{ARIMA}(2,0,0)(0,1,1)_{12} \\ \Sigma_t = \text{fixed ARMA}(1,1) \end{gathered} \tag{14}$$

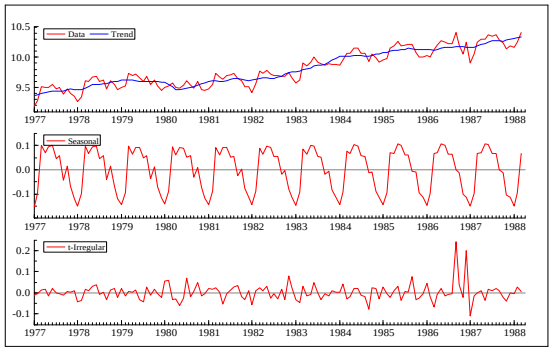


Figure 7: (top) Auto series y_t and trend T_t; (middle) seasonal component S_t; (bottom) combined irregular and outlier component $\tilde{I}_t$.

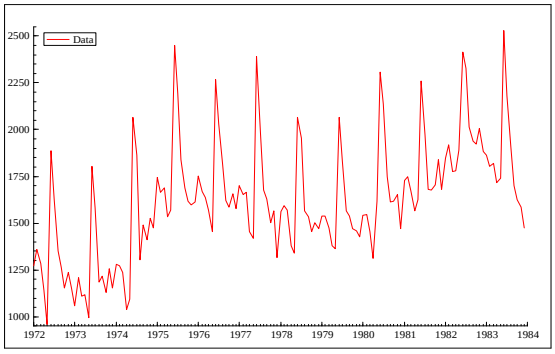


Figure 8: Teenage Unemployment Data 1972-1984. Source: US Bureau of Labor Statistics

where S_t is the seasonal component, T_t the trend, I_t a Gaussian white noise irregular and Σ_t the sampling error component. Here the ARMA model for the sampling error was obtained from prior analysis of the data, as were the associated weights with it, and S_t, T_t, and I_t were derived from the ARIMA model through the canonical decomposition.

The interface for the ARIMA portion of the model can be seen in Figure 9

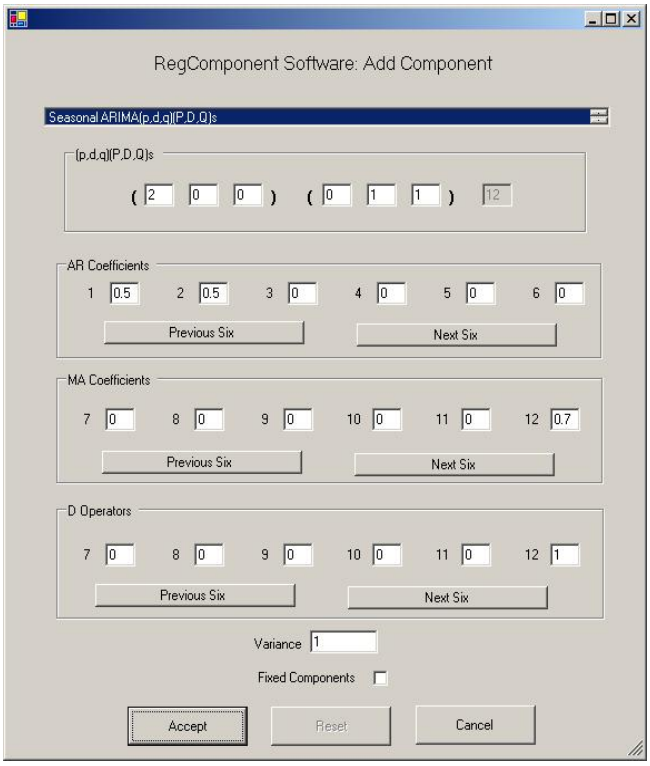


Figure 9: RegComponent ARIMA model Interface

As can be seen in Figure 10, the removal of a sampling error component changes the seasonal adjustment. There is much apparent structural variability (level changes) in the adjustment when sampling error is not accounted for, while removing the sampling error component removes this variation to a great extent, stabilizing the adjustment.

5 Conclusions

Model based seasonal adjustment is increasingly being used to adjust macroeconomic data. However, it is often perceived that some of the most commonly used methods are greatly different. Here a framework has been used to show that the models can all be considered in a similar fashion through the use of state space modeling techniques. This allows many different techniques to be combined into a single software application where all the models can be estimated in a unified way.

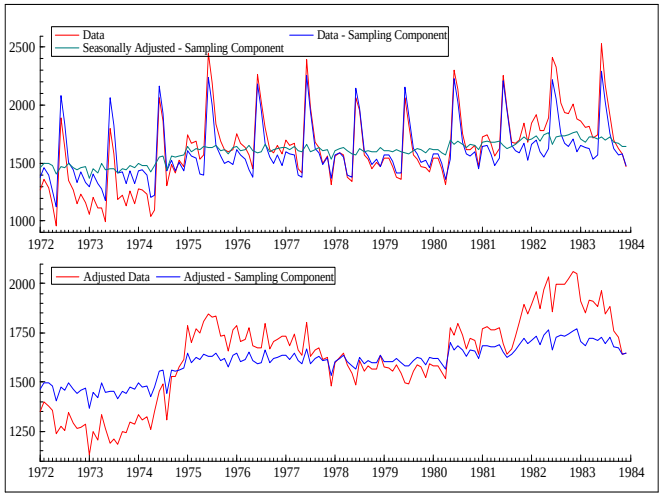


Figure 10: (top) Teen series y_t, series without sampling error $y_t - \Sigma_t$ and seasonally adjusted series without sampling error $y_t - \Sigma_t - S_t = y_t^{\text{sa}} - \Sigma_t$; (bottom) seasonally adjusted series $y_t - S_t = y_t^{\text{sa}}$ and seasonally adjusted series without sampling error $y_t^{\text{sa}} - \Sigma_t$.

Accompanying Software

The software that was used to perform the analysis and to generate the output in the examples will be available from the authors in Early 2004.

Website
http://staff.feweb.vu.nl/koopman/

Acknowledgements

Some of this research was carried out while the second author visited the US Census Bureau as an ASA/NSF/BLS/US Census Research Fellow. The authors would like to thank William Bell, David Findley, Brian Monsell, Marius Ooms and Stuart Scott for help, support and many insightful discussions. All errors in the paper are the authors.

Disclaimer

This paper reports the results of research and analysis undertaken at the US Census Bureau. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are

those of the authors and not necessarily those of the US Census Bureau.

Appendix: ARMA model in State Space

The ARMA model (8), with $D(B) = 1$, is given in state space form with the first element of the state vector α_t equal to y_t and with

$$Z' = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}, T = \begin{bmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \phi_2 & 0 & 1 & & 0 \\ \vdots & \vdots & & \ddots & \\ \phi_{m-1} & 0 & 0 & & 1 \\ \phi_m & 0 & 0 & \cdots & 0 \end{bmatrix}, R = \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{m-2} \\ \theta_{m-1} \end{pmatrix}, \tag{15}$$

with $m = \max(p, q+1)$ and further $H = 0$ and $Q = 1$. The mean vector of the initial state vector is $a = 0$ and the initial variance matrix, P, is the solution of

$$(I - T \otimes T)\,\mathrm{vec}(P) = \mathrm{vec}(RR'). \tag{16}$$

The ARIMA model (8) can also be formulated in state space form in similar ways. State elements associated with $D(B)$, the non-stationary part of the ARIMA model, require special initialisation conditions (see Durbin and Koopman (2001) for more details).

References

Bell, W. R. (2003). On RegComponent time series models and their applications. working paper, January 2003, Statistical Research Division, U.S. Census Bureau.

Bell, W. R. and M. Otto (1992). Bayesian assessment of uncertainty in seasonal adjustment of time series components. Research Report Number 92/12, Statistical Research Division, U.S. Census Bureau.

Bell, W. R. and M. G. Pugh (1990). Alternative approaches to the analysis of time series components. In A. C. Singh and P. Whitridge (Eds.), *Analysis of data in time: Proceedings of the 1989 International Symposium*, pp. 105–116. Statistics Canada.

Burman, J. P. (1980). Seasonal adjustment by signal extraction. *Journal of Royal Statistical Society A 143*, 321–37.

Doornik, J. A. (2001). *Object-Oriented Matrix Programming using Ox 3.0*. London: Timberlake Consultants Press.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). *Journal of Royal Statistical Society B 62*, 3–56.

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press.

Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, and B. C. Chen (1998). New capabilities of the X-12-ARIMA seasonal adjustment program (with discussion). *Journal of Business and Economic Statistics 16*, 127–77. http://www.census.gov/ts/papers/jbes98.pdf.

Gómez, V. and A. Maravall (1997). Programs TRAMO and SEATS : Instructions for the user (beta version: June 1997). Working Paper 97001, Ministerio de Economía y Hacienda, Dirrectión General de Análisis y Programación Presupuestaria, Madrid.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

Hillmer, S. C. and G. C. Tiao (1982). An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association 77*, 63–70.

Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Engineering, Transactions ASMA, Series D 82*, 35–45.

Kitagawa, G. and W. Gersch (1984). A smoothness priors - state space modeling of time series with trend and seasonality. *Journal of the American Statistical Association 79*(386), 378–389.

Koopman, S. J., A. C. Harvey, J. A. Doornik, and N. Shephard (2000). *Stamp 6.0: Structural Time Series Analyser, Modeller and Predictor*. London: Timberlake Consultants.

Koopman, S. J., N. Shephard, and J. A. Doornik (1999). Statistical algorithms for models in state space form using SsfPack 2.2. *Econometrics Journal 2*, 113–66. http://www.ssfpack.com/.

Long-Range Forecasting

Chair: Stephen MacDonald, Economic Research Service, U.S. Department of Agriculture

The Changing Nature of the Links Between Fertilizer and Energy Prices

David Torgerson, Economic Research Service, U.S. Department of Agriculture

As the role of crude oil prices in triggering business cycles has diminished, that of natural gas has increased. There is concern that the natural gas market situation may trim U.S. economic growth for the next several years. In addition, during the time that crude oil prices were influential in economic growth up to the mid-1980s, oil prices also were a key factor determinate of natural gas prices. When the connection between oil and natural gas prices was tight it was reasonable to forecast nitrogen-based fertilizer prices as dependent on crude oil prices. As the interconnection between crude oil and natural gas prices has weakened, the importance of natural gas prices in nitrogenate pricing has risen. At the same time, natural gas prices have also become more volatile. As result of these two recent trends, fertilizer prices have become more volatile. Due to bottlenecks in natural gas supplies, fertilizer prices are likely to rise relative to other major farm inputs and continue to be volatile over the next decade.

Problems in Forecasting the Chinese Agricultural Economy

James Hansen and Fred Gale, Economic Research Service, U.S. Department of Agriculture

China is a key player in the world agricultural commodity markets with large global impacts. Consequently, accurate projections are crucial for analysis. ERS maintains agricultural economic models for baseline projections and policy analysis. This research identifies and addresses problem areas in Chinese agricultural projected simulations. Comparison of historical baseline projections with actual data and projections by various research institutions is conducted. Reasons for deviations of projections from actual data are investigated. Factors contributing to forecast errors include model structure, parameters, data, and inadequate information on China's agriculture, marketing, and policy. Improvements in models and projections are presented.

Quality Assurance Methods for Variance Estimates of a Family of Time Series

Jerry L. Fields, Bureau of Labor Statistics, U.S. Department of Labor

Time series estimates are used as one component of small-area employment estimation in the Current Employment Statistics program at the Bureau of Labor Statistics. This presentation will describe quality assurance procedures used to identify time series that may be harmful to the estimation process. Standard error estimates of long-range projections from time series models are compared against modeled values from generalized linear models to measure quality. A sequence of Bernoulli trials is used to identify series with unstable or pathological variance estimates. A nonparametric method is described to assign a quality score to the time series estimates.

Russian Grain and Meat Production and Trade: Forecasts to 2012

William Liefert, Olga Liefert, Stefan Osborne, Eugenia Serova (Russian Institute for Economy in Transition), and Ralph Seeley, Economic Research Service, U.S. Department of Agriculture

The paper presents forecasts for Russian agricultural production, consumption, and trade of grain and meat for 2012. Using a projections model for Russian agriculture developed at the Economic Research Service, forecasts are made for various scenarios, depending on different assumptions concerning agricultural productivity growth and policy changes. Preliminary projections show that, in the absence of the tariff rate quotas for meat imports that Russia created in spring 2003, Russia's already large imports of meat would grow substantially, and that Russia could become a major exporter of grain.

QUALITY ASSURANCE METHODS FOR VARIANCE ESTIMATES OF A FAMILY OF TIME SERIES

Jerry L. Fields
Bureau of Labor Statistics

Introduction.

The Current Employment Statistics (CES) program at the Bureau of Labor Statistics (BLS) uses long-range (10-12 months beyond the sample horizon) forecasts from time series as one of the weighted components in models of employment for small areas. An automated process is used to identify an ARIMA model for each series and produce employment estimates and error estimates for the desired future values. (In this application, the goal is not to develop a descriptive model of the observed portion of the time series; rather, it is to predict the future.) Any automated modelling process raises the risk of model misspecification. Variance estimates for non-stationary time series may be especially vulnerable to misspecification and result in either inflation or collapse of the variance estimate. Weights obtained from out-of-control variance estimates will place incorrect weight on possibly corrupt or misspecified time series relative to other model components and cause harm to the final small-area estimate.

For a long time series a sequence of out-of-sample estimates can be used to create a quality history of forecast performance. Control charts and other quality assurance procedures can be used to detect suspicious variance estimates in the current era and to identify obstreperous or otherwise ill-behaved series. In this report the creation of a quality history will be discussed. Next, a model will be presented for time series standard error estimates. Finally, quality assurance measures will be described.

Quality history.

The CES program consists of more than 7000 employment time series classified by geographic region and major industrial division. Each series consists of 111 observations of monthly employment from January 1992 through March 2001. For each series there were 21 models created based on subseries of lengths $T = 51, 54, \Lambda\ , 111$. For each model the mean standard error s in the long-range forecast epoch and the mean employment x in the final quarter of the subseries were recorded,

$$s = \frac{1}{3}\sum_{T+10}^{T+12} s_t\ ,$$

$$x = \frac{1}{3}\sum_{T-2}^{T} x_t\ .$$

The sequences of 21 mean standard error, mean employment level pairs (s, x) are a quality history for each series.

Variance models.

Generalized linear models (GLM) [Myers, Montgomery, Vining] were developed to describe the time series standard error estimates for the family of CES employment series. Factors included in the GLMs were mean employment in the final quarter of the sample as a covariate, and industrial division as a categorical effect. The employment covariate was transformed by the common log, $y = \log x$. The industrial divisions are the thirteen NAICS supersectors and their aggregate, δ_i, $i = 1, \Lambda\ , 14$; where $\delta_{14} \equiv 0$ for the aggregate industry effect.

Standard error is a gamma distributed and positive-definite response. The appropriate link is the natural log. Employment and industrial division are significant effects with no evidence of interaction. Other covariates (time of forecast epoch) and effects (geography) were considered and rejected. The GLM for standard error estimate is

$$\ln s_i = -0.956(8) + 1.7041(14)\, y + \delta_i\,,$$

where the uncertainty values are one standard deviation in the in the final digit of the coefficients. The range of the industry effect is $-0.234(5) \le \delta_i \le 0.617(5)$. Hence, employment and industrial division are sufficient for generalized variance functions of the standard error estimate. The GLM estimates will be used as a quality assurance standard for the time series estimates.

Quality assurance.

The GLM produces standardized deviance residuals (SDR) that are analogous to normal z-scores in ordinary ANOVA. For each series then, there is a sequence of 21 standardized deviance residuals; one SDR for each forecast epoch. (GLM is available as **proc genmod** in SAS™. SAS does not as yet, version 8.2, provide prediction intervals for a new observation, only confidence intervals for the mean. The use of SDR here is an alternative to prediction intervals.)

Large, positive SDR indicate an inflated time series variance estimate relative to the family of time series. Large, negative SDR indicate variance collapse. If the sequence of SDRs is sufficiently long, standard control chart methods may be used with appropriate control limits on SDR to identify out-of-control series. For sequences too short for control charts other methods are necessary.

One alternative to control charts is to establish rules based on the empirical distribution of SDR for the entire family of time series. If all series have similar quality, then the frequency that the SDR exceeds a critical value follows a binomial distribution for each series. Another critical value for failure frequency establishes a reject/no-reject criterion for each series. A second alternative is to use a nonparametric test to rank each series according to the distribution of its SDRs relative to the entire family.

Binomial method.

If all series in the family have identical variance quality, then the distribution of SDR z for a single series is the same as the empirical distribution of the SDR for the entire family. Similarly for absolute SDR; upper quantiles of $|z|$ are displayed in Table 1. To give equal weight to testing for variance inflation and collapse obtain the empirical probability p_c that the absolute SDR exceeds some critical value z_c, $p_c = \mathrm{P}_{\mathrm{emp}}\left(|z| > z_c\right)$. Then, each series consists of a sequence of $N = 21$ Bernoulli trials that test $|z|$ against z_c. The rejection frequency n_i within a series is a binomial distribution, $n_i \sim \mathrm{Bin}\left(N, p_c\right)$. Hence, a critical rejection frequency n_c, or a significance level α may be established for the series, $\mathrm{P}_{\mathrm{Bin}}\left(n_i \geq n_c; N, p_c\right) = \alpha$. Alternatively, the p-value for the observed rejection frequency n_i for series i is $p = \mathrm{P}_{\mathrm{Bin}}\left(n \geq n_i; N, p_c\right)$. To give unequal weight to detection of variance inflation and collapse use asymmetric critical values, $p_{c+} = \mathrm{P}_{\mathrm{emp}}\left(z > z_{c+}\right)$, $p_{c-} = \mathrm{P}_{\mathrm{emp}}\left(z < z_{c-}\right)$, $p_c = p_{c+} + p_{c-}$. The administrative procedure is to first choose either z_c or p_c, then choose either n_c or α. Experience will suggest appropriate values that provide acceptable power.

A distinction will be made between common cause variance defects and special cause variance defects. Series that have common cause defects display a pattern of poor variance estimates, frequently much larger or much smaller than expected for the series size. Those series may or may not have an adverse effect on subsequent use of the variance and employment estimates; they should be reviewed by an analyst. For the detection of common cause defects the selected critical values were $z_c = 4$, $p_c = \mathrm{P}_{\mathrm{emp}}\left(|z| > 4\right) = 3.3 \times 10^{-3}$ and $n_c = 4$, $\alpha = \mathrm{P}_{\mathrm{Bin}}\left(n \geq 4; N = 21, p = p_c\right) = 7.0 \times 10^{-7}$. Figure 1a shows an example of a series with common cause variance defects; the first four quarters of long-range forecasts have unusually large confidence intervals. The inflation of the confidence intervals is enhanced for easier viewing in the chart of absolute relative error, figure 1b.

Series that have special cause defects have at least one pathological variance estimate that demands investigation. To detect series with special cause defects the selected critical values were $z_c = 6$, $p_c = \mathrm{P}_{\mathrm{emp}}\left(|z| > 6\right) = 8.2 \times 10^{-4}$ and $n_c = 1$, $\alpha = \mathrm{P}_{\mathrm{Bin}}\left(n \geq 1; N = 21, p = p_c\right) = 0.0170$. Figures 2 and 3 show examples of pathological variance inflation. In figure 2 the defect is in the contemporary era and there are no out-of-sample observations for comparison. In figure 3 out-of-sample observations are available. Figure 4 shows an example of variance collapse. On inspection, the pathological examples appear to be a result of model misspecification by over-differencing.

Nonparametric method.

The binomial method is sensitive to subjective and discrete rejection criteria, $n\left(|z| > z_c\right) \geq n_c$. The studentized deviance residuals from the GLM are continuous measures of error. That continuity could be used to develop a more efficient method of detecting

series with unstable variance estimates. The distribution of SDR is not obvious for a gamma response with a $\ln$ link. (It's probably asymptotically normal—*everything* in statistics becomes asymptotically normal at some time.) However, the distribution is scale-free; and that makes it especially amenable to nonparametric methods.

If all series in the family have similar quality with respect to variance estimates, then the SDRs from an individual series will be uniformly distributed across the universe of SDRs. Series with larger than expected variance estimates will have SDRs that cluster in the upper tail of the universe, whilst series with smaller than expected estimates will cluster in the lower tail. Whence, series that avoid the median of the distribution will be unstable for variance estimation. Those series can be detected by the Siegel-Tukey nonparametric test [Lehman].

The universe of SDRs is ordered and ranks are assigned inward from alternate tails to the median; $R_{(1)} = 1, \quad R_{(N)} = 2, \quad R_{(2)} = 3, \quad R_{(N-1)} = 4, \Lambda$. Rank sums R_i are computed for each series. The null hypothesis is that the family of series has a common median. The null hypothesis is rejected strongly and supports the claim that series quality as measured against the GLM SDR varies across the family.

Rank sums may be rescaled to create a score Q that is asymptotically standard normal (didn't I tell you),

$$Q = \frac{R_i - \overline{R}}{s^*}.$$

Thence, that score is a continuous measure of quality. Series with large negative quality score Q cluster in the tails of the SDR distribution. Those series will have occurrences of very large or very small variance for their size; and, may have stable variance. Series with large positive quality score cluster near the median of the SDR distribution and have very stable variance. Series with quality score near zero display expected random variations of variance around the modelled value.

Experience will guide the choice of a critical value Q_c for the rejection criteria $Q < Q_c$. The selection of $Q = -6.5$ resulted in 194 rejected series out of 7190 (2.7%). Those are analogous to series with common cause defects and should be subject to analyst review rather than automatic rejection. Figures 5-8 display a range of quality characteristics for a set of similar size series from a common industry.

Summary.

Procedures were described to identify series with variance estimates unusual with respect to a large family of series. Series with common cause defects display a trend of unusual variance estimates. Series with special cause defects have at least one pathological variance estimate. A nonparametric method was used to identify series with common cause defects and to assign a relative quality score to each series. A binomial method was used to detect series with special cause defects and may also be used to detect common cause defects. Both methods depend on a generalized linear model to describe the time series estimate of standard error and measure the error of that estimate relative to the GLM estimate.

Tables 2 and 3 show the observed distributions of common cause, $n(|z| > 4)$, and special cause, $n(|z| > 6)$, defects from the binomial method. Table 4 shows the observed quantiles of the quality score Q from the nonparametric method. None of those are in agreement with their respective expected distribution. That is a result of variance quality that cannot be explained by the GLM; *viz*., either lack of fit or unusual variance estimates. When the SDR are randomly reassigned to groups the expected distributions are obtained. Hence, the quality assurance procedures are effective at detecting series with unusual variance estimates.

Other responses may be used in place of the time series standard error estimate. Mean absolute relative error of the out-of-sample long-range employment forecasts was used to detect level shifts and forecast misspecification.

References.

E. L. Lehman, *Nonparametrics: Statistical Methods Based on Ranks*, Holden-Day, San Francisco (1975).

R. H. Myers, D. C. Montgomery, G. G. Vining, *Generalized Linear Models*, John Wiley, New York (2002).

Table 1. Empirical quantiles p of absolute standardized deviance residuals $|z|$ from GLM of time series standard error estimates. The quantiles are the upper tail probabilities. Compiled from 21 observations on each of 7190 series.

$\|z\|$	p
1.87	0.0500
2.00	0.0396
2.40	0.0200
2.88	0.0100
3.00	0.0087
3.56	0.0050
4.00	0.0033
4.70	0.0020
5.00	0.0016
5.67	0.0010
6.01	0.0008

Table 2. Distribution of rate of common cause defects n over the family of 7190 series. The expected distribution is binomial with $N = 21$ and empirical $p = 3.3 \times 10^{-3}$ as discussed in the text. The randomized distribution is the observed rate when z are randomly reassigned.

$n(\|z\| > 4)$	Frequency		
	Observed	Expected	Randomized
0	7017	6704	6695
1	105	470	488
2	16	16	7
3	7	0	0
4+	45	0	0

Table 3. Distribution of rate of common cause defects n over the family of 7190 series. The expected distribution is binomial with $N = 21$ and empirical $p = 8.2 \times 10^{-4}$ as discussed in the text. The randomized distribution is the observed rate when z are randomly reassigned.

$n(\|z\| > 6)$	Frequency		
	Observed	Expected	Randomized
0	7139	7068	7068
1	32	121	121
2	5	1	1
3+	14	0	0

Table 4. Quantiles of nonparametric quality score Q for the standardized deviance residuals z as discussed in the text. The expected distribution is standard normal. The randomized distribution is the observed rate when the z are randomly reassigned.

p	Observed	Expected	Randomized
0.99	-7.21	-2.33	-2.31
0.95	-5.74	-1.65	-1.67
0.90	-4.60	-1.28	-1.31
0.75	-2.32	-0.67	-0.68
0.50	0.29	0	0.01
0.25	2.59	0.67	0.70
0.10	4.09	1.28	1.29
0.05	4.79	1.65	1.64
0.01	5.77	2.33	2.34

Figures.

Figure 1. An example of an historically unstable series, $n(|z| > 4) = 4$. The first four quarters (1997) of forecasts have larger than expected variance estimates for the 10-12 month ahead forecasts. Solid squares are the forecasts, solid circles are the observed levels.

Figure 2. A series with pathological variance inflation in the out-of-sample era (2001.Q3). From time series $\hat{s} = 6500$, from the GLM $\hat{\hat{s}} = 500$ and SDR $z = 9.7$.

Figure 3. A series with pathological variance inflation in the historic era (2000.Q4). Time series $\hat{s} = 2400$; GLM $\hat{\hat{s}} = 305$, $z = 6.7$.

Figure 4. A series with pathological variance collapse in the historic era (2001.Q1). Time series $\hat{s} = 10^{-13}$; GLM $\hat{\hat{s}} = 130$, $z = -9.7$.

Figure 5. A series with consistently larger than expected variance, $Q = -7.5$, $z \sim 3.1$. Time series $\hat{s} \sim 440$, GLM $\hat{\hat{s}} \sim 160$.

Figure 6. A series with consistently smaller than expected variance, $Q = -6.3$, $z \sim -1.8$. Time series $\hat{s} \sim 70$, GLM $\hat{\hat{s}} \sim 145$. The variance estimates are also very stable.

Figure 7. A series with very stable variance estimates near the expected values, $Q = 5.6$, $z \sim 0$. Time series $\hat{s} \sim 140$, GLM $\hat{\hat{s}} \sim 140$.

Figure 8. A typical series, $Q = -0.1$, $z \sim -0.9$. Time series $\hat{s} \sim 95$, GLM $\hat{\hat{s}} \sim 130$.

RUSSIAN GRAIN AND MEAT PRODUCTION AND TRADE: FORECASTS TO 2012

William Liefert, Olga Liefert, Stefan Osborne, Ralph Seeley, Economic Research Service, USDA
Eugenia Serova, Russian Institute for Economy in Transition[1]

Introduction

During the 1980s, Russia (as well as the Soviet Union in the aggregate) was a large importer of grain, soybeans, and soybean meal. Most of the imports were used as feed to support the policy-driven expansion of the livestock sector. When economic reform began in the early 1990s, Western studies (Koopman 1991, Liefert et al. 1993, Tyers 1994) forecasted that effective reform that increased agricultural productivity could turn Russia (and the former USSR as a whole) from a major grain importer into a major grain exporter.[2]

By 2000, however, these forecasts had not been fulfilled. Although Russia had virtually eliminated its imports of soybeans and soybean meal and substantially reduced its grain imports, it remained a grain importer. During 1998-2000, annual grain net imports averaged 3.7 million metric tons (mmt).[3] Rather than importing large amounts of feed to maintain sizeable livestock herds, Russia during reform has slashed its livestock inventories and production, and become a big importer of meat (2.65 mmt in 2001).[4]

In 2001 and 2002, Russia for the first time during the transition period exported a significant amount of grain (7.0 and 8.6 mmt of gross exports in the two years, and 5.3 and 7.0 mmt net). The exportable surplus coincided with rising grain production over 1999-2002, yielding bumper harvests in both 2001 and 2002 of 82 mmt.[5] Although weather was favorable in both two years, there are also signs that Russia might be improving its agricultural system to increase productivity, perhaps presaging a long-term rise in output. For example, new large, vertically integrated producers in the Russian agro-food economy, typically financed and managed by enterprises outside of agriculture, could bring more efficient management to the sector than the former state and collective farms that currently dominate agriculture. Also, in July 2002 the Russian Duma passed a land law that, more than any previous federal legislation, legalizes and codifies the private ownership and market sale of agricultural land, which could also boost productivity. Could rising productivity finally fulfill the forecasts of the early 1990s by turning Russia into a major grain exporter? Might productivity growth also expand the country's livestock sector, such that Russia substantially reduces its meat imports?

Policy changes could also affect Russia's future agricultural trade. In its negotiations for accession to the World Trade Organization (WTO), Russia argues that its current policies with respect to each of the three "pillars" of the Uruguay Round Agreement on Agriculture—market access, export subsidies, and domestic support—are relatively moderate. Russia is therefore negotiating for "bound" (maximum allowable) limits on agricultural import tariffs, export subsidies, and support to producers that exceed current levels.

Russian agricultural interests are particularly concerned about the country's large imports of meat, and in particular poultry, given that imports in 2001 and 2002 supplied about two-thirds of national poultry consumption. In spring 2002, Russia banned imports of U.S. poultry, citing health concerns and irregularities in certification procedures, with unresolved problems remaining as of spring 2003. In spring 2003, the Russian government imposed tariff rate quotas (TRQs) on its meat imports (and for poultry a pure quota), to be maintained for a minimum of three years. The low tariff quotas for all the meats sum to 1.92 mmt, compared to Russia's 2001 total beef, pork, and poultry imports of 2.65 mmt.

This paper examines how changes in Russian agricultural productivity, consumer income, and policies, as well as changes in other variables, could affect the country's production and trade in grain and meat. The paper uses a model for Russian agriculture to forecast Russian production, consumption, and trade for grain (wheat and coarse grains) and meat (beef, pork, and poultry) for the year 2012.[6] Various scenarios are run depending on different assumptions concerning two key variables: (1) the degree of productivity growth in Russian agriculture (reflecting the effectiveness of further agricultural reform); and (2) trade policy developments.

The next section of the paper examines the model used in the forecasts, as well as the assumptions made for key exogenous variables. The subsequent section examines the forecast results, and the conclusion presents the paper's main findings.

The Model and Forecast Assumptions

The forecasting model we use for Russian agriculture was developed by the Economic Research Service (ERS) of the U.S. Dept. of Agriculture. The model is spreadsheet-based in Excel, and utilizes the Country Projections and Policy Analysis (CPPA) model-builder designed at ERS (Hjort and van Peteghem 1991).

The Russia model is dynamic and partial equilibrium in nature, and projects agricultural production, consumption, and trade each year from marketing year (July-June) 2001/2002 for crops and calendar year 2002 for livestock products through 2012. The model consists of supply and demand equations for commodities that use synthetic (rather than estimated) own and cross-price elasticities. Crop production is forecasted using area and yield functions. Each crop area function depends on current and lagged prices for all crops in the model (grain, oilseeds, and sugar), while yield functions depend on lagged prices and an exogenous productivity trend. Livestock products have production functions, which depend on commodity and feed prices and exogenous productivity trends.

Key exogenous variables and parameters for which values must be assumed are: (1) GDP; (2) the real exchange rate; (3) price transmission elasticities; (4) productivity (specifically yields and feed coefficients); and (5) policies (mainly subsidies and trade controls, such as tariffs and quotas).

GDP

Most macro forecasters for Russia (PlanEcon, Oxford Economics) project that GDP will grow during our forecasting period by 4-5 percent a year, which we also assume in our projections. (The Russian Ministry of Economics also forecasts annual growth of 4 percent over 2003-05.) GDP growth will raise consumer income, and thus demand for foodstuffs. Demand for high value products such as meat, which have relatively high income elasticities of demand, should be particularly stimulated.

Real Exchange Rate

Russia's economic crisis of 1998 resulted in major depreciation of the Russian ruble vis-à-vis Western currencies, in both nominal and real terms. From the start of the crisis in August 1998 through the end of 1999, the ruble depreciated in nominal terms by about 75 percent, and in real terms by about 55 percent. In 2000, however, the ruble began appreciating in real terms (as the inflation rate exceeded the nominal rate of currency depreciation), with real appreciation in 2000 and 2001 equaling 13 and 6 percent, respectively (PlanEcon).

In the view of most Western macroeconomic forecasters for Russia (PlanEcon, Oxford Economics), the ruble is still undervalued somewhat relative to Western currencies. For example, PlanEcon predicts that over 2002-2006, the ruble will appreciate in real terms by about 15 percent. Some further real appreciation seems especially likely if the Russian economy continues growing at a (relatively) high rate. We therefore assume that the ruble appreciates in real terms over 2002-2006 by 15 percent, and then remains constant in real terms over the rest of the forecasting period. By lowering the real price of tradable goods, real appreciation should reduce production and stimulate consumption, thereby decreasing exports (for net export commodities) and increasing imports (for net import commodities).

Price and Exchange Rate Transmission

An important element within the model is the degree to which changes in world trade (border) prices for commodities and the real (inflation-adjusted) exchange rate are transmitted to Russian domestic producer and consumer prices. The specific model variables that capture transmission are price and exchange rate transmission elasticities (TEs), which equal the percent change in the Russian domestic price for a commodity divided by the percent change in the border price or real exchange rate. We assume that the TEs for all commodities initially equal 0.5, and then rise through the course of the forecasting period to 0.75.

The low assumed price and exchange rate TEs at the start of the forecasting period reflect a more general assumption that even after ten years of economic reform, Russian agricultural and food markets are still not well integrated into world markets. The main reason is that the internal physical and institutional infrastructure that a well-operating market-oriented agricultural economy needs is still weak. Deficient infrastructure raises domestic transportation and transaction costs, which segments domestic regional markets from each other, as well as cuts these regional markets off from the world market.

Although storage capacity is also inadequate, the main weakness in physical infrastructure is transportation, particularly the poor road system. Major deficiencies in institutional infrastructure are the weak systems of

market information and commercial law (Osborne and Trueblood 2002). Market-oriented producers need a commercial environment that reduces risk and transaction costs, which must include a commercial legal system that protects property and enforces contracts. Wehrheim et al. (2000) argues that undeveloped institutions are the main problem facing Russian agriculture.

Osborne and Liefert (forthcoming) provide empirical evidence that transmission elasticity in the Russian agro-food economy is low. They compute price and exchange rate elasticities for consumer retail prices for beef and pork in 31 separate Russian cities over 1994-99. Their results support our assumption that TEs for all commodities at the beginning of the forecasting period equal 0.5 (half of 100 percent).

The increase in our assumed TE values from 0.5 to 0.75 over the forecasting period reflects the belief that continued reform will improve institutional infrastructure for the agro-food economy. Recent legislation passed by the Russian Duma concerning taxation, the judicial system, and land affairs (the land code) suggests that such improvement is possible, though it is unclear how effectively such legislation will be implemented.

Productivity

We make forecasts for two scenarios involving productivity: one based on the assumption that productivity growth over the forecasting period is relatively low, and one based on the assumption that growth is relatively high. The only variables in the model that reflect agricultural productivity are yields for crops and feed coefficients for livestock production. Feed efficiency is not the sole element that determines the productivity of livestock production, examples of other factors being the productivity of labor and material inputs such as energy. To compensate for the fact that, in our model, feed coefficients are the sole productivity-capturing variables for livestock production, we inflate a bit the values chosen for these coefficients in our forecasting runs.

Since agricultural reform began in the early 1990s, grain yields have fallen. Over 1999-2001, average annual yields for grain in the aggregate equaled 1.51 metric tons, compared to 1.68 metric tons over 1987-1991. In our low productivity growth scenario, we assume that over the forecasting period grain yields rise, such that by the outyear they return to their pre-reform levels. Feed coefficients (tons of non-pasture feed per ton of slaughter weight of meat) have increased substantially during transition, by about 50 percent, revealing markedly worse feed efficiency. In our low productivity growth scenario, we assume that feed coefficients (all types of feed for all types of meat) fall by 1.5 percent a year. In our high productivity growth scenario, we assume that grain yields both recover to pre-reform levels and increase by an additional average annual rate of 1.5 percent, and that feed coefficients decrease at an average annual rate of 3 percent.

The low productivity growth scenario is based on the general assumption that over the forecasting period, Russian agricultural productivity performance does not improve much over that of the 1990s. Osborne and Trueblood (2002) estimate that over 1993-98, total factor productivity (TFP) in Russian crop production declined by 7 percent. Voigt and Uvarovsky (2001) calculate that over this same period, Russian TFP for all agriculture (crops and livestock) decreased by 15 percent. On the other hand, Lerman et al. (2001) computes that TFP in Russian agriculture (crops and livestock) rose over 1992-97 by 7 percent. Although Lerman et al. differs from Voigt and Uvarovsky and Osborne and Trueblood in finding that productivity rose somewhat rather than fell, all the studies show that Russian agricultural productivity performance during the 1990s was disappointing (particularly relative to expectations at the beginning of reform).

We will examine only briefly the current status and problems of Russian agricultural producers, in order to justify our productivity assumptions. The three main types of agricultural producers in Russia are private farms, the former state and collective farms, and household plots.[7] As of 2002, about 280,000 or so private farms existed in Russia. The farms average about 60 hectares in size, account for less than 5 percent of all Russian farmland, and even less of total agricultural output. Private farms face major challenges in establishing stable and cost-effective upstream and downstream linkages (obtaining inputs and marketing output). Evidence of the severe challenges facing such farms is that since 1996 they have stopped growing (in terms of number of farms and share in arable land and output).

The dominant agricultural producers (at least in an institutional sense) continue to be the former state and collective farms inherited from the Soviet period. Although most of these farms have officially reorganized as joint stock companies owned by their workers and managers, they have done little to change their systems of organization and management. Although output on these farms has fallen substantially, virtually none have stopped operating.

More than half of all agricultural output (in particular livestock products, potatoes, and vegetables) is produced on the household plots tended by workers on the large farms. The plots average only about half a hectare in size, and officially comprise less than five percent of all farmland. The plotholders, however, are able to obtain many inputs (fertilizer, fuel, animal feed, and land for grazing) from their parent farms at little or no cost. This relationship helps explain the plots' high productivity, as well as the low productivity of the parent farms. It is hard to imagine such plots becoming the basis for a technologically and commercially modern and dynamic agricultural system.

The high transaction costs discussed earlier that result from poor commercial and institutional infrastructure also hurt the overall productivity of the agro-food economy. In addition to systems of market information and commercial law, farms especially need a financial system that allows fast, affordable access to capital. Private farms are particularly dependent on such infrastructure, and its slow development helps explain the limited growth and success of this form of production.

Certain developments in Russian agriculture since 2000 offer some basis for optimism that productivity performance could substantially improve. New, vertically integrated producers (Rylko 2001) are emerging in the agriculture and food sector, with finance and management often coming from outside the sector. The new operators could stimulate productivity growth by improving both the technology of the country's production and its system of organization and management. On the other hand, the new producers might simply represent the best possible management and production practices within the economy's existing technology and administrative system, with any productivity gains coming mainly from strengthening vertical ties for production and distribution of output, rather than from real technological or systemic change.

In 2001, the Russian Duma passed legislation reforming the tax and judicial systems, which could simplify the working environment for Russian businesses. In June 2002, the Duma also passed legislation that clarified and sanctioned at the federal level property rights in agricultural land and the market sale of land (though with some qualifications, such as the requirement that purchased agricultural land must be farmed, and a ban on ownership of agricultural land by foreigners). Prior to this legislation, a bewildering mass of laws at both the federal and regional levels "governed" land affairs. Our high productivity growth scenario is based on the assumption that these recent developments involving institutional and farm level changes move agriculture to a qualitatively higher level of performance.

Policy

The two areas of state policy that could heavily impact future Russian agricultural production and trade are market access (involving such policies as tariffs, quotas, and tariff rate quotas) and subsidies that support production and exports. Russia is currently negotiating entry into the World Trade Organization (WTO),[8] which means that the terms of its accession could affect and constrain its future agricultural production and trade policies.

With respect to each of the three main "pillars" of the Uruguay Round Agreement on Agriculture—market access, export subsidies, and domestic support—Russia is arguing that its current policies are fairly moderate. It is therefore negotiating for "bound" (maximum allowable) limits on agricultural import tariffs, export subsidies, and support to producers that are above current levels.[9] Russia's tariffs for most agricultural imports currently range from 10 to 20 percent. In its WTO accession negotiations, Russia is asking for an initial average bound tariff of 35 percent, to fall over six years to an average of 25 percent. In comparison, the average bound tariff on agricultural products for WTO members exceeds 60 percent, while the average bound tariff for Japan, EU countries, and the United States equal 58, 30, and 12 percent, respectively (Gibson et al. 2001).

Before 2003, the only agricultural product on which Russia imposed an import quota was sugar. In spring 2003, however, the Russian government created tariff rate quotas (TRQs) for imports of beef and pork, and for poultry imports a pure quota, to be maintained for a minimum of three years. The annual quota for poultry will be 1.05 mmt, for beef 0.42 mmt, and for pork 0.45 mmt. These figures compare to Russia's 2001 poultry, beef, and pork imports of 1.44, 0.67, and 0.55 mmt. For poultry no imports will be allowed above the quota (with quota imports assessed the current 25 percent tariff), while for beef and pork the tariff for above-quota imports will rise from the existing 15 percent to 60 and 80 percent, respectively. 90 percent of the quota shares for beef and pork, as well as all the quota shares for poultry, will be distributed by country according to their shares in meat exports to Russia during the years 2000-2002. The remaining 10 percent of the quota shares for beef and pork will be auctioned off.

During the transition period, Russia has not used any export subsidies for agricultural and food products. In

its negotiations, Russia is proposing that its bound export subsidies be based on levels over the period 1990-1992, which covers the last two years of the Soviet regime (1990-91). It is asking for bound annual export subsidies of $726 million, which would then drop over six years to $465 million a year.

Countries negotiating WTO accession are expected to ground their bound level of domestic support to agriculture on a base period, typically the three most recent years of available data. Over 1998-2000, Russia's total budgetary transfers to agriculture, from federal and regional governments combined, annually averaged 38 billion rubles, or $1.9 billion (Russian Federation State Committee for Statistics (a) 2001, p. 530). This value equaled 11 percent of agriculture's GDP, and less than one percent of total GDP. Russia argues it is unfair to require it to base its bound support on recent support levels. During the Soviet period, agricultural support was high, with Soviet budget subsidies to agriculture in 1990 equaling about 11 percent of GDP (World Bank 1992, p.138). Russia received a share in these subsidies generally equal to its 47 percent share in total Soviet agricultural output. Russian support to agriculture has fallen steadily during the transition period, mainly because of diminishing state finances rather than the desire from a policy perspective to shrink subsidies.

Russia is therefore asking to base its support on the period 1991-93. In 2001 it proposed bound annual support of $16.2 billion, to fall over a six-year implementation period to $12.9 billion. In 2002 Russia lowered its proposed annual bound support level to $9 billion. WTO members regard even this reduced level as excessive and unwarranted. They point out that it would be more than four times the 2000 support figure, more than half of Russia's agricultural GDP, and about 2.5 percent of total GDP (GDP figures from PlanEcon).

What are the most likely changes in Russia's grain and meat production and trade policies over the forecasting period which could be incorporated into the forecasting scenarios? As mentioned before, in 2003 Russia imposed TRQs on its beef and pork imports and a pure quota for poultry, to be maintained for a minimum of three years. (Throughout the rest of this paper, the phrase "meat import TRQs" for Russia will cover both the TRQs created in 2003 for imports of beef and pork *and* the pure quota created for imports of poultry.) We run two scenarios involving meat import TRQs, in both cases with the assumption that the TRQs remain throughout the entire forecasting period, at the specific terms created in 2003. The first scenario combines the TRQs with our low productivity growth assumptions, and the second scenario combines the TRQs with our high productivity growth assumptions.

If Russia succeeds in negotiating export subsidies, it would most likely use them for grain. Nonetheless, we believe it is unlikely that over the forecasting period Russia would in fact enact grain export subsidies. First, given the strong opposition by the United States and Cairns countries to export subsidies in general, Russia would face particular difficulty getting export subsidies approved as part of their WTO accession terms. Second, even if Russia negotiates this right, it is more likely to use its limited budgetary funds to support producers directly rather than to subsidize exports, which would have the disadvantageous effect of worsening the country's terms of trade. Third, Russian livestock producers would lobby strongly against grain export subsidies (especially of feed grains) while meat imports remain high. We therefore do not run any scenarios involving the use of export subsidies.

As mentioned before, Russia is pushing for bound domestic support high above current levels. Even if Russia were allowed higher bound domestic support, would it in fact raise its agricultural support substantially over the forecasting period? We assume in all our scenario runs that GDP rises over the forecasting period at an average annual rate of 4 percent. Such growth would increase state budgetary revenue, which could fund rising subsidies to agriculture. Yet, although the agricultural establishment lobbies strongly for more support, there are many other claimants on the public purse (in particular the extremely underfunded health, education, police and judicial, and welfare systems). Given the uncertainty about the future of state support to agriculture, and our interest in limiting our scenario runs to a manageable number, we do not run any scenarios with the assumption that domestic support to agriculture increases over the forecasting period. Thus, the only policy change we make in our scenario runs is that which Russia had already made as of early 2003—the imposition of meat import TRQs.

In summary, we make forecasts for four different scenarios:

1. Low productivity growth and no meat import TRQs or other production and trade policy changes.

2. High productivity growth and no meat import TRQs or other production and trade policy changes.

3. Low productivity growth and meat import TRQs over the forecasting period.

4. High productivity growth and meat import TRQs over the forecasting period.

Forecast Results

Scenario #1: Low Productivity Growth and no Meat Import TRQs or Other Policy Changes

Total grain and meat production rise over the forecasting period in the aggregate by 13 and 15 percent, respectively (tables 1 and 2).[10] (Unless indicated otherwise, all forecast figures are for 2012, and percent changes give the change from the base period to the outyear. The base period for grain is 1999-2001 (average annual values), while the base period for the meats is 2001.) A major change is the growth in meat consumption, which rises by 28 percent. Average annual GDP growth of 4 percent substantially increases consumer income, which because of the relatively high income elasticity of demand for meat, significantly raises domestic meat demand. With meat consumption growing in percent terms at almost twice that of production, meat imports surge, increasing by 48 percent. Poultry imports expand to 1.91 mmt.

The meat import forecasts, however, are upper bound projections. The reason why involves the question of the degree of Russia's integration into world agricultural markets. The model forecasts that most of the growth in consumer demand for meat is satisfied by imports. Implicit in this result is the assumption that Russian domestic meat markets are well integrated into world markets, such that world trade prices largely determine domestic producer and consumer prices for meat. Because an increase in demand does not raise Russian domestic producer prices, Russian meat production does not expand by much in response to the growth in demand. Thus, imports must satisfy most of the demand growth.

As discussed earlier, the variables within the model that capture the degree of integration into world agricultural markets are the price and exchange rate transmission elasticities. Our assumption is that the TEs for all commodities equal 0.5 at the beginning of the forecasting period, and rise to 0.75 by the outyear 2012. Thus, even by the outyear, integration into world markets, as represented by the TEs, is less than that implied by the large increase in Russian meat imports. The technical reason behind this problem is that the TEs come into effect only if world prices *change*. In our forecasting scenario, however, the income-driven surge in consumer demand for meat occurs without any change in meat border prices. A conflict therefore exists between the degree of world market integration as implied by the TE values, and the high degree of world market integration as implied by the model's mechanics, which particularly come into play when domestic demand and supply curves shift without any change in border or domestic prices. Because Russia over the forecasting period will probably remain imperfectly integrated into world markets, growth in meat demand could spark an increase in domestic output as well as imports. The consequence for our meat import forecasts is that they are upwardly biased.

Given that meat production rises only moderately, demand for (and thus consumption of) feed grain also increases modestly. The growth in grain production is such that Russia moves from net grain imports of 1.2 mmt to net exports of 3.7 mmt. (Though Russia imported grain during 1999-2001, it was mainly to build up stocks following the disastrous 1998 harvest of 46 mmt. As table 1 shows, during 1999-2001 production in fact exceeded consumption.)

Scenario #2: High Productivity Growth and no Meat Import TRQs or Other Policy Changes

Higher productivity growth increases output of total grains and meat over the forecasting period by 33 and 25 percent, respectively (table 3). Output of wheat and coarse grains rises by 39 and 26 percent, while production of beef, pork, and poultry grows by 15, 28, and 41 percent, respectively. The ranking of the three types of meat in terms of percent growth reflects the relative degree of feedstuffs (as opposed to pasture and fodder) in animals' total diet: poultry and beef are the most and least feedstuff-intensive, and pork is intermediate.

The country becomes a major net exporter of grain of about 20 mmt in the outyear, for two reasons. The first is productivity growth in grain production, and the second is that improved feed efficiency (the variable in the model that captures productivity growth in meat production) reduces grain consumption. Russia thereby becomes a major grain exporter of both wheat and coarse grains, with wheat exports rising to 11.8 mmt. The growth in Russian wheat exports from 2.1 mmt in the low productivity scenario to 11.8 in the high productivity scenario has the isolated effect of lowering world wheat prices in the outyear by 5 percent.

The rise in meat production lowers meat imports (compared to the low productivity growth scenario). Yet, with net imports of 3.6 mmt, Russia remains a big meat importer.

Scenario #3: Low Productivity Growth and Meat Import TRQs

In this scenario, we impose the TRQs on meat imports enacted by Russia in spring 2003. The TRQs increase domestic producer and consumer prices for the meats by 24 and 29 percent, respectively (the aggregate price increases being weighted averages), thereby raising production and reducing consumption (table 4). Compared to the low productivity growth alone scenario, meat output in the outyear is 12 percent higher and consumption 13 percent lower. Beef, pork, and poultry production rises by 6, 11, and 26 percent, respectively (compared to the low productivity alone scenario). The relative output growth of the three types of meat reflects the ranking of their price elasticities of supply: high and low for poultry and beef, and intermediate for pork. Meat producers benefit at the expense of meat consumers, who suffer from both higher prices and lower consumption. Among the three types of meat, the domestic consumer price for poultry rises the most, and consumption falls the most, by 22 percent (all compared to the low productivity alone scenario). The isolated effect of the TRQs (for poultry pure quota) on world prices for poultry, pork and beef is a decline of 7.5, 7.3, and 4.7 percent, respectively.[11]

The growth in meat output increases domestic demand for feed grain, which raises grain consumption 7 percent (compared again to the low productivity growth alone scenario). Increased demand for feed grain results in net grain imports in the outyear of 1.2 mmt. Large reduction of meat imports through TRQs therefore results in the country becoming a small grain importer.

Scenario #4: High Productivity Growth and Meat Import TRQs

Output of both grain and meat is stimulated by rising productivity, while meat production receives an additional boost from import TRQs (table 5). Grain and meat output increase over the forecasting period by 33 and 36 percent, respectively. Beef, pork, and poultry production rises by 20, 39, and 73 percent, respectively (compared to the base period). Poultry production expands by a much greater percentage than beef and pork both because it is the most intensive in the use of feedstuffs (and therefore benefits the most from our assumption about improved feed efficiency), and because it has the highest price elasticity of supply (and therefore responds the most to the domestic price increase resulting from the import quota/TRQs). Grain consumption is higher than in the scenario involving only high productivity growth, because the meat import TRQs stimulate meat production, and thereby demand for feed grain.

Russia is a major net grain exporter of about 16.5 mmt. As opposed to the scenario involving low productivity growth and meat import TRQs where Russia becomes a small grain importer, with higher productivity growth, production is sufficiently stimulated to produce large grain exports. Yet, grain exports are lower than in the scenario involving high productivity growth alone. The stimulus to meat production from the meat import TRQs increases domestic demand for feed grain, thereby reducing grain exports.

Conclusion

Using a commodity forecasting model for Russian agriculture created at the Economic Research Service of USDA, we make projections for Russian grain and meat production, consumption, and trade for the year 2012. Assumptions are made for the following key variables: (1) GDP; (2) the real exchange rate; (3) price and exchange rate transmission elasticities; (4) productivity (yields and feed coefficients); and (5) meat import TRQs. Forecasts are presented for four scenarios: (1) low productivity growth and no change in policies affecting production and trade; (2) high productivity growth and no change in policies; (3) low productivity growth and imposition of TRQs for meat imports; and (4) high productivity growth and meat imports TRQs.

Depending on how productivity and policies change over time, Russia could become either a major or minor grain exporter, or a small importer. In the scenario involving low productivity growth and no meat import TRQs or other policy changes, the stimulus to grain production is enough to make Russia a modest 2012 net grain exporter of 3.7 mmt. However, in the scenario involving high productivity growth and no meat import TRQs, the increase in grain production turns Russia into a large grain exporter (19.6 mmt). In the scenario involving high productivity growth and meat import TRQs, Russia remains a big grain exporter (16.6 mmt). Productivity growth raises grain output sufficiently to create surpluses for export, though export volumes are lower than in the second scenario (high productivity growth alone). The reason is that the import TRQs stimulate meat production, which by increasing domestic demand for feed grain cuts into the exportable grain surplus. In the scenario involving low productivity growth and meat import TRQs, however, Russia becomes a small grain importer (1.2 mmt). The substantial rise in meat production from the TRQ

protection increases consumption of feed grain beyond the country's level of production.

Two main conclusions concerning Russia's grain trade follow from the scenarios. The first is that high productivity growth would generate large surpluses for export regardless of whether meat import TRQs exist throughout the forecasting period. The second is that meat TRQs without productivity growth will make the country a small grain importer, as meat imports are reduced at the expense of rising inflows of feed grain.

A major conclusion concerning the meat sector is that, in the absence of meat import TRQs or any other major policy intervention, Russia would likely remain a big importer of meat—beef, pork, and especially poultry. In the low productivity growth scenario, meat imports by 2012 equal 3.9 mmt. In the high productivity scenario, meat imports by 2012 are lower, but not substantially so, at 3.6 mmt. In spring 2003, however, Russia imposed TRQS on its meat imports, with the low tariff quota volumes summing to 1.92 mmt, compared to Russia's total meat imports in 2001 of 2.65 mmt. The TRQs will expand meat production at the expense of meat consumers, who will face reduced total supplies at higher prices. The meat TRQs therefore could pit the interests of meat consumers against producers.

Endnotes

[1] William Liefert is a senior agricultural economist, Stefan Osborne and Ralph Seeley agricultural economists, and Olga Liefert a consultant, all with the Market and Trade Economics Division of the Economic Research Service, U.S. Dept. of Agriculture. Eugenia Serova is President of the Analytical Centre "Agrifood Economics" at the Russian Institute for Economy in Transition. The authors thank Ed Allen, Cheryl Christensen, John Dunmore, James Hansen, Gregory Pompelli, Randy Schnepf, David Skully, James Stout, and Michael Trueblood for helpful comments. The authors bear responsibility for any remaining deficiencies. The views expressed are the authors' alone and do not in any way represent official USDA views or policies.

[2] For example, Tyers predicted that by 2000 Russia could be exporting about 40 million metric tons of grain a year. Tyers provides forecasts for Russia, Ukraine, and other countries and regions of the former USSR, while Koopman's and Liefert et al.'s forecasts are for only the former Soviet Union as a whole. Although he does not use a forecasting model in his analysis, Johnson (1993) also argues that successful reform that reduces waste and raises productivity could transform the former Soviet Union into a major grain exporter.

[3] Liefert and Swinnen (2002) examine why the studies misforecast NIS agricultural production and trade during reform, the key point being that they underestimated the severity and duration of the decline in agricultural output during the transition. In fairness to the forecasting studies, though, their projections were based to a large degree on the assumption that Russia and the other countries of the former USSR would adopt more ambitious agricultural reform programs than they in fact have implemented.

[4] Liefert (2002) finds that in the late 1990s, Russia had a comparative disadvantage in the production of meat relative to grain, such that the substitution of meat imports for feed imports during the transition period appears rational. For an analysis of how reform has changed agricultural production, consumption, and trade in the transition economies, see Liefert and Swinnen (2002). Cochrane et al. (2002) examines the restructuring specifically of the livestock sector in the countries of the former Soviet bloc during transition.

[5] In this report, all annual production and trade figures reported for grain are in marketing year (July-June). Also, figures for grain production and trade do not include rice, buckwheat, or pulses.

[6] Cochrane et al. (2002) uses models created for the livestock sectors of Russia, Ukraine, Poland, Hungary, and Romania to examine the effects of changes in various policies, resource endowments, and factor prices on these countries' livestock production and trade.

[7] Liefert (2001) provides a review of the current status of each of these three types of producers.

[8] Although Russia officially began its bid for WTO membership in 1993, progress in its accession negotiations up to 2001 had been slow. Recent developments, however, have motivated both Russia and WTO members to quicken the pace of negotiations. China's accession in 2001 left Russia as the largest remaining nonmember (in both GDP and population), while geopolitical developments in 2001 increased interest in both Russia and the West for integrating the country more strongly into international political and economic institutions.

[9] Russia's "current" negotiating positions as identified in this article are those given in its English-language website *Russia and WTO* (Russian Federation). According to the website, Russia's official positions on

agricultural issues were last identified in a proposal submitted to the WTO Secretariat in March 2001. Since that time, the press has reported possible changes in some of Russia's negotiating stances. Yet, none of these changes (even if made) would involve major shifts in position, with the one major exception that in October 2002 the Russians lowered their figure for annual bound domestic support from $16.2 billion to $9 billion.

[10] In tables 1-5, the commodity category *Grains* is the sum of wheat and coarse grains (thereby excluding rice). *Meats* is the sum of beef, pork, and poultry (thereby excluding mutton). In 2001, mutton production equaled only 4 percent of total output of beef, pork, and poultry, and little mutton was traded.

[11] The drop in world meat prices shows that Russia has market power in world meat trade. In 2001, Russia accounted for 24, 16, and 12 percent of world imports of poultry, pork, and beef (USDA), respectively.

References

Cochrane, Nancy, Britta Bjornlund, Mildred Haley, Roger Hoskin, Olga Liefert, and Philip Paarlberg. *Livestock Sectors in the Economies of Eastern Europe and the Former Soviet Union: Transition from Plan to Market and the Road Ahead*. Economic Research Service, U.S. Dept. of Agriculture, Agricultural Economic Report No. 798, February 2002.

Gibson, Paul, John Wainio, Daniel Whitley, and Mary Bohman. *Profiles of Tariffs in Global Agricultural Markets*. Economic Research Service, U.S. Dept. of Agriculture, Agricultural Economic Report No. 796, January 2001.

Hjort, Kim, and Pierre van Peteghem. *The CPPA Model-Builder: Technical Structure and Programmed Options in Version 1.3*. Economic Research Service, U.S. Dept. of Agriculture, Staff Report No. AGES 9144, 1991.

Interfax. *Food and Agriculture Report*. Moscow, weekly.

Johnson, D. Gale. "Trade Effects of Dismantling the Socialized Agriculture of the Former Soviet Union." *Comparative Economic Studies* 35, 4:21-33, Winter 1993.

Koopman, Robert. "Agriculture's Role During the Transition from Plan to Market: Real Prices, Real Incentives, and Potential Equilibrium." In *Economic Statistics for Economies in Transition: Eastern Europe in the 1990s*, U.S. Dept. of Labor, Washington DC, 1991.

Lerman, Zvi, Yoav Kislev, Alon Kriss, and David Biton. "Agricultural Output and Productivity in the Former Soviet Republics." *Economic Development and Cultural Change* 51, 4:999-1018, July 2003.

Liefert, William. "Agricultural Reform: Major Commodity Restructuring but Little Institutional Change." In *Russia's Uncertain Economic Future*, Joint Economic Committee, U.S. Congress, Washington, DC, December, 2001.

Liefert, William. "Comparative (Dis?)Advantage in Russian Agriculture." *American Journal of Agricultural Economics* 84, 3:762-767, August 2002.

Liefert, William, Robert Koopman, and Edward Cook. "Agricultural Reform in the Former USSR." *Comparative Economic Studies* 35, 4:49-68, Winter 1993.

Liefert, William, and Johan Swinnen. *Changes in Agricultural Markets in Transition Economies*. Economic Research Service, U.S. Dept. of Agriculture, Agricultural Economic Report No. 806, February 2002.

Osborne, Stefan, and William Liefert. "Price and Exchange Rate Transmission in Russian Meat Markets." *Comparative Economic Studies* (forthcoming).

Osborne, Stefan, and Michael Trueblood. *Agricultural Productivity and Efficiency in Russia and Ukraine: Building on a Decade of Reform.* Economic Research Service, U.S. Dept. of Agriculture, Agricultural Economic Report No. 813, July 2002.

Oxford Economics. Subscription service at www.oef.com.

PlanEcon. *Review and Outlook for the Former Soviet Republics*. Washington, DC, biannual.

Russian Federation. *Russia and World Trade Organization*. www.wto.ru/russia.asp.

Russian Federation State Committee for Statistics (a). *Rossiiskii Statisticheskii Ezhegodnik* (*Russian Statistical Yearbook*). Moscow, annual.

Russian Federation State Committee for Statistics (b). *Tseni v Rossii (Prices in Russia)*. Moscow, 1996, 1998, and 2000.

Rylko, Dmitri. "New Agricultural Operators, Input Markets, and Vertical Sector Coordination." Presented at Workshop on Russian Agricultural Input Markets, Golitsino, Russia, July 2001.

Tyers, Rod. *Economic Reform in Europe and the Former Soviet Union: Implications for International Food Markets*. Washington, DC: International Food Policy Research Institute, Research Report 99, 1994.

U.S. Dept. of Agriculture (USDA). *Production, Supply, and Distribution Database (PS&D).* www.ers.usda.gov/Data/PSD/.

Voigt, Peter, and Vladimir Uvarovsky. "Developments in Productivity and Efficiency in Russia's Agriculture: The Transition Period." *Quarterly Journal of International Agriculture* 40, pp. 45-66, 2001.

Wehrheim, Peter, Klaus Frohberg, Eugenia Serova, and Joachim von Braun, editors. *Russia's Agro-food Sector: Towards Truly Functioning Markets*. Dordrecht, Netherlands: Kluwer Academic Publishers, 2000.

World Bank. *Food and Agricultural Policy Reforms in the Former USSR: An Agenda for the Transition*. Studies of Economies in Transition, Paper No. 1, Washington, DC, 1992.

Table 1: Russian Agriculture: Grain and Meat Production, Consumption and Trade, Base Period

	Production	Consumption	Trade Balance*
	(million tons)		
Grains**	65.83	63.19	-1.20
Wheat	37.45	36.17	-0.82
Coarse grains	28.38	27.02	-0.37
Meat***	4.00	6.65	-2.65
Beef	1.77	2.44	-0.67
Pork	1.53	2.08	-0.55
Poultry	0.70	2.14	-1.44

*Positive (negative) numbers are net exports (imports).
**Grain numbers are annual averages over marketing year (July-June) 1999-2001. Because of stocks, net trade for grain in this table does not equal the difference between production and consumption.
***Meat numbers are for calendar year 2001.
Source: USDA.

Table 2: Russian Agriculture: Forecast with Low Productivity Growth

	Production	Consumption	Trade Balance*
	(million tons)		
Grains	74.10	70.41	3.69
Wheat	43.30	41.17	2.13
Coarse grains	30.80	29.24	1.56
Meat	4.60	8.51	-3.91
Beef	1.97	3.00	-1.02
Pork	1.76	2.74	-0.98
Poultry	0.87	2.77	-1.91

*Positive (negative) numbers are net exports (imports).
Source: Authors' calculations.

Table 3: Russian Agriculture: Forecast with High Productivity Growth

	Production	Consumption	Trade Balance*
	(million tons)		
Grains	87.76	68.12	19.64
Wheat	51.93	40.11	11.84
Coarse grains	35.83	28.01	7.80
Meat	4.98	8.61	-3.63
Beef	2.03	3.00	-0.97
Pork	1.96	2.80	-0.84
Poultry	0.99	2.81	-1.82

*Positive (negative) numbers are net exports (imports).
Source: Authors' calculations.

Table 4: Russian Agriculture: Forecast with Low Productivity Growth and Meat Import TRQs

	Production	Consumption	Trade Balance*
	(million tons)		
Grains	74.02	75.17	-1.16
Wheat	43.13	43.25	-0.12
Coarse grains	30.89	31.92	-1.04
Meat	5.15	7.38	-2.22
Beef	2.09	2.81	-0.71
Pork	1.96	2.42	-0.46
Poultry	1.10	2.15	-1.05

*Positive (negative) numbers are net exports (imports).
Source: Authors' calculations.

Table 5: Russian Agriculture: Forecast with High Productivity Growth and Meat Import TRQs

	Production	Consumption	Trade Balance*
	(million tons)		
Grains	87.65	71.02	16.63
Wheat	51.78	41.40	10.40
Coarse grains	35.87	29.62	6.23
Meat	5.45	7.63	-2.18
Beef	2.12	2.80	-0.67
Pork	2.12	2.57	-0.46
Poultry	1.21	2.26	-1.05

*Positive (negative) numbers are net exports (imports).
Source: Authors' calculations.

Social Programs and Forecasting Measures of Well-Being

Chair: Karen S. Hamrick, Economic Research Service, U.S. Department of Agriculture

Forecasting Nonmetro Use of Food Stamps with a GDP Per Capita Forecast and a Model of Wage and Salary Income Distribution

John Angle, Economic Research Service, U.S. Department of Agriculture

A parsimonious model is fit to the conditional distribution, wage and salary income, nonmetro and metro, by five levels of education for 1961-2001. This model takes the national mean of wage and salary income as an exogenous variable. This mean can be estimated under the model from the conditional medians. However, to forecast income distribution this relationship is inverted and the close linear relationship between mean and GDP per capita is used to forecast the mean. Food stamp usage is closely related to the left tail of the wage and salary income distribution.

Price and Regional Economic Convergence

Qingshu Xie, MacroSys Research and Technology

In economic convergence analysis, regional price variation is often ignored mainly because of the absence of regional price deflators. There is a debate over whether regional price differences have an impact on regional economic convergence. One argument suggests that prices greatly impact measurement of poverty and the standards of living. This paper analyzes the patterns of interstate income convergence 1963-2000 based on the data deflated with the national and implicit state-level price deflators, respectively. Using various inequality measures, the results indicate that regional price variation has a great impact on the patterns of regional convergence analysis. This reflects a need for the development of explicit regional price deflators.

Analyzing the Demand for Non-Alcoholic Beverages

Annette Clauson, Economic Research Service, U.S. Department of Agriculture

Because USDA is the lead Federal agency for nutritional information and programs of nutritional benefits for children and low-income households, ERS organized a team of experts to address the complex issues of why milk and more nutritious juices have been displaced by other non-alcoholic beverages. This presentation will outline the steps necessary to mine the raw database (the 1999 A.C. Nielsen Homescan Panel), select the tools for working with patterns in the data, and choose the decision trees and econometric methods for obtaining information on the drivers of demand for milk and other non-alcoholic beverage products. The findings should result in developing policies for analyzing, forecasting, and setting priorities in food programs.

Disability Benefit Applications, Economic Trends, and Projections

Mikki D. Waid and Frederick L. Joutz
Department of Economics, The George Washington University

The current Social Security program solvency issues suggest revisiting the topic of disability applications. Previous research suggested that the number of applications is affected by macroeconomic variables such as the unemployment rate as well as supply side changes such as workforce overall health. Modeling the number of applications will aid in determining future program costs and will assist in determining how long the program will be solvent. This paper presents results from an experimental model for the number of annual applications. Annual data from 1962 to 1997 are used to analyze the factors contributing to the growth of applications. The model's with-in-sample and out-of-sample forecasting performance is evaluated.

PRICE AND REGIONAL ECONOMIC CONVERGENCE

Qingshu Xie, MacroSys Research and Technology

Introduction

Economic convergence has widely been studied in economic research since the mid-1980s. It includes the convergence of per capita income levels and economic growth rates across economies. The decline of per capita income dispersion is referred to as σ (sigma) convergence; and the convergence of economic growth rates as β (beta) convergence (Sala-i-Martin, 1990; Barro and Sala-i-Martine, 1991). Although there has been a sizable amount of literature on economic convergence at both the international and interregional levels, researchers hardly reach a consensus on whether economies are converging or not and empirical studies often lead to controversy. Temple (1999) shows that the controversy in across-country convergence studies is attributed to the quality of output data, the measurement approach of growth rates, and several common econometric problems such as omitted variable, parameter heterogeneity, outliers, model uncertainty, and endogeneity. The studies of regional economic convergence are not immune to these problems either. Moreover, they are affected by other factors such as geographic scale and regional price differences which are often not fully explored in empirical studies.

It is well known that there exist significant variations in cost of living across regions because of regional price differences (e.g., McMahon and Melton, 1978; McMahon, 1991). In practice, American Chamber of Commerce Research Association publishes estimates of cost of living for some cities in the United States that reflect geographic differences in cost of living. However, there are not a series of regional price indexes that include both temporal inflation and spatial price differences. Therefore, researchers often have to ignore regional price differences in regional economic convergence.

Do regional price differences have an impact on regional economic convergence? If any, but how much is the impact? The study of these issues has important policy implication. It helps to understand the actual trend of regional economic convergence and thus provide useful information for policy makers on regional economic development policy. Using state-level price deflators, this paper aims to examine whether regional price differences have an impact on interstate income convergence in the United States. The hypothesis is that regional price differences affect the pattern and degree of interstate income convergence in the United States. Following the introduction, previous studies on the impact of regional price dispersion on economic convergence is briefly reviewed in the second section. Data and methodology are discussed in the third section. The findings of the analysis are presented in the fourth section. Conclusions are provided in the final section.

Previous Studies

There are two opposite arguments on whether regional price differences have an impact on the result of regional economic convergence analysis. On the one hand, Sala-i-Martin (1996) argues that regional price dispersion is not likely to affect the pattern of convergence. On the other hand, several studies suggest that regional price differences have a significant impact on regional economic convergence in the United States (Black and Dowd, 1997; Deller, Shields, and Tomberlin, 1996; Slesnick, 2002).

Using national price deflators, Sala-i-Martin (1996) finds that the convergence speeds of Canadian and Japanese regional economies are not different from those of two other corresponding studies using regional price deflators (Coulombe and Lee, 1993; Shioji, 1992), which are all about 2 percent per year. Thus, he argues that interregional price dispersion does not impact the pattern of regional income convergence. However, this argument is not convincing for two reasons. First, the generalization based on only two cases is not sufficient to prove it is true for regional income convergence in other countries. Second, no proof is given on the impact of interregional price dispersion on σ convergence.

Several studies show that regional price differences affect the pattern and degree of regional economic convergence in the United States. Using the national price deflators, many studies show that there is a reversal of interstate inequality in per capita personal income in the United States in the 1980s (e.g., Amos, 1988; Bernat, 2001; Fan and Casetti, 1994). Deller, Shields, and Tomberlin (1996) argue that the divergence of interstate inequality in per capita personal income in the 1980s disappears when not only temporal but also spatial price differentials are taken into account. Conversely, a study by Black and Dowd (1997) suggests that the divergence in per capita personal income was underestimated without eliminating spatial price differences and real regional

income inequality has steadily increased since 1981. Slesnick (2002) finds that the convergence of regional welfare measured with per capita consumption is amplified after regional price differences are accounted for.

Table 1 summarizes the three studies on US regional economic convergence that deflate data with regional price deflators. Deller *et al*. (1996) and Black and Dowd (1997) use the same economic indicator and unit of analysis, but they differ in the use of regional price deflators and thus the conclusions. Notably, it is questionable that both Bureau of Labor Statistics (BLS) regional CPI and state level cost of living indexes are simultaneously used in Deller *et al.* (1996). There is actually double deflation in doing so because BLS regional CPI not only accounts for temporal inflation but also partly accounts for the differences in regional prices for states across four different regions. Slesnick (2002) constructs a series of regional prices and uses per capita consumption as economic indicator to measure economic convergence across four census regions. Despite the differences, there is one thing in common in the three studies: regional price differences have an impact on region economic convergence.

Table 1. Studies on US Regional Economic convergence with Regional Price Indexes

Study	Indicator	Study Period	Unit of Analysis	Regional Price	Conclusion
Deller *et al*. (1996)	Personal income per capita (PCPI)	1969-91	State	BLS CPI for four BLS regions and cost of living index	The divergence of interstate income in the 1980s disappears.
Black and Dowd (1997)	PCPI	1963-89	State	Implicit state price deflators based on BEA GSP data	Interstate income inequality is underestimated without consideration of regional price differences and interstate income inequality has steadily increased since 1981.
Slesnick (2002)	Consumption per capita	1960-2000	Four Census regions plus rural households	Estimated regional price level series	Regional price amplifies the pattern of convergence, has no impact on inequality across households but affect the threshold of poverty rate.

Data ad Methodology

This paper studies the convergence of per capita personal income across 48 contiguous states in the United States in 1963-2000. National and state-level implicit price deflators are used to adjust data for inflation and regional price dispersion. Per capita personal income data and national implicit price deflators are from Bureau of Economic Analysis (BEA). State level implicit price deflators are derived based on BEA's data on gross state product (GSP). The GSP data include two series: GSP for 1963-1986 in current and 1982 constant dollars and GSP for 1977-2000 in current and 1996 constant dollars. The GSP data for 1963-1976 and for 1977-2000 are combined to form a series of GSP for 1963-2000 in current and 1996 constant dollars. As in Black and Dowd (1997), state level price deflators are derived through the division of GSP in current dollars by GSP in constant dollars.

Both σ and β convergence are measured and the impact of regional price differences on the pattern and extent of interstate income convergence is evaluated. Four inequality measures including standard deviation, coefficient of variation (CV), the Gini coefficient, and neighborhood disparity index (NDI) are used in measuring σ convergence. The first three measures are commonly used in convergence analysis. The neighborhood disparity index is created by Chakravorty (1996) for spatial income inequality analysis. The basic idea of NDI is first to measure the differences in an indicator between each observation and the average of its close neighboring counterparts, and then to obtain a summary inequality measure by normalizing the sum of the individual differences with a double product of the number of observations and the average value of the study area. First order contiguity is considered in this analysis. The formula for NDI is:

$$NDI = \frac{\sum_{i=1}^{N} \left| \sum_j y_j / n_j - y_i \right|}{2Ny_a}$$

where, y_i = per capita income for a spatially central state, state i; y_j = per capita income of states neighboring state i; n_j = number of sates contiguously neighboring the central state; N = total number of states, i.e. 48 in this research; and y_a = the average U.S. per capita income. The value of NDI falls between 0 and 1, where 0 suggests a perfect equality of spatial income distribution, and 1 indicates maximum neighborhood disparity.

Cross-section regression is used to measure β convergence. The model for β convergence is:

$$\frac{1}{T} Ln(\frac{Y_t}{Y_o}) = a - bLn(Y_o) + e$$

Where Yt is the per capita income at time t, Y_o is the initial per capita income, and T is the length of study period. A negative coefficient of the initial income level, $Ln(Y_o)$, indicates β convergence, which means poor economies grow faster than rich ones. Convergence speed is computed with the following equality: $b = [(1 - e^{-\beta T}) / T]$. Then, convergence peed is: $\beta = -[Ln(1 - bT)] / T$ (Sala-i-Martin, 1996).

This study is different from previous studies in several aspects. First, the data are extended to better reflect the pattern of regional economic convergence. Second, the impact of regional price differences on both σ and β convergence is measured. Third, multiple inequality measures including a spatial inequality measure are used in measuring the pattern of interstate income convergence. Black and Dowd (1997) use state level price deflators but a single inequality measure which is not commonly used in convergence analysis.

Research Findings

Sigma convergence

Figures 1 – 4 indicate that in general the dispersion in per capita personal income declined for the period 1929-2000 measured with data adjusted with the national price deflators and for the period 1963-2000 measured with data adjusted with state-level price deflators. However, the patterns of σ convergence with state-level price deflators are clearly different from those with the national price deflators. These results suggest that regional price differences do matter in regional economic analysis.

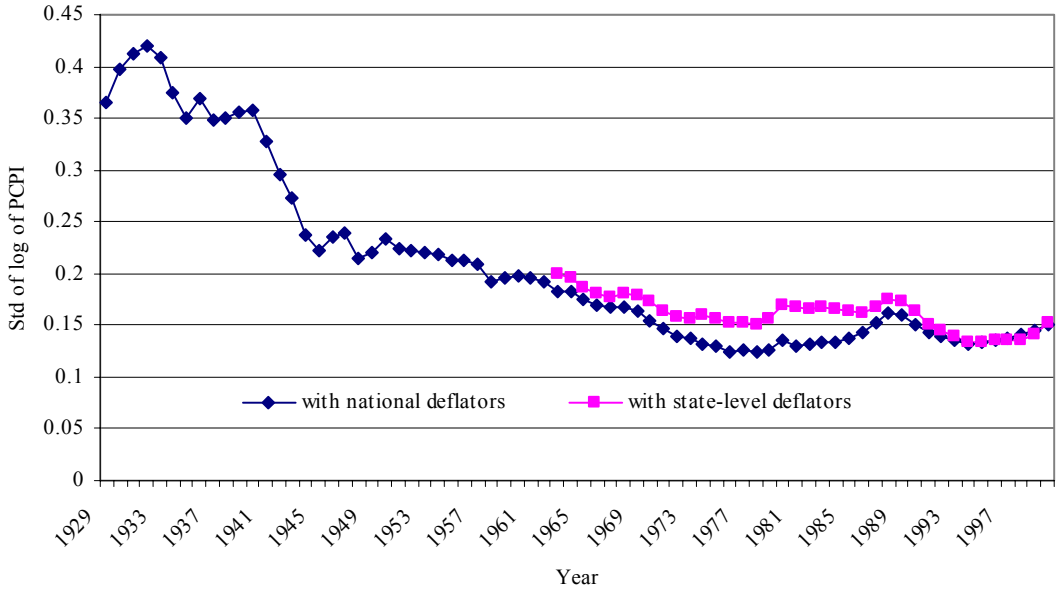


Figure 1. Effect of State-level Deflators on Standard Deviation of Log of PCPI

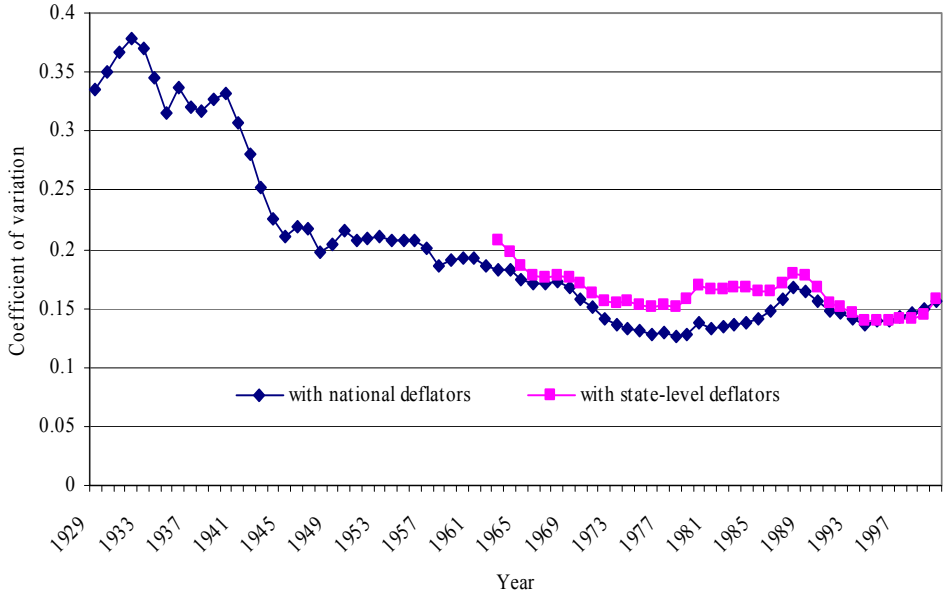


Figure 2. Effect of State-level Deflators on Coefficient of Variation of PCPI

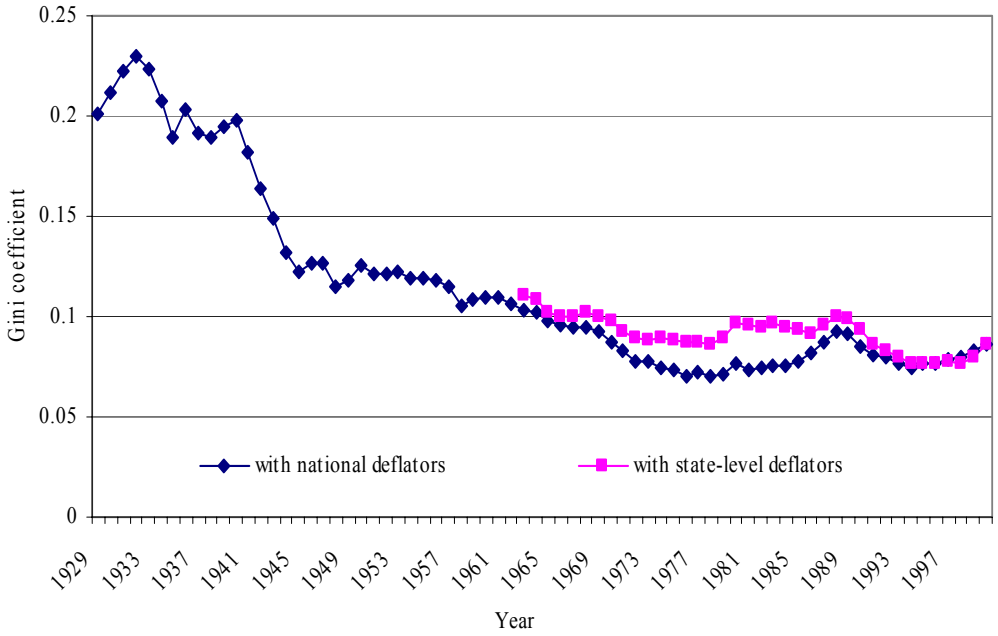


Figure 3. Effect of State-level Deflators on the Gini Coefficient of PCPI

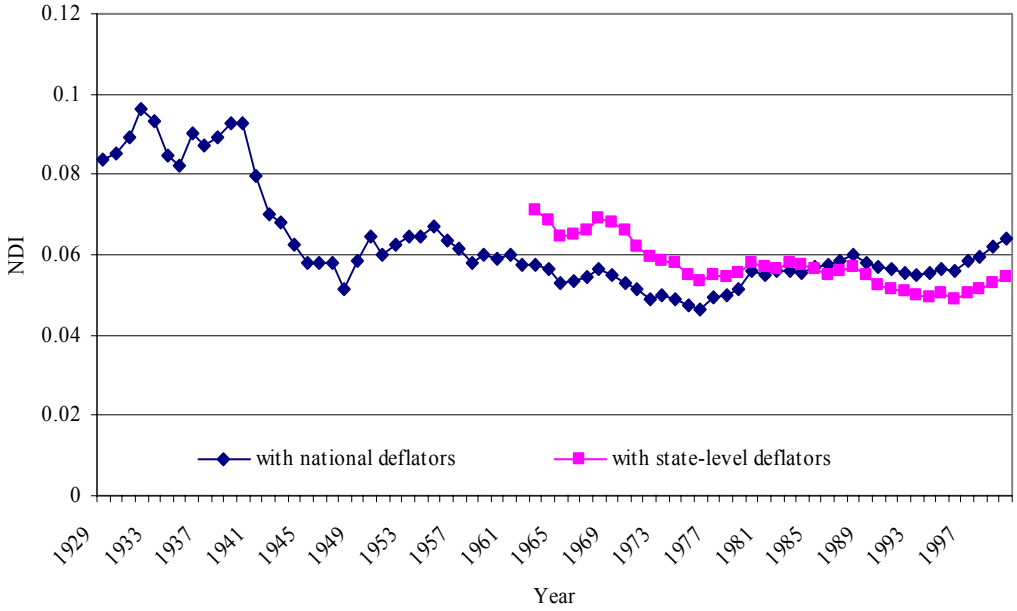


Figure 4. Effect of State-level Deflators on NDI of PCPI

Figure 5 shows the extent of underestimation and overestimation of interstate income inequality without consideration of regional price differences. The effects of state-level deflators on the four inequality measures are different. While the patterns of underestimation and overestimation in standard deviation, CV and the Gini coefficient are very similar, the pattern of underestimation and overestimation in NDI is quite different. In regard to standard deviation, CV and the Gini coefficient, Figure 5 suggests that, without adjustment for spatial price differences, interstate inequality in PCPI is underestimated in the period 1963 – 1994 and slightly overestimated during 1997-1999. The underestimations increased from 3-5 percent in the late 1960s to the maximums 25-30 percent in 1981. It reduced after 1981 but remained at a level of greater than 5 percent until 1991. The underestimation was smaller than 5 percent from 1992-1994. Relative to the underestimation, the overestimation is trivial. The maximum overestimation was in 1998 and was less than 4 percent. As for NDI, it is underestimated in the period before 1985 and overestimated afterwards. The maximum underestimation is in 1970, greater than 25 percent. The extent of underestimation of NDI is far smaller than those of other three measures during the period 1974 – 1985. In terms of NDI, interstate inequality in PCPI is largely overestimated after 1985. The overestimation stably increased from 4 percent in 1986 to 15 percent in 2000.

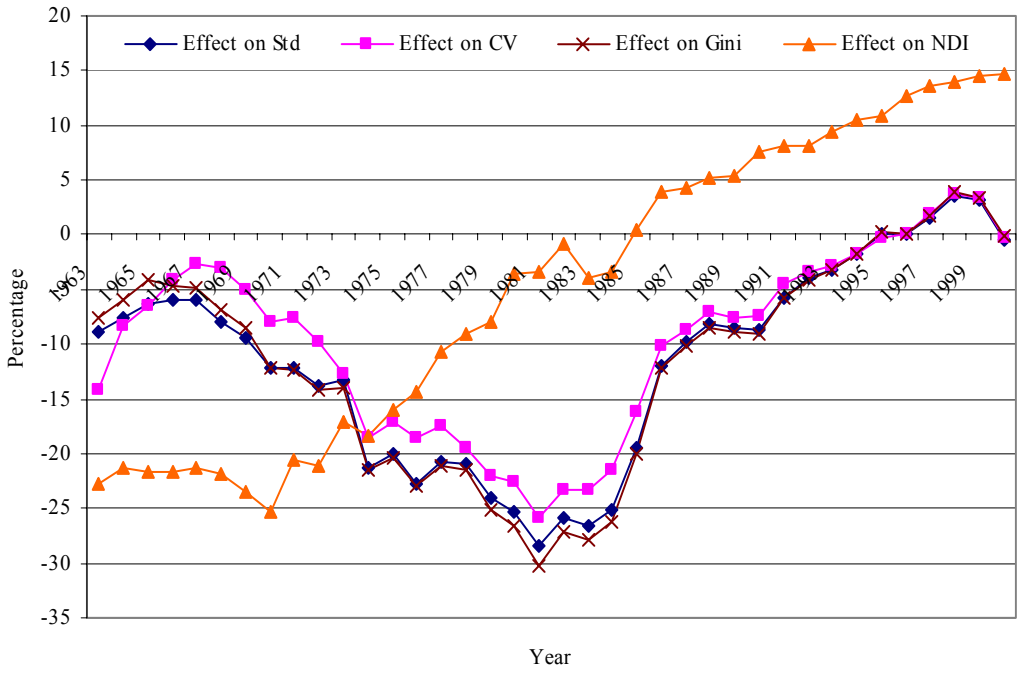


Figure 5. Effect of State-level Deflators on the Measures of Regional Income Inequality

To compare the patterns of σ convergence with different inequality measures, the values of the inequality measures are normalized with their own maximum values, respectively, and the results are presented in Figures 6 and 7. Figure 6 presents the results with the data adjusted with the national price deflators and indicates an overall pattern of convergence in PCPI among the states during 1929-2000. Since the mid-1950s, all four measures reveals a very similar pattern: income inequality in PCPI declined from the mid-1950s to the late 1970s (around 1976-1978), then had a sustained increase through the late 1980s (1988), declined again through the mid-1990s (1994) but remained above the lowest level in the late 1970s, and then rose again up to the end of the study period. The argument on the reversal of regional income convergence at state level seems to be confirmed.

Figure 7 presents the results with the data adjusted with state-level price deflators. Measured with the three conventional measures, interstate inequality in PCPI declined from the late 1960s to 1978, increased from 1979 to 1980, then slightly declined over 1980-86, rose in the next two years (1987-88), reverted to decline during 1989-94, stagnated from 1995 to 1998, and then started to rise afterwards. It is worth noting that the interstate inequality in PCPI in most years of the 1990s was below the lowest level in the 1970s. In comparison with other three measures, NDI shows a sharper decline from 1968 to 1976, relatively small fluctuation from 1976 to 1988, and slower decrease from 1988 to 1994.

Further, the universal pattern of sustained strong reversal of interstate income convergence in the 1980s found earlier with the data adjusted with the national price deflators is not firmly supported with the data adjusted with state-level deflators. Though the interstate income inequality in 1988 is still greater than in 1978, the increase of interstate income inequality is generally smaller than in the data adjusted with national deflators and it is not stable. In contrast to the sustained reversal of regional income convergence shown in Figure 6, there is an overall slight decline of interstate inequality measured with all the four measures during the period 1981 – 1986. Then, there is a sharper increase from 1986 to 1988. This suggests that the argument for sustained regional income divergence in the 1980s is only a pattern that is based on the data without adjustment for spatial price differentials.

In general, interstate income inequality in PCPI declined from the late 1960s to the late-1970s, increased with fluctuation till 1988, decreased again and started to rise in the late-1990s. This pattern seems to generally match the pattern in interstate

income inequality that is based on the data adjusted with the national deflators. Nevertheless, the impact of state-level deflators on interstate income inequality is also partially reflected by the fluctuations of interstate income inequality in the 1980s and by the differences between the pattern of NDI and those of other three measures. The argument for sustained strong regional income divergence in the 1980s is not well supported by the data adjusted for spatial price differentials. Furthermore, interstate inequality in most years of the 1990s was below rather than, as in the data adjusted with national deflators, above the lowest level in the 1970s.

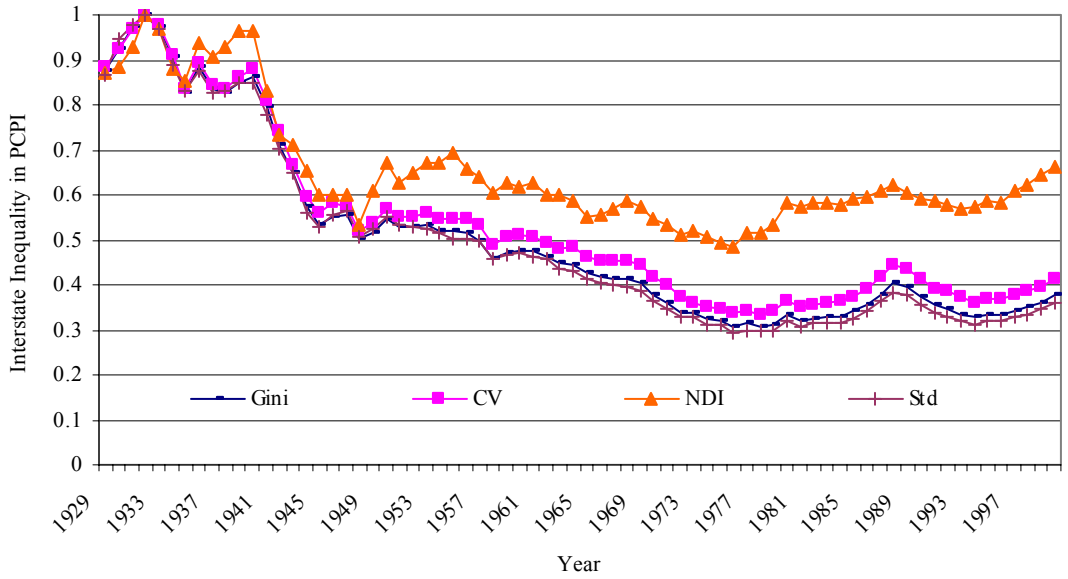


Figure 6. Comparison of Interstate Inequality in PCPI
Measured with Different Measures (1929-2000)

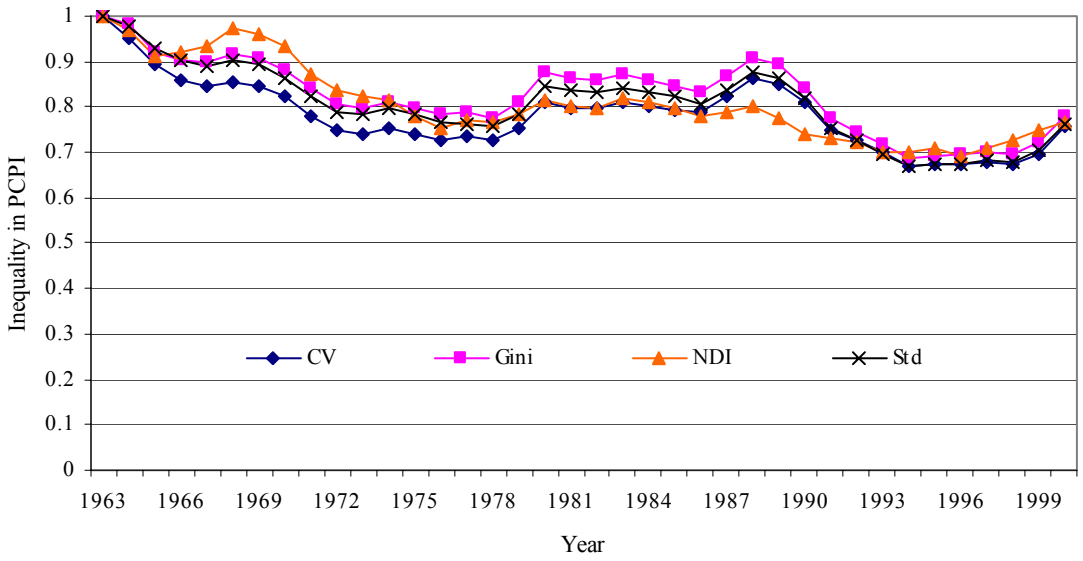


Figure 7. Comparison of Interstate Inequality in PCPI
with Different Measures (1963-2000)

Beta convergence

Beta convergence is measured for three time periods, 1929-2000, 1963-2000, and 1970-2000. All results of the three periods indicate the presence of β convergence, i.e. poor states grow faster than rich states. The convergence speeds and half life of convergence time are shown in Table 2. For the period 1929-2000, a convergence speed of about 2 percent per year is obtained with the data adjusted with the national price deflators and the half life of convergence is 42 years. To measure the effect of regional price deflators on β convergence, the convergence speeds measured with the data adjusted with the national and state-level price deflators are calculated for 1963-2000 and for 1970-2000. As shown in Table 2, in 1963-2000 the convergence speed with the data adjusted with the national price deflators is only 1 percent per year while the convergence speed with the data adjusted with state-level price deflators is about 2 percent per year. The half life of convergence is 64 years and 42 years, respectively. If the study period is shorter by 7 years (1970-2000), this reduces the convergence speed with the data adjusted with the national price deflators by half and doubles the half life of convergence. But the convergence speed and half life of convergence are barely affected when the data is deflated with the state-level price deflators. Obviously, convergence speed is affected by the use of price deflators and the length of study period. Using state-level price deflators, the trend of β convergence is enhanced. In other words, poor states grow even faster than rich ones when regional price differences are accounted for. This finding contradicts the argument that regional price dispersion does not have an impact on regional economic convergence (Sala-i-Martin, 1996).

Table 2. The Impact of Regional Price on Convergence Speed

Time Period	Convergence Speed (β coefficient)		Half Life of Convergence (years)	
	With national price deflators	With State Price Deflators	With national price deflators	With State Price Deflators
1929-2000	0.016396	NA	42.28	NA
1963-2000	0.010864	0.016491	63.80	42.03
1970-2000	0.005328	0.015519	130.1	44.66

Note: Half life of convergence time, $H = Ln(2)/\beta$

Conclusions

This paper analyzes the impact of regional price differences on interstate income convergence in the United States over the period 1963-2000 during which GSP data are available for the construction of state-level price deflators. The impact of regional price difference on both σ and β convergence is analyzed and evaluated. The results indicate that regional price differences affect both the extent and pattern of regional income convergence in the United States. Without accounting for regional price differences, interstate income inequality is greatly underestimated before 1995 and slightly overestimated after 1995, according to the results of standard deviation, coefficient of variation, and the Gini coefficient. It is underestimated before 1985 and overestimated after 1985 based on the results of neighborhood disparity index. The argument for a sustained strong reversal of interstate income convergence in the 1980s based on the data adjusted with the national price deflators is not supported by the data adjusted with state-level price deflators. The speed of convergence is underestimated without accounting for regional price differences. These findings contradict the argument that regional price dispersion does not have an impact on regional economic convergence. To better understand the development of regional economies, it is necessary to develop explicit regional price deflators in the future. Without regional price deflators, the pattern of regional economic development would probably be misunderstood.

References

Amos, O. M. (1988) Unbalanced regional growth and regional income inequality in the latter stages of development. *Regional Science and Urban Economics* 18, 549-566.

Barro, R. and Sala-i-Martin, X. (1991) Convergence across States and Regions. *Brookings Papers on Economic Activity* (1). 107-182.

Bernat, G.A., Jr. (2001) Convergence in state per capita personal income, 1950-1999. Presented at the 2001 Annual Meeting of the

Southern Regional Science Association, Austin, Texas, April 5-7, 2001. (A shortened version of this paper is published in *Survey of Current Business*, 81 (6), 2001 (June)).

Black, D. C. & Dowd, M. R. (1997) Measuring real interstate income inequality in the United States. *Economics Letters* 56, 367-370.

Chakravorty, S. (1996) A measurement of spatial disparity: the case of income inequality. *Urban Studies*, 33 (9), 1671-1686.

Coulombe, S. and Lee, F. (1993) Regional economic disparities in Canada. Mimeo. (University of Ottawa, Ottawa). Cited in Sala-i-Martin (1996).

Deller, S., Shields, M., and Tomberlin, D. (1996) Price differentials and trends in state income levels: a research note. *The Review of Regional Studies*, 26 (1), 99-113.

Fan, C. C. and Casetti, E. (1994) The spatial and temporal dynamics of US regional income inequality, 1950-1989. *Annals of Regional Science* 28, 177-196.

McMahon, W. W. and Melton, C. (1978) Measuring cost of living variation. *Industrial Relations* 17, 324-332.

McMahon, W. W. (1991) Geogrphical cost of living differences: an update. *American Real Estate and Urban Economics Association Journal* 19 (3), 426-449.

Sala-i-Martin, X. (1990) On Growth and States. PhD Dissertation, Harvard University.

Sala-i-Martin, X. (1996) Regional cohesion: evidence and theories of regional growth and convergence. *European Economic Review* 40, 1325-1352.

Shioji, E. (1992) Regional growth in Japan, Mimeo (Yale University, New Haven, CT), Cited in Sala-i-Martin, X. (1996).

Temple, J. (1999) The New Growth Evidence, *Journal of Economic Literature*, 37 (1), 112-56.

Slesnick, D.T. (2002) Prices and regional variation in welfare. *Journal of Urban Economics* 51, 446-468.

ANALYZING THE DEMAND FOR NON-ALCOHOLIC BEVERAGES

Annette Clauson, Economic Research Service, U.S. Department of Agriculture

Abstract

Because USDA is the lead Federal agency for nutritional information and programs of nutritional benefits for children and low-income households, ERS organized a team of experts to address the complex issues of why milk and more nutritious juices have been displaced by other non-alcoholic beverages. This presentation will outline the steps necessary to mine the raw database (the 1999 A.C. Nielsen Homescan Panel), select the tools for working with patterns in the data, and choose the decision trees and econometric methods for obtaining information on the drivers of demand for milk and other non-alcoholic beverage products. The findings should result in developing policies for analyzing, forecasting, and setting priorities in food programs.

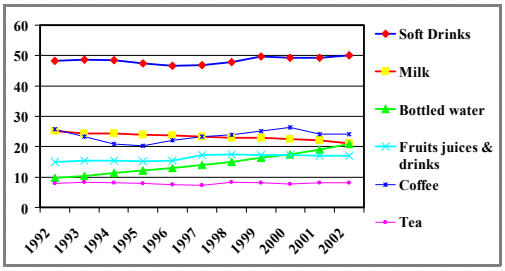
Analyzing the Demand for Non-Alcoholic Beverages

Annette Clauson, ERS, USDA

Project Cooperators: Texas A&M,Oral Capps, Jr., Grant Pittman, Matthew Stockton; ERS, Joanne Guthrie

1

U.S. Beverage Consumption gallons per person



2

Nonalcoholic Beverage Consumption ERS/USDA & AC Nielsen, 1999

	All Consumption Gallons/Person (ERS/USDA)	At Home Gallons/Person (AC Nielsen)	Percent of Consumption at Home
Soft Drinks	50.8	20.2	39.8%
Milk	23.6	13.2	55.9%
Bottled Water	18.1	5.6	30.9%
Fruit Juices	9.6	7.9	82.3%
Coffee	25.7	16.8	65.4%
Tea	8.4	5.8	69.0%

3

Project Objectives

(1) To analyze household consumption patterns of non-alcoholic beverages

(2) To analyze nutrient intake (per person per day) of calories, calcium, vitamin C, and caffeine derived from the consumption of non-alcoholic beverages

(3) To understand the drivers of demand for non-alcoholic beverages for poverty and non-poverty households

(4) To obtain own-price, cross-price, and expenditure elasticities of demand for non-alcoholic beverages for poverty and non-poverty households

4

Data Description

- 1999 AC Nielsen Homescan Panel
  - Tracked 7,195 households across the U.S. for entire year
  - Recorded expenditures and quantities for every individual food purchase at retail level
  - Demographics given for each household
    - Income
    - Household Size
    - Age, Employment Status, Education of Female Head
    - Age and Presence of Children
    - Race
    - Region
    - Hispanic Origin

5

Data Description

- Five data files
  - Dry Goods 4,111,719 records
  - Dairy Goods 873,899 records
  - Frozen Goods 1,002,851 records
  - Random Weights 507,306 records
  - Demographics 7,195 records

6

Demographic Description

- Region: East 20.3%, West 20.0%, South 34.3%, Central 25.3%
- Presence of Children: 30% of households have children under 18
- Race: 83.5% White, 10.2 % Black, 1.3% Oriental, 5% Other
- Hispanic Origin: 6.4% Hispanic
- Poverty Status: 94.1% above 130% poverty
- Household Size
  - 1-member household, 21.9%
  - 2-member household, 37.6%
  - 3-member household, 16.2%
  - 4-member household, 14.9%
  - 5+-member household, 9.4%

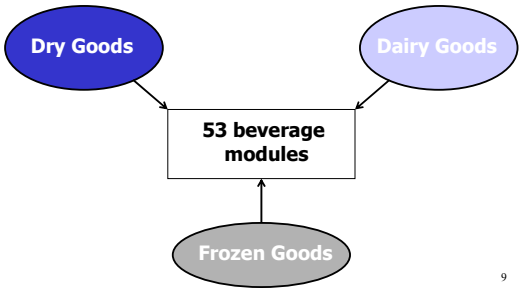
7

Selection of Modules

- Using Data Description File
  - Determine applicable nonalcoholic beverage categories
- Used 3 Files
  - Dry goods
  - Dairy goods
  - Frozen goods
- Found 53 product modules pertaining to non-alcoholic beverages

8

Selected Modules (Goods)



9

Units of Measurement

- Different modules had different units of measurement
  - Quarts
  - Gallons
  - Bags
  - Dry ounces
  - Concentrated ounces
- Convert all to gallons

10

Price Computation

- Each record contains
  - Quantity purchased (Q, in gallons)
  - Total expenditure (TE, in $)
- Calculate average price ($ per gallon) derived as (TE/Q)
  - Including discounts/coupons
- Each record now contains
  - Total quantity in gallons
  - Average price paid per gallon
  - Total expenditure

11

Problem Noticed

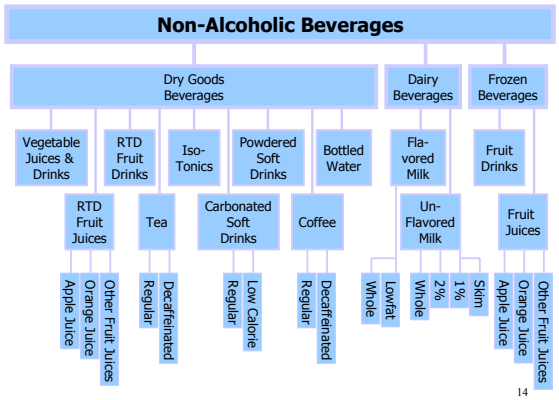
- Mean and Frequency checks exposed a problem
  - Some records had TE=0
  - But, Q > 0
- Why?
  - Don't know
- Small percentage had this problem
  - Dropped the records with this problem
    - .127 % of data (1,257 records out of 989,062)
- Concentrated in milk records

12

Combining Modules

- Each module now in gallons
- Aggregate into nonalcoholic groups
- Example: Coffee
  - Module #1463 Ground coffee
  - Module #1464 Soluble flavored coffee
  - Module #1465 Soluble coffee
  - Module #1466 Liquid coffee
- Create
  - All Coffee
  - Decaffeinated Coffee
  - Regular Coffee
  - Maintain individual modules as well

13



14

Data Clean Up

- Mean and Frequency procedures show outliers
  - Not all prices feasible due to generalized conversion factors
- Use Chebyschev's Inequality(+,- 5 st. dev)
  - At most drop 4 % of records
  - Overall dropped .218 % (2,154 records)
  - Packaged tea 1.2 % (conversion difficulty)

15

Creating Final Data Set

- Merging all modules
  - By household identification number
  - Add demographics in at this time
- Have a record for every purchase
- Aggregate records by time element
  - Quarterly
  - Annually
- Nearly 80% of households purchase non-alcoholic beverages of some type every month
- Nearly 97% of households purchase non-alcoholic beverages of some type 10 months of the year

16

Final Data Sets

- For each nonalcoholic beverage
  - Average price the households paid per gallon
  - Total Quantity purchased in time frame
  - Total expenditure in the time frame
- 77 Groupings of beverages
  - Aggregate categories
    - All Milk
  - Specific categories from initial modules (53)
    - 2% flavored milk
- Ready to Analyze (Quarterly, Annually)

17

Summary Statistics

Averages per consuming household (1999)			
All Tea	Price per Gallon	Gallons	Dollars Spent
5302 (73.7%)	1.24	15.00	18.58
All Coffee	Price per Gallon	Gallons	Dollars Spent
5584 (77.6%)	1.02	42.62	43.57
All Carbonated Soft Drinks	Price per Gallon	Gallons	Dollars Spent
7041 (97.9%)	2.34	51.87	121.19
Bottled Water	Price per Gallon	Gallons	Dollars Spent
4898 (68.1%)	1.24	14.33	17.73
All Milk	Price per Gallon	Gallons	Dollars Spent
7036 (97.8%)	2.76	33.93	93.50

18

Market Penetration for 1999, Selected Non-Alcoholic Beverages

All Tea	73.7%
All Coffee	77.6%
All Carbonated Soft Drinks	97.9%
Bottled Water	68.1%
All Milk	97.8%

19

Probit Analysis

Development of demographic profile for households who consume non-alcoholic beverages

- Household size
- Age, employment status, and education of female head
- Presence of children
- Race
- Hispanic origin
- Region
- Poverty status

20

Example: Bottled Water

Who is most likely to purchase bottled water?

- Black households
- Households located in the West
- Households with incomes above 130% poverty threshold
- Age of female head 25 to 49
- Female head with at least a high school education

21

Heckman Sample Selection Analysis

Issue: Ascertain drivers of consumption of non-alcoholic beverages

Example: Bottled Water

Once the decision is made to purchase bottled water, what are the determinants associated with the amount bought?

- Price is the only statistically significant factor

22

Nutrition Analysis

- Concentration on calories (kcal), calcium (mg), vitamin c (mg), caffeine (mg)
- Conversion of gallons to calories and milligrams using nutritive value of foods publication (Home and Garden Bulletin Number 72), October 2002

Groups of Non-alcoholic Beverages

- All nonalcoholic beverages (calories, calcium, vitamin C, caffeine)
- Carbonated soft drinks, fruit drinks, powdered soft drinks (calories, vitamin C)
- Ready-to-drink fruit juices, frozen fruit juices (calories, vitamin C)
- Milk (calories)
- Carbonated soft drinks (caffeine)
- Coffee (caffeine)
- Tea (caffeine)

Cross-tabulations with demographic characteristics

23

Average Nutrients Per Person Per Day Derived From The Consumption of Non-Alcoholic Beverages in 1999

Beverage Category	Calories (kcal)	Calcium (mg)	Vitamin C (mg)	Caffeine (mg)
All non-alcoholic beverages	194.60	196.16	41.42	87.68
Carbonated soft drinks, fruit drinks, powdered soft drinks	86.95		14.39	
Ready-to-drink fruit juices, frozen fruit juices	35.87		24.62	
Milk	65.92			
Carbonated soft drinks	Not Calculated			23.45
Coffee	Not Calculated			59.04
Tea	Not Calculated			5.09

24

Cross Tabulations
Average Nutrients Per Person Per Day
Derived From The Consumption of
All Non-Alcoholic Beverages in 1999

Demographic Factor	Calories (kcal)	Calcium (mg)	Vitamin C (mg)	Caffeine (mg)
Ethnicity				
Hispanic	190.11	165.45	39.39	67.56
Non-Hispanic	194.90	198.25	41.56	89.05
Region				
East	187.33	183.54	45.49	95.56
Central	208.75	217.80	39.81	91.31
South	197.94	187.34	42.99	83.36
West	178.33	196.75	36.64	82.48
Race				
White	196.22	210.90	39.70	94.55
Black	190.99	107.43	55.65	51.30
Oriental	135.83	133.77	36.99	42.37
Other	190.38	146.77	42.37	60.07

25

Nutrient Demand Analysis

- Regression analysis of nutrients per person per day derived from non-alcoholic beverages as a function of demographic variables
- Major findings about calories, calcium, vitamin C, and caffeine intake

26

Key Findings About Calorie Intake From Non-Alcoholic Beverages

- Intake lower for females 65+
- Intake lower for employed females
- Intake lower for females with some college
- Intake lower for orientals
- Intake higher in central and southern regions
- Intake lower in the West
- No differences by poverty threshold

27

Key Findings About Calcium Intake From Non-Alcoholic Beverages

- Intake lower for females 25 to 39
- Intake lower for employed females
- Intake lower for blacks, orientals, other
- Intake higher in central region
- Intake lower for households below 130% poverty threshold (by 21 mg)

28

Key Findings About Vitamin C Intake From Non-Alcoholic Beverages

- Intake decreases as females get older
- Intake lower for employed females
- Intake higher for blacks
- Intake higher in the East
- Intake lower for households below 130% poverty threshold (by 7 mg)

29

Key Findings About Caffeine Intake From Non-Alcoholic Beverages

- Intake increases as females get older
- Intake lower for employed females
- Intake lower for females with some college education
- Intake lower for blacks, orientals
- Intake higher in the East
- Intake lower in southern and western regions

30

Major Points About Nutrition Analysis

- On average, 10 percent of recommended daily intake of calories comes from non-alcoholic beverages; major contributors carbonated soft drinks, fruit drinks, powdered soft drinks (4.3 percent)
- On average, about 20 percent of recommended daily intake of calcium comes from non-alcoholic beverages
- On average, close to 70 percent of recommended daily intake of vitamin C comes from non-alcoholic beverages
- On average, daily intake of caffeine from non-alcoholic beverages is equivalent to almost two 12 oz cans of Coca-Cola or about one 7 oz cup of coffee or roughly a 1.25 12 oz serving of iced tea
- Intake of calcium and vitamin C derived from the consumption of non-alcoholic beverages is significantly lower for households below 130% poverty threshold

31

Demand System Analysis

Use of LA/AIDS Model

Censored Demand System Estimators

Milk, carbonated soft drinks, fruit juices, bottled water, other non-alcoholic beverages

Below 130% poverty threshold

Above 130% poverty threshold

All households

32

Censored Demand System Estimators

- Heien & Wessells (1990)
- Shonkwiler & Yen (1999)
- Perali & Chavas (2000)
- Yen, Lin, & Smallwood (2003)

33

LA/AIDS Model

Deaton and Meullbauer (1980)

Let W_i denote the expenditure share of product *i*

P_i denote the price of product *i*

X denote the total expenditure on all products in the system

$$w_i = \alpha_i + \sum_{j=1}^{N} \gamma_{ij} \ln p_j + B_i \ln X - B_i \left(\sum_j w_j \, hp_j \right) + \epsilon_i$$

Conditions for Adding-up

$$\sum_{i=1}^{N} \alpha_i = 1, \; \sum_{i=1}^{N} B_i = 0, \; \sum_{j=1}^{N} \gamma_{ij} = 0$$

Conditions for Homogeneity

$$\sum_{j=1}^{N} \gamma_{ij} = 0$$

Conditions for Symmetry

$$\gamma_{ij} = \gamma_{ji}$$

34

Descriptive Statistics

	Expenditure Shares	Average Gallons	Average Price per Gallon ($)
Milk	0.243	33.18	3.08
Bottled Water	0.032	9.76	1.85
Carbonated Soft Drinks	0.316	50.76	2.45
Fruit Juices	0.173	15.42	4.40
Other Non-Alcoholic Beverages	0.235	62.17	1.81

Price Imputations necessary for those cases where there are zero levels of consumption.

Price Imputations accomplished through the use of auxiliary regressions of reported prices on demographic variables.

35

Elasticity Measures Own-price Elasticities
(% change in consumption / % change in price)

	Below 130% Poverty Threshold	Above 130% Poverty Threshold	All Households
Milk	-0.877	-1.402	-1.375
Carbonated Soft Drinks	-0.751	-1.118	-1.091
Fruit Juices	-0.908	-0.938	-0.946
Bottled Water	-1.667	-1.505	-1.518
Other Non-Alcoholic Beverages	-1.019	-1.080	-1.079

Expenditure Elasticities
(% change in consumption / % change in total expenditure)

Milk	0.976	0.970	0.969
Carbonated Soft Drinks	1.194	1.189	1.188
Fruit Juices	0.854	0.743	0.753
Bottled Water	0.970	0.945	0.952
Other Non-Alcoholic Beverages	0.879	1.007	0.996

36

Major Competitors in Terms of Significant Cross-Price Elasticities

	Below 130% Poverty	Above 130% Poverty	All Households
Milk	Other non-alcoholic beverages	Carbonated soft drinks	Carbonated soft drinks
	Carbonated soft drinks	Fruit juices	Fruit juices
		Other non-alcoholic beverages	Other non-alcoholic beverages
		Bottled water	Bottled water
Carbonated Soft Drinks	Milk	Milk	Milk
	Other non-alcoholic beverages	Other non-alcoholic	Other non-alcoholic
		Fruit juices	Fruit juices
		Bottled water	Bottled water
Fruit Juices	Other non-alcoholic beverages	Milk	Milk
		Other non-alcoholic	Other non-alcoholic
		Carbonated soft drinks	Carbonated soft drinks
Bottled Water	Other non-alcoholic beverages	Milk	Milk
		Other non-alcoholic	Other non-alcoholic
		Carbonated soft drinks	Carbonated soft drinks
Other Non-Alcoholic	Fruit juices	Carbonated soft drinks	Carbonated soft drinks
Beverages	Milk	Milk	Milk
	Carbonated soft drinks	Fruit juices	Fruit juices
	Bottled water	Bottled water	Bottled water

37

DISABILITY BENEFIT APPLICATIONS, ECONOMIC TRENDS, AND PROJECTIONS

Mikki D. Waid and Fred Joutz
Research Program on Forecasting
The George Washington University
Department of Economics

Presented at the Federal Forecasters Conference
October 2003

Abstract

The current solvency issues regarding the Social Security program suggest that this is a good time to re-visit the topic of disability applications. More specifically, previous research has suggested that the number of disability (DI) applications is affected by macroeconomic variables such as the unemployment rate. During times of high unemployment, more individuals may apply for disability benefits to replace lost earnings. In addition, other economic and demographic trends may cause the number of applications to increase or decrease. For example, supply side changes, the overall health of the workforce, the number of workers with taxable earnings, labor force participation rates, the number of disability awards, and the value of disability benefits may all have an effect on the number of applications. Modeling the number of disability applications will assist in strategic planning regarding administration costs of the Social Security program and in determining how long the disability program will be solvent. This paper presents preliminary results from an experimental model for the number of annual Social Security disability applications. Annual data from 1962 to 2001 is used to analyze the factors contributing to the growth of disability applications. The model's with-in sample and out-of-sample forecasting performance will be evaluated.

Achowledgements: The authors wish to thank Federal Forecaster's Conference participants in the Social Programs Session and the Applied Econometrics Workshop at George Washington University.

INTRODUCTION

The *2002 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Disability Insurance Trust Funds* states that the combined OASI (Old Age and Survivors Insurance) and DI (Disability Insurance) Trust Funds are expected to become exhausted in the year 2041. In addition, the DI Trust Fund, alone, is expected to become exhausted in the year 2028. The costs of the OASDI Trust Fund are expected to increase rapidly in the near future because of the large number of baby-boomers (born between 1946 and 1964) who are expected to receive benefits. In 2001, the ratio of workers to OASDI beneficiaries was estimated to be 3.4. By 2030, this figure is predicted to fall to 2.1. Being a pay-as-you-go system, the large number of beneficiaries compared to the small number of workers contributing to the system should put a strain on the OASDI Trust Funds. Given that the DI Trust Fund is projected to become exhausted in approximately 25 years, now is the time to revisit issues surrounding the DI program and costs to the program.

One important topic regarding the DI program is the vast increase in the number of DI applications over the past several decades. The number of DI applications has increased from 418,600 in 1960 to 1,489,600 in 2001. The increase in applications increases the costs to the DI program in several ways. First, it is likely that an increase in applications will correspond to an increase in the number of beneficiaries. Second, an increase in applications will correspond to an increase in the number of employees needed to review the applications and possibly to an increase in the amount of time used to review the applications (the amount of time used to review an application can increase because individuals who initially have their application rejected can appeal the rejection). Thus, the number of disability applications will

have a major impact on the DI Trust Fund through the number of beneficiaries and the costs associated with the application process (such as the costs of appeals).

Therefore, to gain a better understanding of the future of the DI Trust Fund, this paper will present a model of the number of DI applications. More specifically, this paper will attempt to determine the factors that are significant in calculating the number of annual DI applications. In addition, the model will test whether certain macroeconomic variables are significant in determining the number of applications and whether these variables should be considered exogenous. Such information is needed for projecting the future number of applications and for addressing future costs to the DI program. In addition, the number of DI applications will be forecasted to the year 2010.

There are six sections to the paper. The first section reviews the literature on macroeconomic factors influencing the disability program. The second section discusses disability insurance in terms of an aggregate model. Supply side effects are characterized by institutional and legislative changes in the program. Demand side effects are determined by macroeconomic activity, the opportunity cost of disability benefits, eligible workers, and previous awards. The data is discussed in the next section. The specification and estimation of the model is described in section our. Forecasts are made in section five and followed by the conclusion.

I. LITERATURE REVIEW

There have been a variety of reports which attempt to link the disability program with macroeconomic variables. Early reports used time series data whereas the more recent reports have used cross-section and "pooled" data. Lando, Coate, and Krause (1979) use time series data to determine the effects of macro variables on the number of DI applications. Using quarterly data from 1964-1978, they find that a one percentage point change in the unemployment rate increased the number of DI applications received by district offices by approximately 11 thousand per quarter. In addition, they find that a one percentage point increase in a replacement rate measure (defined as the value of a disability cash benefit for new awardees divided by the Bureau of Labor Statistics data for spendable average earnings for a worker with three dependents) increases the number of DI applications by 375 thousand.

Halpern (1979) uses quarterly data from 1964-1978 and also attempts to find a relationship between macroeconomic variables and the number of DI applications. Unlike Lando et al., Halpern does not find the unemployment rate to be significant in determining applications. However, she admits that the insignificance in unemployment could result from an omitted variable bias. She finds the benefit replacement rate (average monthly benefits divided by the average spendable earnings for a worker with three dependent children), the population insured for disability, the percentage of the insured population older than 45, a dummy for the year 1968, and a dummy for the year 1974 to be all significant[1].

Stapleton, Coleman, et al. (1998) have conducted one of the more recent studies regarding the number of DI applications. Using annual, pooled, cross-section/time series state level data from 1988-1992, they find that the unemployment rate has a significant effect on the number of DI-only[2] applications. A one percentage point increase in the unemployment rate increases DI applications by 4 percent. They also find that the effect of the unemployment rate is stronger for men than women. The other significant variable in their model was an estimate of the percentage increase in DI applications from 1988 to 1992 that was not accounted for by the explanatory variables.

One possible criticism of the above studies is that the results do not take into account possible causality issues. For example, Halpern (1979) suggests that her insignificant macroeconomic effects may be misleading. She states that "problems with the data and the specification of the model … make it undesirable to draw a firm conclusion on the interaction of unemployment and DI applications". To be more specific, she states that "the simultaneous decline in the number of applications and the unemployment rate between 1976 and 1978 suggests a causal relationship between these two variables". Halpern was not the only individual to suggest that there may be causal relationships between the social security program and other

[1] Reasons for the 1968 and 1974 dummy variables will be explained later in the paper.
[2] DI-only refers to individuals who do not simultaneously apply for SSI (Supplemental Security Income). All results in this paper are for DI applications only.

variables. Bound (1989) estimates a cross-sectional model with data from the 1972 Survey of Disabled and Non-Disabled Adults and the 1978 Survey of Disability and Work. He looks at rejected DI applicants and finds that DI beneficiaries were, for the most part, disabled and many would not have been working even in the absence of the DI program. Thus, the DI program served as a substitute for the "more meager state-run programs" that were in existence. Thus, he casts doubt on the large disincentive labor-force effects presented in other cross section studies. Bound suggests that the results from other studies may suffer from causality issues. He states that:

> "To study behavioral responses to social programs (for example, disability insurance, unemployment insurance, workers' compensation), researchers have often used replacement rates, potential benefits, or other program parameters as explanatory variables, even when these variables could not plausibly be taken to be exogenous. This paper should underline the potential dangers in such exercises. The results of such exercises simply cannot be informative about any causal relationship between program design and behavioral response" [500].

Thus, if one attempts to model variables associated with social programs (such as disability applications), one should consider the possibility that the independent variables may actually be endogenous. Although, studies generally have not used disability applications to determine labor force participation, studies have used the value of disability benefits. For example, Parsons (1980) uses data from the National Longitudinal Surveys of Older Males (aged 45-59 in 1966) to model the labor force participation decision in 1969. He finds that the replacement ratio (social security benefits divided by the wage rate) has a significant negative effect on labor force participation. On average, he finds a 10 percent rise in benefits (without a simultaneous increase in the wage rate) will reduce labor force participation by 6 percent.

Autor and Duggan (2001) use social security administrative data matched to the Current Population Survey monthly files for 1978-1999 to find that liberalizations in DI legislation reduced the U.S. unemployment rate. A change in legislation in 1984 (which allowed more individuals to be eligible for disability benefits) and more changes in the system from the 1980's through the 1990's (which increased the replacement rate of benefits for low skilled workers) reduced the unemployment rate by 0.64 percentage points. Thus, they find that liberal legislation in the disability program allowed the low-skilled unemployed to exit the labor force.

The final two papers suggest that social security benefits may have an effect on labor force participation and the unemployment rate. If applications affect the level of benefits, this suggests that applications may also have an effect on labor force participation and the unemployment rate and that models of disability applications should consider that these variables may have feedback effects.

II. A SIMPLE AGGREGATE MODEL OF DISABILITY APPLICATIONS

In this section a simple aggregate model of disability applications is proposed. The number of disability applications can change due to supply side changes as the result of legislative actions affecting who is covered and how much they are entitled to. Demand side effects come from changes in macroeconomic variables (such as the unemployment rate and the labor force participation rate), the change in the value of benefits, the overall health of the population, and the number of workers with earnings taxed by the Social Security program.

Supply Side Changes

There have been legislative changes to the disability program which have made individuals more/less likely to be eligible for disability benefits. These changes may, in turn, affect an individual's decision to apply for benefits. For example, a liberal change in the disability program may induce individuals to apply for disability benefits who would otherwise have not applied. Table 1 lists legislative changes to the Disability Insurance program. Although not all changes appear to have had dramatic affects on the number of

disability applications, a few legislative changes have had noticeable effects.

The 1967 legislation made it easier for young workers to become insured for disability benefits. This legislation is thought to have encouraged young workers to apply for disability benefits. In 1968, the number of applications increased by 26 percent from its 1967 levels. The institutionalization of the Supplemental Security Income (SSI) program in 1972 is also thought to have increased the number of applications. The SSI program provides income support for disabled individuals (as well as individuals age 65 and over). The SSI program is thought to have greatly disseminated information regarding the DI program. The increase in knowledge of the program encouraged more individuals to apply for disability benefits. The number of disability applications increased by 40 percent between 1972 and 1974.

Legislation in 1980 limited disability benefit levels and tightened administration of the Social Security and SSI disability programs. From 1980 to 1984, the number of disability applications decreased by 18 percent. The Ticket to Work and Work Incentives Improvement Act, initiated in 1999, allows individuals to work for a certain period of time without loosing their disability benefits (before the initiation of this program, a beneficiary could loose disability benefits if his/her job earnings were above a certain amount). From 1999 to 2000, the number of disability applications increased 11 percent and increased another 13 percent from 2000 to 2001. The change in the number of applications indicates that these legislative changes could have a lasting impact on the number of individuals who decide to apply for disability benefits. The 1967, 1972, and 1999 legislative changes are thought to have increased the number of applications while the 1980 legislative change had the opposite effect.

Macroeconomic Variables

Some individuals are disabled and currently working. However, if these individuals were to lose their jobs, they may decide to apply for disability benefits. Therefore, in times of high unemployment, more individuals may decide to apply for disability benefits as an alternative to working. Thus, the unemployment rate should be positively correlated with the number of disability applications.

The female labor force participation rate may also be an indicator of the number of disability applications. First, an increase in the labor participation rate, itself, suggests that more individuals could be eligible for disability benefits (because more individuals are working and paying into Social Security). Thus, an increase in the female labor force participation rate can increase the number of disability applications. Second, women tend to live longer than men (see, for example, U.S. Department of Health and Human Services). Thus, at any given time, women may be in better health than men. Therefore, an increase in the female labor force participation rate may suggest that the workforce is generally in better health (thus, fewer in the workforce will apply for disability benefits). These two effects work in opposite directions and cause the effect of the female labor force participation rate to be ambiguous.

The Value of Disability Benefits

In this paper, the value of disability benefits is defined as the average annual Social Security disability benefit divided by the average disposable income. If the value of disability benefits increases, individuals (especially those with low income) may decide that collecting a disability benefit is more valuable than working. Thus, an increase in the value of disability benefits should correspond to an increase in the number of applications.

The Overall Health of the Population

It is difficult to measure the overall health of the population. As a proxy for the health of the population, this paper will use the average number of bed disability days per person[3]. An increase in the number of bed disability days can indicate that the population, for that year, is in poorer health. Thus, a decrease in health (i.e. an increase in the number of bed disability days) of the population should correspond to an increase in the number of disability applications.

The Number of Workers with Taxable Earnings

Individuals can not receive disability benefits unless he/she is insured. In order to be insured, an individual must have enough quarters

[3] In the National Health Interview Survey, respondents were asked how many times in the past 12 months an illness or injury had kept them in bed for more than half of the day. They were instructed to include days on which they were an overnight patient in a hospital.

of coverage. A quarter of coverage is obtained by earning a certain amount of money from a job. For example, in 2001, the earnings required for 1 quarter of coverage were $830. A person can earn up to 4 quarters of coverage each year. In order to be insured for disability benefits, a person must have earned at least 20 quarters of coverage during the past 40 quarters (different rules apply for individuals who become disabled before age 31). In order to receive quarters of coverage an individual must have wages which are subject to the Social Security tax. In other words, the individual must have taxable earnings. The larger the number of individuals with taxable earnings, the greater is the number of individuals who will eventually be insured for disability. Thus, an increase in the number of individuals with taxable earnings could correspond with an increase in the number of applications.

The Number of Individuals Awarded a Disability Benefit

The number of disability awards could encourage other individuals to apply for disability. An increase in awards could encourage individuals to apply for DI benefits because the award increase may signify an increase in the probability of being awarded a DI benefit. In addition, individuals who were previously awarded benefits could educate those who would like to apply (or even those who have not thought about applying). In addition, a large number of awards in the previous year could signify a liberalization in the awarding of disability benefits. This too, may encourage, more individuals to apply for benefits. Thus, there should be a positive correlation between the number of DI awards and the number of DI applications.

III. EXAMINATION OF THE DATA

This section will present the data series as well as analyze the time series properties of the data. The data are annual for the years 1962 to 2001.

Figure 1 shows a plot of SSI Disability Application from 1962 to 2001 along with the major legislative changes. Annual applications grew from 400,000 to 1.3 million in the first decade then declined as many in the baby-boom were entering the labor force, the relative SSI benefits declined, and awards and coverage became tighter. In 1987, they began to rise again and picked during the recession and slow economic recovery in the early 1990s before declining again as the economy. Applications pick up again in 1997 and reach their peak of 1.7million in 2001. They are expected to increase as the labor force ages.

Figure 2 suggests that disability applications and the female labor force participation rate have increased over time. However, the female participation rate has less variation than the number of DI applications. In addition, the growth rate of female labor participation decreases beginning in 1989.

Not surprisingly, the number of DI awards closely follows the number of DI applications (Figure 3). This graph suggests that the increase in awards reflects an increase in applications rather than differences in legislation. The number of awards, however, shows slightly less variation than the number of applications, although the number of awards decreases much more than the number of applications from the years 1977 to 1982. The number of awards reaches a peak in 1975 (a year later than DI applications reach a peak) and does not reach another peak until 1992 (two years before the number of applications reaches a peak).

The benefit replacement rate appears to decline over time (Figure 4). Benefit replacement varies the most between 1962 and 1984 and decreases afterward. The replacement rate is also highest in 1972 when there was an increase in benefits due to legislation (see Table 1). Contrary to what is expected, the number of applications and the replacement rate appear to move in opposite directions up until 1994.

The number of days (per person) spent in bed due to a disability fluctuates a great deal over time (Figure 5). The series appears to cycle approximately every two years and appears to cycle around 6.3 days. However, there is a sudden drop in 1997. It is not clear why this drop occurs.

As previously stated, the unemployment rate and the number of applications should be positively correlated. However, a visual inspection of Figure 6 gives mixed results. The number of applications increases between 1963 and 1970 while the unemployment rate is decreasing. Both measures increase between 1970 and 1972 and then the two series appear to move in opposite directions between 1975 and 1986. After 1986, the number of DI applications and the unemployment rate begin to move together once again.

Also previously stated, the number of workers with taxable earnings increases the number of individuals who can apply for disability benefits. Thus, an increase in the number of workers with taxable earnings should correspond to increases in DI applications. Figure 7 supports this idea. Both the number of workers with taxable earnings and the number of disability applications have increased over time. However, the number of workers with taxable earnings shows much less variation over time.

IV. MODEL SPECIFICATION AND ESTIMATION

We employ the general-to-specific modeling approach advocated by Hendry (1986)). It attempts to characterize the properties of the sample data in simple parametric relationships which remain reasonably constant over time, account for the findings of previous models, and are interpretable in an economic and financial sense. Rather than using econometrics to illustrate theory, the goal is to "discover" which alternative theoretical views are tenable and test them scientifically.

The approach begins with a general hypothesis about the relevant explanatory variables and dynamic process (i.e. the lag structure of the model). The general hypothesis is considered acceptable to all adversaries. Then the model is narrowed down by testing for simplifications or restrictions on the general model.

The first step involves examining the time series properties of the individual data series. We look at patterns and trends in the data and test for stationarity and the order of integration. Second, we form a Vector Autoregressive Regression (VAR) system. This step involves testing for the appropriate lag length of the system, including residual diagnostic tests and tests for model/system stability. Third, we examine the system for potential cointegration relationship(s). Data series which are integrated of the same order may be combined to form economically meaningful series which are integrated of lower order. Fourth, we interpret the cointegrating relations and test for weak exogeneity. Based on these results a conditional error correction model of the endogenous variables is specified, further reduction tests are performed and economic hypotheses tested.

Unit Root Tests

In order to avoid potential problems with the regressions, it is important to determine the stationary order of integration of the variables. Tables 2 and 3 present results from the Augmented Dickey-Fuller tests for a unit root in levels and first differences, respectively. The rejection of a unit root suggests that the series is stationary and that the hypothesis tests from least squares can be used. The numbers in parentheses is the number of lags used as based on the Akaike Info Criterion (AIC). The specifications for the ADF tests use a maximum of four lags in levels and three lags in first differences. We cannot reject the null hypothesis of a unit root process for all the series in levels. However, for all variables with the exception of bed disability days, we can reject the null hypothesis of a unit root for the series in differences. As can be seen, none of the series are stationary in levels. All of the series contain a unit root with the exception of the number of disability bed days which contains two unit roots. Thus, in order to obtain stationary series, the number of disability days will have to be differenced twice and all other variables will have to be differenced once.

The VAR Model

This section will present a long run model and will test for cointegrating vectors. The variables most likely to affect the number of disability applications, in the long run, are the benefit replacement rate and the number of workers with taxable earnings. As previously mentioned, individuals cannot receive disability benefits unless they are “insured” and the number of “insured” depends on the number of individuals with taxable earnings. Thus, the number of workers with taxable earnings is likely to correspond to the number of those who are insured. In a similar manner, the benefit replacement rate should affect the number of applications in the long run. A high benefit replacement rate should encourage more individuals to apply for disability benefits because more of their employment earnings can be potentially replaced by benefits. The VAR is specified below.

$$Z_t = \Pi_0 + \sum_{i=1}^{k} \Pi_i Z_{t-i} + \sum_{j=0}^{m} \Theta_j X_{t-j} + E_t$$

where

$$Z_t = \begin{bmatrix} Z_{1t} \\ Z_{2t} \\ Z_{3t} \\ \end{bmatrix} = \begin{bmatrix} Log\ Disability\ Applications \\ Log\ Benefit\ Replacement \\ Log\ Workers\ with\ Taxable\ Earnings \\ \end{bmatrix}$$

where

$$X_t = \left[Female\ Part.Rate, Lagged\ Beddays, \quad TotalEmployment, Dum67, Dum72 \right]'$$

where last term, $\mathbf{E_t}$, is a matrix of random disturbances assumed to be approximately normally distributed.

The VAR system consists of the log of disability applications, the benefit replacement rate and the log of the number of workers with taxable earnings with four lags. The model is conditioned on female labor force participation, lagged number of hospital bed-days, total employment, and dummy variables for 1967 and 1972. We did not find evidence for the legislative change in 1980 in terms of a dummy variable, bust suspect that it is captured in the declining bebfit replacement variable.

Before checking for cointegration, a vector autoregression is run (VAR) and the correct lag length is specified. Table 4 gives mixed results for the appropriate lag length. The Likelihood Ratio, the Schwarz Information Criterion, and the Hannan-Quinn Information Criterion select two lags. However, the Final Prediction Error and Akaike Information Criterion both select four lags. In order to pin down the correct lag specification, Wald exclusion tests are also calculated. The tests do not reject the null of two lags (Table 5). Thus, the criteria indicate that the system should be estimated using differences.

Testing for Cointegration an the Long Run Model

Table 6 provides the results from the cointegration analysis. The Trace and Max-Eigenvalue tests indicate two cointegrating equations at both the 5% and 1% levels. The coefficients for the benefit replacement and the number of workers with taxable earnings are positive and significant at the 1% level. Thus, a one-percentage point increase in the number of workers with taxable earnings will raise the number of applications by 13.9 percent percentage points. This seems large, but it is a long-run estimate of the number of applications for disability insurance not the number of awards. In addition, a one-percentage point increase in the benefit replacement rate will increase the number of DI applications by 1.93 percent. In addition to significant coefficients, the speed of adjustment or alpha coefficient for disability applications is –0.06 and significant from zero. Thus, the relationship appears to be stable.

Analysis of the residuals suggests they are approximately normal and not serially correlated. Finally, with the exception of the cross-correlogram for benefit replacement and workers with taxable earnings, all estimated residuals in the VAR appear to be within two times the asymptotic standard errors of the lagged correlations (Figure 8).

Additional tests for weak exogeneity were not entirely satisfying. All three variables do not appear to be exogenous. However, theory and common sense suggest that the number of workers in the economy is not related to disability applications. Also, the benefit replacement differential is set by the congress and is not determined by the model in a strict sense. Thus, we drop these two equations and estimate a single error correction equation for benefit applications.

The Short Run or Error Correction Model

The VAR analysis suggests that the model should be analyzed in first differences. Thus, the short-run model is converted to first differences. In the process we add several other variables which may not be important in the long-run, but are important in a short-run sense. Theses include the recent change in the number of awards (an "announcement effect"), the change in the unemployment rate, and the policy dummy variables. The following general equation is analyzed:

$$D(LogDIApp)_t = \pi_0 + \pi_1 \Delta FemaleLabor_t + \pi_2 \Delta NumAwards_{t-1} +$$
$$+ \pi_3 \Delta Benefit\text{Rep}_{t-1} + \pi_4 \Delta BedDays_{t-1} + \pi_5 \Delta UnemRate_{t-1} +$$
$$+ \pi_6 \Delta WorkTaxEarn_{t-1} + \pi_7 Dummy1967 + \pi_8 Dummy1972 +$$
$$+ \pi_9 Dummy1980 + \pi_{10} ECM_{t-1} + \varepsilon_t$$

$$where\ ECM_{t-1} = LOGDIAPP(-1) - 1.93 * BENREPLCE(-1) -$$
$$- 13.98 * LOG(WRKTAXE(-1)/1000) + 63.84$$

The model is fit from 1963 to 2001. There are 11 estimated coefficients and 38 observations. Thus, there are not as many observations as one would like for the estimation.

Table 7 shows the results of the short-run general model. The difference in awards and the difference in disability bed days have the correct sign. In addition, the female participation rate has the correct sign if an increase in women in the labor force causes the labor force to be healthier, in general. The other variables, however, have incorrect signs. For example, an increase in the benefit replacement rate should cause an increase in disability applications; however, the short-run general model predicts that the benefit replacement rate will have a negative impact. Similarly, the unemployment rate, the number of workers with taxable earnings, the 1967 dummy, the 1972 dummy, and the 1980 dummy all have the incorrect sign. The female labor participation rate, the number of awards, the benefit replacement rate, and the error correction term are all significant, although the coefficient for the difference in the number of awards is very close to zero. The adjusted R-square is 0.31 and the null hypothesis that all coefficients are jointly zero can be rejected.

In order to find a better fitting model, Wald tests were conducted and variables were dropped one-by-one to determine whether any variables can be used for the specific short-run model. The following equation is the final short-run model:

$$D(LogDIApp)_t = \pi_0 + \pi_1 \Delta FemaleLabor_t + \pi_2 \Delta NumAwards_{t-1} +$$
$$+ \pi_3 \Delta Benefit\text{Rep}_{t-1} + \pi_4 ECM_{t-1} + \varepsilon_t$$

$$where\ ECM_{t-1} = LOGDIAPP(-1) - 1.93 * BENREPLCE(-1) -$$
$$- 13.98 * LOG(WRKTAXE(-1)/1000) + 63.84$$

The significant variables are the change in the female labor participation rate, the change in the number of awards, the change in the benefit replacement rate, and the ECM. Table 8 shows that the female labor rate, the change in the benefit replacement, and the error correction term are all significant at the 1% level. The change in the number of awards is significant at the 5% level. Similar to the general model, the female labor rate and the number of awards have the predicted signs. However, the benefit replacement rate still is negatively correlated to the number of disability applications. It is unclear why an increase in the benefit replacement rate difference would cause a decrease in the change in DI applications. The results suggest that a 1% point increase in the female labor force partition rate will cause a 2% decrease in the number of disability applications. In addition, the regression suggests that an increase in the change of awards by one will increase the number of applications by one. Finally, the ECM term, -0.04, is negative and significant. Thus, the deviation from the "equilibrium" long run number of disability applications is significant. In addition, the unemployment rate does not appear to affect the number of DI applications in differences.

There is no loss in explanatory power from the original model with the specific model. Figure 9 presents residual diagnostics and suggests the estimated error is approximately

normally distributed. There is no serial correlation. Finally the model is stable; Figure 10 presents a graph of the recursive N-step Chow test for the Final Model. There are no rejections. The recursive coefficient estimates are plotted in Figure 11 and they appear to be fairly stable.

V. A SIMPLE FORECAST

We develop a preliminary forecast model for 2002 through 2010 in this section. Projections for the explanatory variables were developed using simple univariate techniques ARIMA(3,0,0) models. The female labor force participation rate uses a ARIMA(1,1,0) with trend term which was negative to level off the rate well below unity. The forecast is presented graphically in Figure 12 and reveals that applications could more than double in the next 8 years. This seems large but not out the realm of possibility, because of demographic changes in the labor force with early baby-boomers aging and increasing their propensity to apply for disability insurance.

VI. CONCLUSION

Modeling the number of disability applications is a very challenging process. The long-run model suggests that the benefit replacement rate and the number of workers with taxable earnings are significant and weakly exogenous variables. Thus, the number of disability applications must be modeled with one cointegrating vector. For the short-run model, the significant variables are the female labor rate, the difference in the number of awards, the difference in the benefit replacement rate, and the error correction variable. Thus, it appears that the benefit replacement rate is important in both the long and short-runs. However, the sign of the replacement rate appears to be contradictory to what one would expect for both equations. In addition, the deviation from the log-run "equilibrium" is significant in the short-run. Thus, the short-run model suggests that the female labor fore participation rate is the only macroeconomic variable significant in determining the number of applications. The unemployment rate was not significant in any of the models.

In addition, forecasts suggest that the number of DI applications could increase to almost 4 million by the year 2010. This is a preliminary model. We want to continue testing it against alternative models.

BIBLIOGRAPHY

Autor, David and Duggan, Mark (2001). The Rise in Disability Recipiency and the Decline in Unemployment. *National Bureau of Economic Research, Working Paper No. 8336*.

Bound, John (1989). The Health and Earnings of Rejected Disability Insurance Applicants. *The American Economic Review*, Vol. 79, No. 3, 482-503.

De Brouwer, Gordon and Ericsson, Neil (1998). Modeling Inflation in Australia. *Journal of Business & Economic Statistics*, Vol. 16, No. 4, 433-449.

Gruber, Jonathan (1996). Disability Insurance Benefits and Labor Supply. *National Bureau of Economic Research, Working Paper No. 5866*.

Halpern, Janice D. (1979). The Social Security Disability Insurance Program: Reasons for Its Growth and Prospects for the Future. *New England Economic Review*, May/June 1979, 30-48.

Halpern, Janice and Hausman, Jerry (1985). Choice Under Uncertainty: A Model of Applications for The Social Security Disability Insurance Program. *National Bureau of Economic Research, Working Paper No. 1690*.

Joutz, Frederick L. and Maxwell, William (2002). Modeling the yields on noninvestment grade bond indexes: Credit risk and macroeconomic factors. *International Review of Financial Analysis*, 11, 345-374.

Lando, Mordechai, and Coate, Malcolm and Kraus, Ruth (1979). Disability Benefit Applications and the Economy. *Social Security Bulletin*, Vol. 42, No. 10, 3-10.

Parsons, Donald (1980). The Decline in Male Labor Force Participation. *Journal of Political Economy*, Vol. 88, No. 11, 117-134.

Rupp, Kalman and Stapleton, David (1995). Determinants of the Growth in the Social Security Administration's Disability Programs – An Overview. *Social Security Bulletin*, Vol. 58, No. 4, 43-70.

Social Security Administration (2001). Annual Statistical Report on the Social Security Disability Insurance Program.

Social Security Administration (2001). Annual Statistical Supplement, 2001: to the Social Security Bulletin.

Song, Jae (2002). Pre-Disability Earnings and Labor Force Attachment of Applications for Social Security Disability Benefits. *Mimeo*, September 2002.

Stapleton, David, Coleman, Kevin et al. (1998). *Growth in Disability Benefits*. Kalamazoo, Michigan: W.E. Upjohn Institute for Employment Research (Chapter 2). Rupp, Kalman and Stapleton, David eds.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, *National Vital Statistics Reports*, Vol. 51, No. 3.

TABLE 1: LEGISLATION CHANGES FOR SOCIAL SECURITY'S DISABILITY INSURANCE PROGRAM

Year of Legislation	Description of Legislation	Effect on Applications
1954	Initiation of Disability Insurance (DI) program (did not offer cash benefits)	+
1956	Allowed benefits for disabled workers aged 50-64 (after a 6 month waiting period) and to adult children of retired, disabled, or deceased workers if the children had been disabled before age 18.	+
1958	The reduction for workers' compensation was eliminated.	+
	Provided benefits for the dependents of disabled workers.	+
1960	Allowed disabled workers aged 18-49 to qualify for benefits.	+
	Instituted a 12-month trial work period with the continuation of benefits.	+
1965	Broadened the definition of disability from "long and indefinite" to "at least 12 months"	+
1967	Redefined disability to consider the ability to hold jobs that exist in substantial numbers in the local or national economy.	+
	Reduced the number of quarters of coverage needed to be eligible for benefits for individuals under the age of 31.	+
	Provided benefits for disabled widow(er)s aged 50-64 at a reduced rate.	+
1972	Reduced the waiting period from 6 months to 5.	+
	Increased from 18 to 22, the age before which a "childhood disability" must have begun.	-
	Extended Medicare coverage to persons who had been receiving disability benefits for 24 consecutive months.	+
	Established the needs-based Supplemental Security Income (SSI) program to replace the Old-Age Assistance, Aid to the Blind, and Aid to Permanently and Totally Disabled programs.	+
	Eliminated the disability insured requirement for blind workers.	
	Increased benefit levels and provided for automatic CPI-indexing.	+
		+

1980	Limited disability benefits levels. Tightened administration of the Social Security and SSI disability programs by instituting a review of initial disability decisions and by establishing a periodic review of continuing disability requirements. Enhanced rehabilitation and work incentive provisions. Withheld benefit payments to incarcerated felons.	- - + -
1982	Responding to the 1980 legislation, persons who appeal that their disability has ceased could elect to have their benefits and Medicare coverage continued pending review by an administrative law judge and could have an opportunity for a face-to-face evidentiary hearing at the reconsideration level of appeal.	+
1983	Gradually increased the age at which full retirement benefits were payable from 65 to 67. Thus, disabled workers and widow(er)s may remain on the DI rolls for an additional 2 years before converting to age-based benefits. This may induce additional older workers to apply for and become entitled to disability based benefits. Improved benefits to disabled widow(er)s were improved by decreasing the benefit reduction for beneficiaries under age 60 and by continuing payments to certain disabled widow(er)s who remarried.	+ +
1984	Revised the mental impairment listings and considered the combined effect of all impairments when determining eligibility for benefits. This was done, in part, to liberalize some of the 1980 amendments.	+
1984-1998	Provided additional Medicare protection for the disabled. Made the definition of disability for disabled widow(er)s the same as those for disabled workers. Prohibited eligibility for individuals whose drug addiction or alcoholism was a contributing factor to their impairment. Modified the provisions for a trial work period.	+ ? - ?
1999	Initiation of the Ticket to Work and Work Incentives Improvement Act.	+

Sources: Social Security Administration, *Annual Statistical Report on the Social Security Disability Insurance Program*, 2001.
Halpern (1979)
Social Security Administration, *Annual Statistical Supplement*, 2001.

TABLE 2: ADF test for unit root in levels, t-statistics with lag selection based on AIC

Variable	Constant		Constant and Trend	
Log of Disability Applications	-2.06	(1)	-2.39	(1)
Female Labor Rate	-2.46	(2)	1.65	(0)
Annual Number of DI Awards	-1.02	(1)	-2.25	(2)
Benefit Replacement Rate	-0.94	(0)	-1.96	(0)
Number of Bed Disability Days	-0.42	(3)	-0.63	(4)
Unemployment Rate	-2.74*	(1)	-2.73	(1)
Number of Workers with Taxable Earnings	-0.64	(2)	-4.84***	(1)

***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

TABLE 3: ADF test for unit root in first differences, t-statistics with lag selection based on AIC

Variable	Constant		Constant and Trend	
Log of Disability Applications	-4.10***	(0)	-4.10**	(0)
Female Labor Rate	-3.16**	(0)	-4.23***	(0)
Annual Number of DI Awards	-3.65***	(0)	-3.61**	(0)
Benefit Replacement Rate	-7.44***	(0)	-7.34***	(0)
Number of Bed Disability Days	-0.52	(2)	-1.51	(2)
Unemployment Rate	-4.94***	(1)	-4.89***	(1)
Number of Workers with Taxable Earnings	-4.77***	(1)	-4.72***	(1)

TABLE 4: VAR lag order selection criteria

Lag	LogL	LR	FPE	AIC	SC	HQ
0	275.6876	NA	8.17E-11	-14.72500	-13.92511	-14.44888
1	326.7074	75.80094	7.59E-12	-17.12614	-15.92630	-16.71195
2	353.9427	35.79498*	2.82E-12	-18.16816	-16.56837*	-17.61591*
3	365.3559	13.04358	2.69E-12*	-18.30605*	-16.30632	-17.61574
4	371.0995	5.579562	3.75E-12	-18.11997	-15.72029	-17.29160

*indicates lag order specified by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

TABLE 5: Lag Exclusion Wald Tests
Chi-squared test statistics for lag exclusion:
Numbers in [ ] are p-values

	LOGDIAPP	BENREPLCE	LOG(WRKTAXE /1000)	Joint
Lag 1	13.79158	11.91294	1.311357	38.31324
	[0.003203]	[0.007687]	[0.726438]	[1.53E-05]
Lag 2	6.645763	13.74608	0.110494	19.13496
	[0.084088]	[0.003272]	[0.990549]	[0.024070]
Lag 3	1.385572	4.544769	0.799213	7.280634
	[0.708920]	[0.208332]	[0.849655]	[0.607927]
Lag 4	4.090379	0.734288	1.799410	6.210060
	[0.251869]	[0.865111]	[0.615063]	[0.718724]
df	3	3	3	9

TABLE 6: Cointegration analysis of the log of disability applications, the log of the number of insured, and the labor rate

Unrestricted Cointegration Rank Test

Hypothesized No. of CE(s)	Eigenvalue	Trace Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.861463	114.1957	29.68	35.65
At most 1 **	0.640259	41.06082	15.41	20.04
At most 2	0.083672	3.233085	3.76	6.65

*(**) denotes rejection of the hypothesis at the 5%(1%) level
Trace test indicates 2 cointegrating equation(s) at both 5% and 1% levels

Hypothesized No. of CE(s)	Eigenvalue	Max-Eigen Statistic	5 Percent Critical Value	1 Percent Critical Value
None **	0.861463	73.13489	20.97	25.52
At most 1 **	0.640259	37.82774	14.07	18.63
At most 2	0.083672	3.233085	3.76	6.65

*(**) denotes rejection of the hypothesis at the 5%(1%) level
Max-eigenvalue test indicates 2 cointegrating equation(s) at both 5% and 1% levels

Unrestricted Cointegrating Coefficients (normalized by b'*S11*b=I):

LOGDIAPP	BENREPLCE	LOG(WRKTAXE/1000)
-6.224476	12.00827	87.02229
-22.48060	52.58618	25.81022
-14.73468	-13.90331	12.89395

Unrestricted Adjustment Coefficients (alpha):

D(LOGDIAPP)	0.008952	0.005603	0.007177
D(BENREPLCE)	0.008352	-0.007567	0.000594
D(LOG(WRKTAXE/1000))	-0.010268	0.000324	0.000853

TABLE 6 Continued: Cointegration analysis of the log of disability applications, the log of the number of insured, and the labor rate

1 Cointegrating Equation(s): Log likelihood 348.0538

Normalized cointegrating coefficients (std.err. in parentheses)

LOGDIAPP	BENREPLCE	LOG(WRKTAXE/1000)
1.000000	-1.929203	-13.98066
	(0.50346)	(0.97190)

Adjustment coefficients (std.err. in parentheses)

D(LOGDIAPP)	-0.055719
	(0.03118)
D(BENREPLCE)	-0.051986
	(0.01227)
D(LOG(WRKTAXE/1000))	0.063913
	(0.00609)

2 Cointegrating Equation(s): Log likelihood 366.9676

Normalized cointegrating coefficients (std.err. in parentheses)

LOGDIAPP	BENREPLCE	LOG(WRKTAXE/1000)
1.000000	0.000000	-74.36575
		(5.97910)
0.000000	1.000000	-31.30055
		(2.64705)

Adjustment coefficients (std.err. in parentheses)

D(LOGDIAPP)	-0.181688	0.402158
	(0.11410)	(0.26385)
D(BENREPLCE)	0.118117	-0.297610
	(0.03097)	(0.07163)
D(LOG(WRKTAXE/1000))	0.056632	-0.106269
	(0.02276)	(0.05264)

TABLE 7: Estimate of the general short-run model using the change in the log of the number of disability applications (Δlog DI applications) as the dependent variable (with standard errors in parentheses)

Independent Variables	Coefficients
Constant	0.80**
	(0.32)
ΔFemale Labor Rate	-0.02**
	(0.01)
ΔNumber of Awards	3.59(e-4)**
	(1.43(e-4))
Δbenefit replacement	-0.53*
	(0.30)
ΔNumber of Days Sick in Bed	-0.01
	(0.02)
Δunemployment rate	-0.01
	(0.01)
ΔWorkers With Taxable Earnings	-9.10(e-4)
	(0.01)
Dummy1967	-0.04
	(0.03)
Dummy1972	-0.03
	(0.03)
Dummy1980	0.03
	(0.03)
ECM_{t-1}	-0.04**
	(0.02)
R^2	0.56
Adjusted R^2	0.33

* = 10% level of significance ***=1% level of significance
** = 5% level of significance

TABLE 8: Estimate of the specific short-run model using the change in the log of the number of disability applications (Δlog DI applications) as the dependent variable (with standard errors in parentheses)

Independent Variables	Coefficients
Constant	0.86***
	(0.22)
Female Labor Rate	-0.02***
	(0.00)
ΔNumber of Awards	2.60(e-4)**
	(1.20(e-4))
ΔBenefit Replacement	-0.60***
	(0.23)
ECM_{t-1}	-0.04***
	(0.01)
R^2	0.47
Adjusted R^2	0.40

* = 10% level of significance ***=1% level of significance
** = 5% level of significance

Figure 1

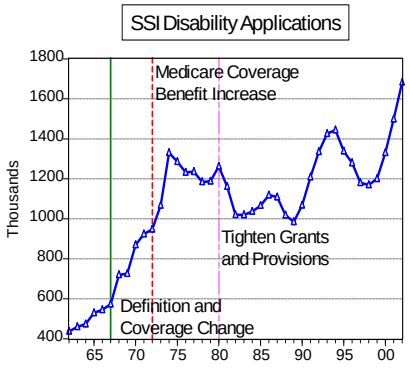


Figure 2: The log of disability applications (left vertical axis) vs. the female labor participation rate (right vertical axis)

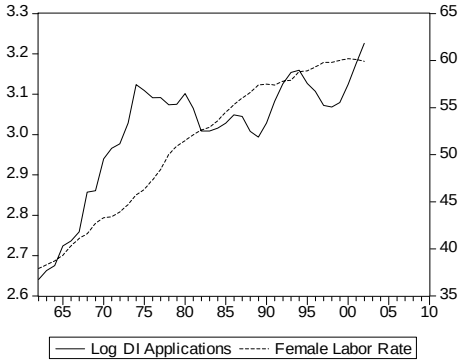


Figure 3: The log of disability applications (left vertical axis) vs. the number of DI awards (right vertical axis)

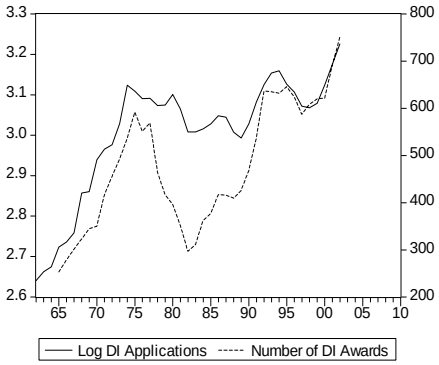


Figure 4: The log of disability applications (left vertical axis) vs. the isability benefit replacement rate (right vertical axis)

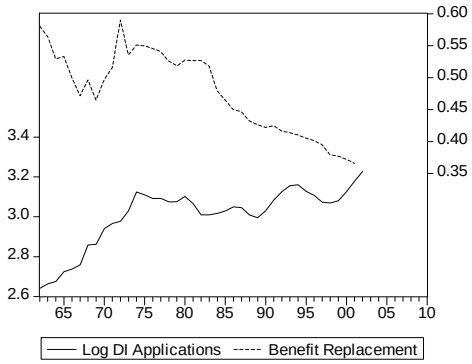


Figure 5: The log of disability applications (left vertical axis) vs. the number of bed disability days per person (right vertical axis)

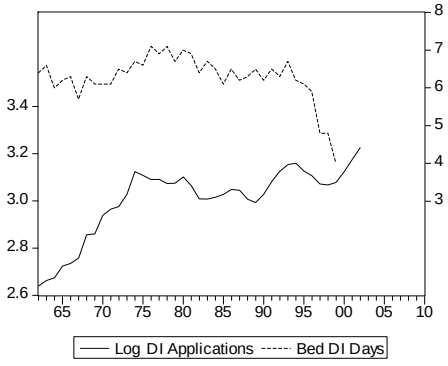


Figure 6: The log of disability applications (left vertical axis) vs. the unemployment rate (right vertical axis)

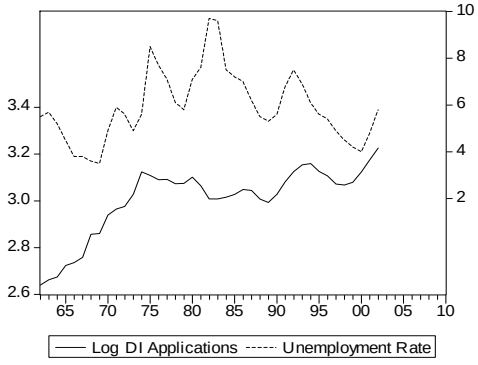


Figure 7: The log of disability applications (left vertical axis) vs. the number of workers with taxable earnings (right vertical axis)

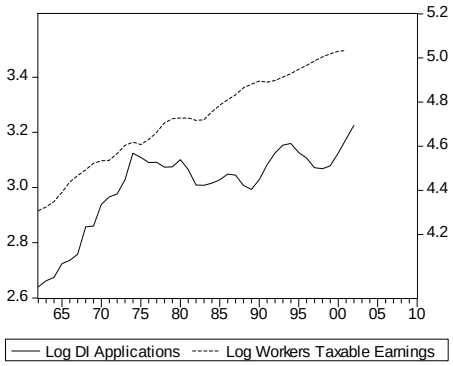


Figure 8: The correlograms of the VAR residuals

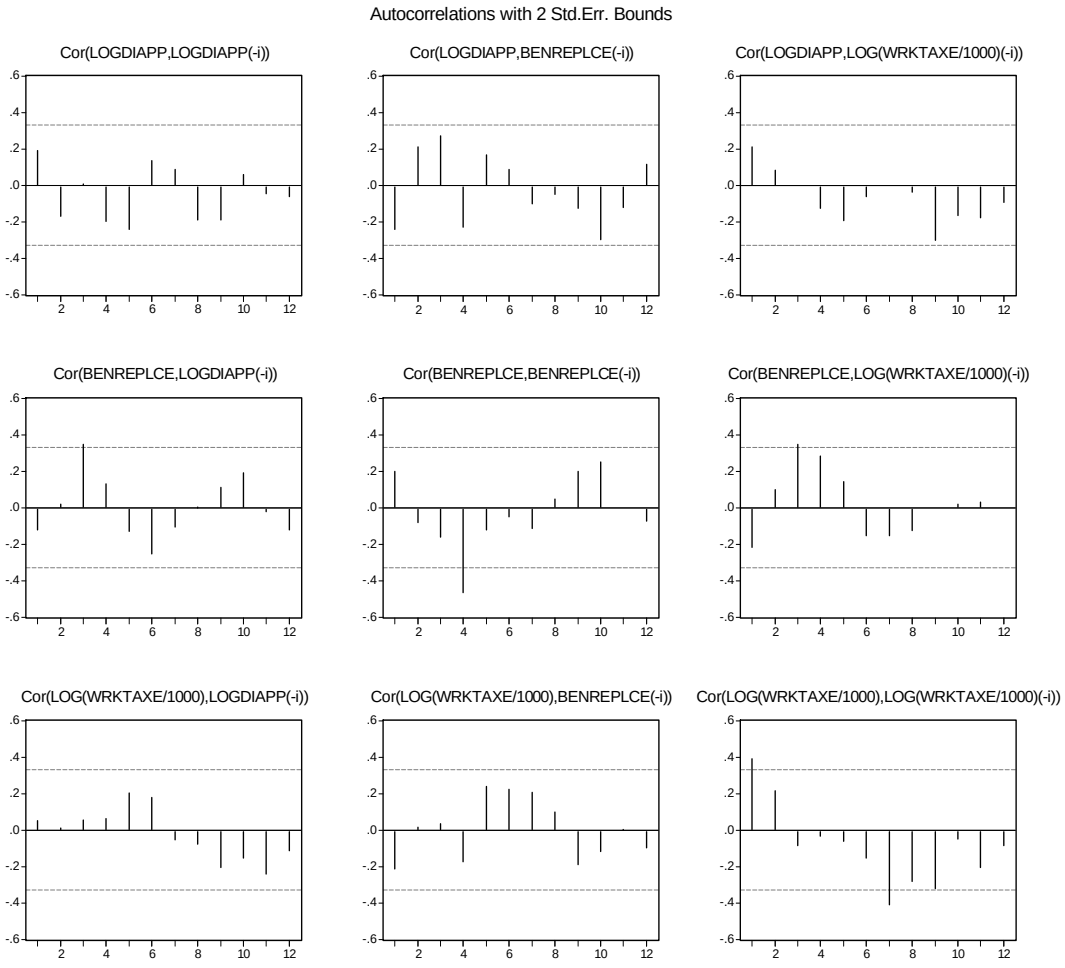


Figure 9: Residual Summary Statistics and Histogram

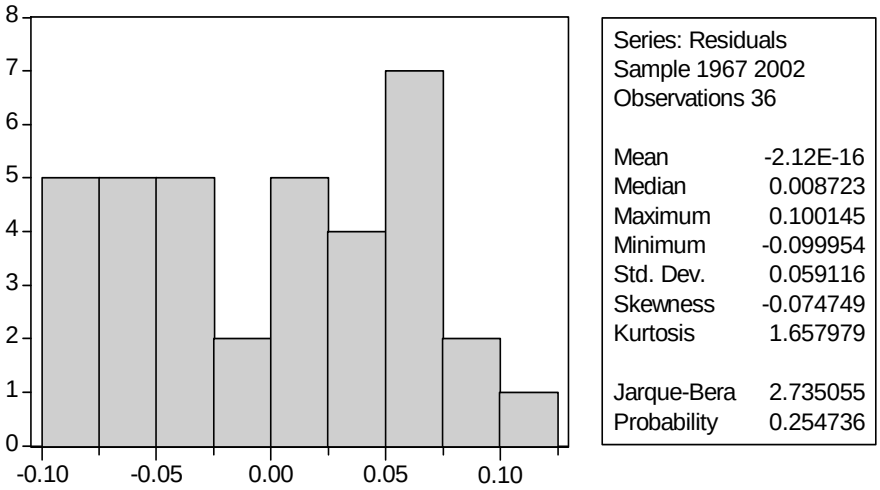


Figure 10:

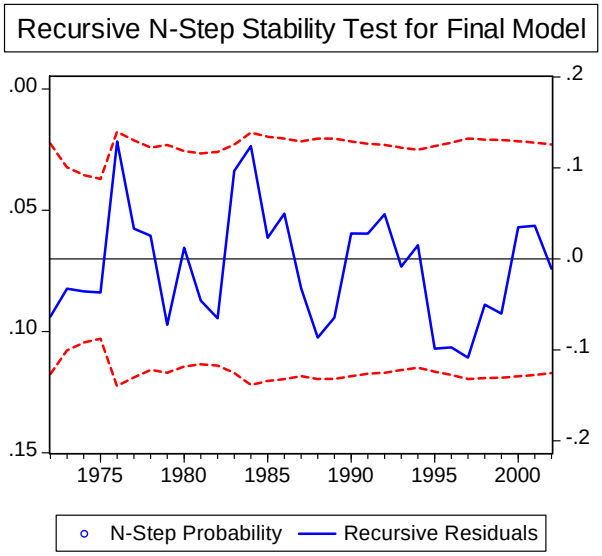


Figure 11:

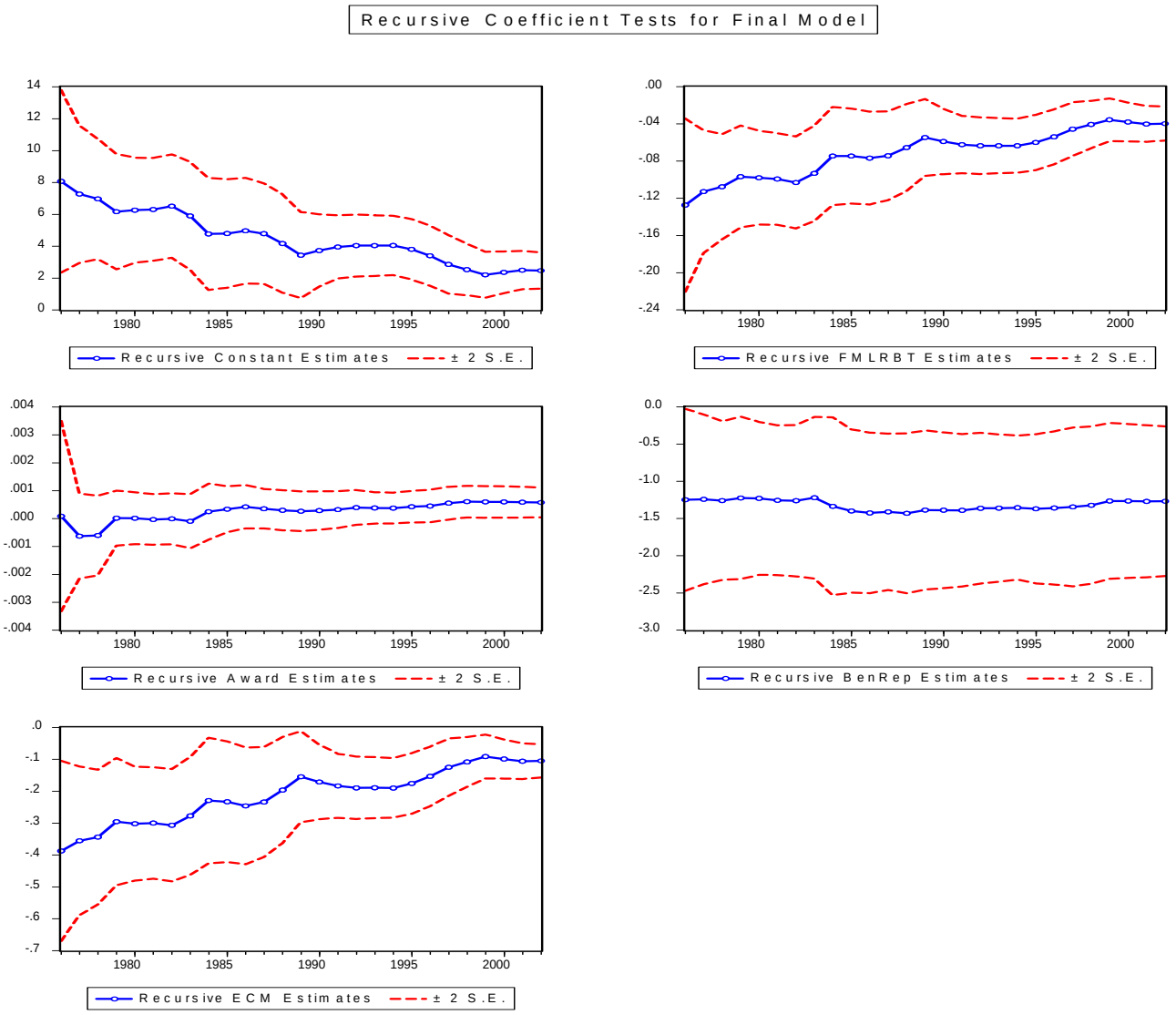
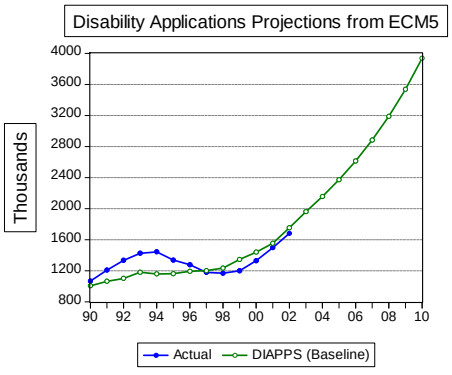


Figure 12:



Transportation Forecasting

Chair: Peg Young, Bureau of Transportation Statistics, U.S. Department of Transportation

FAA Forecasts and Data
Roger Schaufele, Federal Aviation Administration, U.S. Department of Transportation

The FAA provides forecasts of levels of aviation demand and activity at both the national and airport level. These forecasts are important determinants of staffing levels and capital expenditures required to maintain a safe, secure, and efficient environment. The quality and timeliness of the aviation data FAA uses in its forecasting process is an ongoing and important issue to FAA forecasters. This paper provides a brief overview of the aviation data used by the FAA in its forecasting process and highlights both the positive and negative aspects regarding the data. The paper also examines the methods that FAA forecasters employ to maximize the value of the aviation data available.

Estimation of International Trade Traffic Attributes on U.S. Highways

Caesar Singh, Bureau of Transportation Statistics, U.S. Department of Transportation

As the world's largest trading nation, the United States is both the largest importer and exporter of merchandise. With the growth of international trade, the condition and suitability of the nation's freight transportation infrastructure continue to be a transportation challenge. This presentation illustrates procedures that were adopted to estimate freight movement attributes such as ton-miles and value-miles on U.S. highways through use of existing data and application of mathematical modeling techniques. Furthermore, it addresses the reliability and accuracy of these estimates along with data improvement recommendations.

Building the Timetable from Bottom-Up Demand: A Micro-Econometric Approach
Dipasis Bhadra, Jennifer Gentry, Brendan Hogan, and Michael Wells
Center for Advanced Aviation System Development, The MITRE Corporation

The aviation community has a rich collection of tools that simulate the operational flows of the National Airspace System (NAS). In nearly all cases, modeled operational flows of aircraft in the NAS begin with a schedule generated outside of the model. In the past, the schedule has been derived by translating the Federal Aviation Administration's (FAA's) Terminal Area Forecast (TAF) into flights. The downside to this, however, is that NAS operations are made up of specific airport-to-airport flows, which may be different from terminal area growth attributable to those airports. The challenge is to move from a generic traffic count at a specific terminal to a schedule of flights that includes a "when" and a "where" dimension.

Modeled NAS operational performance is highly dependent on the characteristics of the forecasted operations; hence it is critical that the traffic schedule be created correctly. The top-down approach based on TAF projections achieves its goal of replicating the intended volume of flights at each airport, but it does not necessarily achieve the desired operational-level integrity. In other words, the existing method is not capable of forecasting route-specific growth in operational flows.

At the MITRE Corporation's Center for Advanced Aviation System Development (CAASD), we are building a framework which attempts to fill in the gaps mentioned above using a bottom-up, demand-driven micro-econometric approach. Our ultimate goal is to produce a schedule of flights that is linked with origin and destination (O&D) operations via passenger route choice. It should thus be in sync with the Official Airline Guide (OAG), but not driven by it. Our method is comprised of six basic steps, beginning with estimation and forecasts of traveler demand between O&D city pairs, and culminating with the creation of a forecasted schedule that incorporates all major aspects of passenger demand.

Building the Timetable from Bottom-Up Demand: A Micro-Econometric Approach

Dipasis Bhadra, Jennifer Gentry, Brendan Hogan, and Michael Wells
The MITRE Corporation's Center for Advanced Aviation System Development
7515 Colshire Drive, McLean, Virginia 22102

Keywords: Aviation, Decision Support Tools, Timetable, Econometrics, Forecasting

ABSTRACT

The aviation community has a rich collection of tools that simulate the operational flows of the National Airspace System (NAS). In nearly all cases, modeled operational flows of aircraft in the NAS begin with a schedule generated outside of the model. In the past, the schedule has been derived by translating the Federal Aviation Administration's (FAA's) Terminal Area Forecast (TAF) into flights. The downside to this, however, is that NAS operations are made up of specific airport-to-airport flows, which may be different from terminal area growth attributable to those airports. The challenge is to move from a generic traffic count at a specific terminal to a schedule of flights that includes a "when" and a "where" dimension.

Modeled NAS operational performance is highly dependent on the characteristics of the forecasted operations; hence it is critical that the traffic schedule be created correctly. The top-down approach based on TAF projections achieves its goal of replicating the intended volume of flights at each airport, but it does not necessarily achieve the desired operational-level integrity. In other words, the existing method is not capable of forecasting route-specific growth in operational flows.

At the MITRE Corporation's Center for Advanced Aviation System Development (CAASD), we are building a framework which attempts to fill in the gaps mentioned above using a bottom-up, demand-driven micro-econometric approach. Our ultimate goal is to produce a schedule of flights that is linked with origin and destination (O&D) operations via passenger route choice. It should thus be in sync with the Official Airline Guide (OAG), but not driven by it. Our method is comprised of six basic steps, beginning with estimation and forecasts of traveler demand between O&D city pairs, and culminating with the creation of a forecasted schedule that incorporates all major aspects of passenger demand.

INTRODUCTION

The aviation community has developed numerous tools for simulating the operational flows of the NAS[1] [1–3]. Some of these modeling capabilities are quite detailed in approximating the metrics they set out to depict. For example, CAASD's Detailed Policy Assessment Tool (DPAT) measures queuing delays occurring in the NAS throughout the various phases of flight. Taken together, these delays can reach significant levels on a bad weather day. Alternatively, other models have been developed that simulate airline schedule evolution to mitigate the effects of congestion. For instance the National Aeronautics and Space Administration Logistic Management Institute's (NASA/LMI) model provides airlines with a series of actions they can take in response to congestion, including depeaking, off-hours operations, use of secondary airports, and using larger aircraft [3].

In all cases, these tools must be fed a schedule of flights, or timetable, which is generated outside of the model. A typical timetable contains columns for the origin airport, destination airport, departure time, arrival time, equipment type, and possibly the carrier. (Note we use the term "timetable" because it also estimates itineraries for unscheduled traffic.) Table 1 provides an example.

Timetable *input* strongly influences the resulting modeled *output*. For example, modeling an airport with only ten scheduled operations a day will produce

[1]The NAS is a large network of airports and air traffic control facilities (ATC). ATC facilities are classified into three categories: airport towers, terminal radar approach control facilities (or, TRACONs), and air route traffic control centers (ARTCCs or, en route centers). Towers are located at airports and direct airport traffic on the ground and within approximately 5 nautical miles of the airport to altitudes of about 3000 feet. There are 496 towers, of which 266 are under FAA direct control and 230 are managed under contract. TRACON facilities sequence and separate aircraft as they approach and leave airports beginning approximately 5 nautical miles and ending approximately 50 nautical miles from the airport and at altitudes up to about 10,000 feet. En route centers control aircraft in transit and during approaches to TRACONs. The airspace that most en route centers control extends above 18,000 feet for commercial aircraft. At present, there are 22 en route centers.

drastically different results than when modeling the same airport with 1000 scheduled operations a day. The timetable determines the level and general directional flow of these operations. Using a realistic timetable of aircraft operations is therefore a critical component to the operational modeling effort.

In the past, timetables have been derived by extrapolating counts and forecasts of airport terminal operations into individual flights. The FAA measures overall NAS traffic in terms of annual operational counts at each terminal. It publishes a forecast of this traffic each year in the Terminal Area Forecast. Since the FAA forecasts these counts into the future, they make a logical data set to use for growing terminal area traffic, and thus growing any hypothetical schedules as well. The downside to this method, however, is that in reality NAS operations are made up of flows which are associated with a particular location and time of day (similar to a flight plan). This is not necessarily equivalent to extrapolating terminal area growth into individual operations. Thus, the challenge is to move from a generic traffic count at a specific terminal to a timetable of flights that includes a "when" and a "where" dimension.

The top-down approach described above achieves its goal of replicating the intended volume of flights at each airport (i.e., predicted TAF levels), but it does not necessarily achieve the desired operational level of integrity. In other words, the existing method is not capable of forecasting route-specific growth in operational flows. By simply matching terminal traffic forecasts with the OAG schedule (published schedule of flights submitted by commercial airlines), the existing method clearly misses out on rich route-specific information.

The information that is missing in this process is the origin and destination of the passengers on the flights, and information about the routes over which they fly. For example, if some city pairs are expected to experience above- or below-average growth, then some routes (and thus some specific airport pairs) will experience above- or below-average growth. A similar story can be told for hub airports, some of which may add capacity in the near future, and others of which may remain capacity constrained.

PROCESS DESCRIPTION

At CAASD, we are building a framework which attempts to fill in the gaps mentioned above using a bottom-up, demand-driven microeconomic approach [4–5]. Our ultimate goal is to produce a timetable of flights that is linked with O&D operations via passenger route choice and carrier equipment choice. Our output should thus be consistent with the OAG, but not driven by it. Our method is comprised of six basic steps.

Table 1. Sample Timetable

DEP APT (step 2)	ARR APT (step 2)	EQUIP (step 3)	DEP TIME (step 4)	ARR TIME (step 4)
JFK	BOS	AEST	10:30	11:30
JFK	IAD	CRJ1	10:15	11:31
JFK	KIN	A343	10:15	14:05
JFK	LAX	B762	10:10	13:15
JFK	LAX	B763	10:30	13:23
JFK	MCO	B752	10:25	13:11
JFK	PAP	A306	10:00	13:53
JFK	PHX	A320	10:00	13:24

Frequency (step 4) **Block Time (step 4)**

The first step of the process lays the foundation upon which all the other steps will be built. This step determines where people ultimately want to travel. Once we know where people want to go, we use a logit model to determine how to get them there. This second step produces the actual segments that will be listed in our timetable. For instance, a person planning to travel from Seattle to New York may have a stopover in Chicago (see Figure 1). This process translates a single trip into two separate flights, one from Seattle to Chicago, and the other from Chicago to New York.

The third step determines what type of aircraft will be flown on each flight segment. For instance, the flight from Seattle to Chicago may require a different type of plane, because the distance from Seattle to Chicago is twice as great as the distance from Chicago to New York.

The remaining steps assign arrival and departure times to the flights and also take into account flight activity that is not driven by domestic passenger demand (i.e., cargo, international, and general aviation). While the methodology does not encompass all of the complexities airlines must account for when creating their schedules, it does attempt to capture the same passenger demand element that is the primary driver of their schedules.

METHODOLOGY

1. Estimating and Forecasting Domestic O&D Passenger Demand

People fly because they want to go to places for business and leisure reasons. These decisions are primarily driven by local economic and demographic characteristics. In addition, industry characteristics such as fare and market share of major carriers, seasonality, and the structure of airport hubs all play important roles in eventually determining the O&D

demand. Differentiating the NAS by distances, we estimate a set of econometric relationships that define these relationships on O&D data [4].

To estimate those relationships, data from the Department of Transportation's (DOT's) Origin Destination Survey ("10% ticket sample") are matched with local economic and demographic information for each origin and destination airport.[2] Using this information and drawing on well-established econometric methodologies in demand estimation, we estimate O&D passenger demand between city pairs. In this framework, O&D demand is estimated based on local metropolitan variables as opposed to national economic and demographic conditions, and hence called bottom-up demand. Note this is an ongoing process; and as new data become available, we plan to re-estimate these relationships.

Finally, by combining these estimated relationships with commercially available forecasts of local economic and demographic variables, we end up with yield forecasts of passenger flows by O&D metropolitan areas.

2. Assigning O&D Passengers to Routes

We now have forecasts of passenger demand for travel between metropolitan areas. But choosing to travel somewhere is not the only decision a consumer must make. They must also choose how and when they will fly. Unfortunately, good data on passenger flows by *day* or *time* does not exist. However, the 10% ticket sample does have data on passenger routes. This is important, because over one-third of itineraries involve at least one connecting flight. Knowing that a certain number of people want to go from Seattle to New York is only part of the story, and obscures the fact that many of these passengers will change planes in a hub such as Chicago O'Hare or Dallas-Fort Worth. Furthermore, flights through these hubs are filled with passengers going to and from a variety of O&D Metropolitan Statistical Areas (MSAs).

As suggested by [6], we use the following process to convert O&D passenger flows into airport-to-airport (or "segment") passenger flows. First, using data from the 10% ticket sample, we estimate how route characteristics such as travel time, fare, and connections affect the likelihood of passengers choosing a given route from among the set of routes available. We do this using a multinomial logit model. Once the model is calibrated using historical data, the resulting equation is applied to each O&D pair of MSAs. The result is a distribution of O&D passengers among available routes.

After we have passengers assigned to routes, the model goes through all routes and sums passenger counts by segment. Figure 1 illustrates this using a stylized example. As you can see, when completed we have estimated quarterly passenger flows by airport pair.

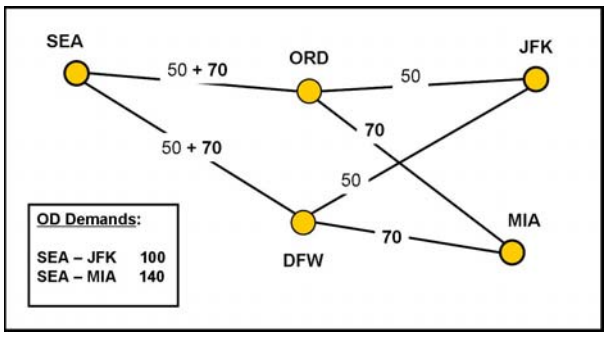


Figure 1. Example of Adding Segment Traffic

As with the calibration of the model, the routes available for passengers to choose from are those which are observed in the actual data. While this does preclude the model from "thinking outside the box" to determine other potentially feasible routes, the converse is also true—we do not end up with too many connecting flights going through small, non-hub airports.

3. Determining Aircraft Equipment Mix

Our third step is critical—taking the airport-to-airport passenger flows and translating total seat demand into the likely set of aircraft types that will fly each route. This step is necessary due to the variation in actual aircraft sizes, which implies that a given number of passengers do not uniquely determine the number of aircraft operations. Estimated passenger counts must be combined with estimated aircraft size to determine a likely equipment mix for each given route.

Choice of aircraft thus emerges as a function of passengers, frequency, trip distance, and other route characteristics. We can therefore estimate a multinomial logit model that enables us to determine the most likely choice of aircraft type. To do this, we turn to DOT's T-100 "Segment" data, which combines historical passenger counts with equipment type, along with flight characteristics such as distance.

An investigation into aircraft utilization over the last 5 years indicates that the most utilized aircraft in

[2]Primary data for this analysis is based on the 10 % O&D sample obtained from the Bureau of Transportation Statistics (BTS) [see http://ostpxweb.dot.gov/aviation for details]. In addition, we use T-100 schedule data collected by the BTS. We combine the O&D travel data with local economic, demographic and spatial variables collected by the Bureau of Economic Analysis (BEA) (see [4] for more details).

the NAS has been narrow-bodies, which transport approximately 60% of scheduled passengers. Narrow-body is a broad classification usually representing single aisle aircraft (e.g., 737 100 through 500 series; and A320). This class of aircraft has a seat range of 90–162 passengers; best cruise speeds at 550–625 miles per hour, and is observed to have been in maximum use for flights ranging from 500 to 750 miles. We have examined actual data and classified the vast majority of aircraft based on these characteristics. In all, there are five natural groupings of aircraft, or categories. Each category has particular performance and capacity characteristics. Category 1 primarily consists of turboprops, that on average, typically fly segments that are less than 250 miles (e.g., SF-340, ATR-42/72, etc.). Category 2 consists of regional jets (e.g., ERJs and CRJs) that fly an average distance of 250-500 miles between MSAs. Category 3 is made up of narrow-bodies (737-100 to 500, A320s, and 727-200, etc.) that fly an average distance of 750-1500 miles between city pairs. Category 4 is the narrow-bodies that tend to fly longer distances, on average 750-1500 miles (i.e., 737 700/LR, A330, etc.). The 5th category is the wide-body category that fly the long haul flights (e.g., 747, 757, 767, 777, L-1011, etc.). As noted, these distances are averages, and many aircraft within one category also travel distances defined under other categories. Each category is also associated with an average number of seats (passenger capacity) and a best cruise speed. Together, the five classifications account for more than 93% of all scheduled passenger activities (see Figure 2).

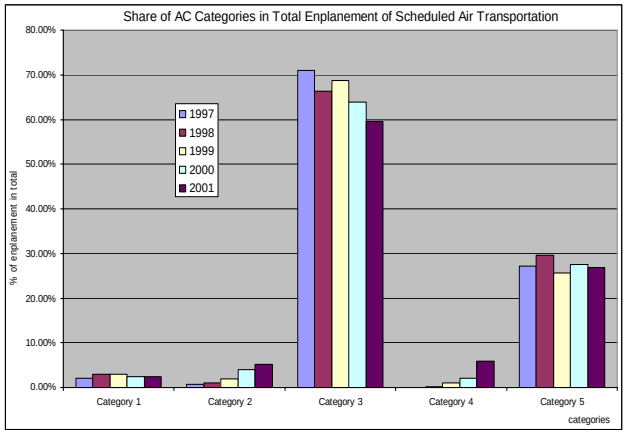


Figure 2. Aircraft Categorization

Modeling aircraft choice, based on historical passenger activities, is a tricky task, especially during a transitional time, like the one resulting from the terrorist attacks on September 11, 2001. The U.S. aviation industry is currently undergoing serious structural changes. At the end of 2002, more than 1400 aircraft were temporarily parked in the Mojave Desert. This is a relatively large percentage considering that the current aircraft inventory consists of only 3623 wide- and narrow-bodies, and around 1020 regional jets (RJs). Routes are being rationalized based on individual profitability in an attempt to improve aircraft utilization. Despite all these changes, the central element of scheduling flights remains intact: carry passengers between two points in the most efficient way.

4. Assigning Times for Scheduled Domestic Flights

Our fourth step determines when the flights will occur. The first part of this task is to determine exactly how many flights will occur. To do this we combine passenger movements between airport pairs (determined in step 2), with the aircraft that are predicted to fly between those airports (determined in step 3), and then apply a load factor. The load factor applied will be specific to each city pair and will be derived from BTS data. For example, if 1000 passengers are predicted to fly between LaGuardia and O'Hare on a given day, in a category 3 aircraft (that holds approximately 133 passengers) with a load factor close to 75%, the resulting frequency is 10 flights a day. This task is critical to the calibration of our model since the frequency calculation determines the total level of operations at an airport. By fine tuning the frequency calculation during the model validation stage, the overall number of operations in the timetable can be adjusted up or down to make it more accurate.

The next task is to take the actual flights and assign arrival and departure times. The timetable will contain commercial operations for 292 airports. The number of airports in the timetable with unscheduled or general aviation (GA) activity will be considerably larger (see step 6). Both commercial and unscheduled departure and arrival times will be assigned using historical data, when available.

Historical data for commercial traffic is obtained from the OAG. This data is then transformed into an arrival and departure distribution of operations over a given day. Several different "days" were obtained; one weekday, and one weekend, from each of the four quarters, for a total of eight representative days. These different "days" represent specific patterns in seasonal and weekday passenger travel.

Current baseline OAG operations will be used as the timetable starting point for scheduled times between city pairs. This data will be processed and altered to accommodate changes in forecasted equipment and international traffic. The historical airport operational distributions will be used to determine time assignment for additional flights. For example, if eight flights

currently operate between LaGuardia and O'Hare each day, and ten frequencies are predicted, then the two additional flights will need to be assigned departure and arrival times based on historical data. The initial eight flights will receive the departure and arrival times that already exist in the current OAG schedule, with some minor adjustments.

To assist in the process of assigning departure and arrival times, the airports have been grouped into four tiers. The airports with the most dominant schedules were assigned to tier 1 (the FAA's capacity critical airports), and the airports with the least dominant schedules were assigned to tier 4 (typically airports only served a couple times a day by just one carrier), with the other airports falling somewhere in between. This categorization will help to determine a flight's scheduled arrival and departure times .

Once an arrival or departure time is established at an airport for a given flight, there is very little "slack" left in the schedule on the other end because the equation—departure time, plus block time[3], equals arrival time—is fairly tight. This causes a problem when a departure is created at an airport that then dictates an arrival at the destination airport during an unlikely time or vice versa.[4] To mitigate this problem, an airport's tier will be used to determine how strictly additional arrivals and departures should conform to the airport's historical distribution of flights. For instance, a hub airport traditionally has strong arrival and departure banks, thus additional flights should conform to those times when banks occur. On the other hand, adding flights to small spoke airports that have only a handful of operations should not necessarily comply with a historic distribution (i.e., If there are only five flights a day at an airport and a sixth is added, that sixth flight should not necessarily be added at a time when flights have historically occurred). Hence when scheduling between tiers, the more dominant airport's schedule will prevail. Intra-tier flights (i.e. flights to and from tier one airports) will undergo additional logic in order to find a departure and arrival time that fits with the airport's historical distributions. This preference will also be enforced by the order in which these flights are assigned in the timetable. Flights between tier 1 airports will be scheduled first, then flights between tier 1 and the remaining tiers. Next flights between tier 2 airports will be scheduled and so on, until all commercial flights have been scheduled.

5. Adding in Scheduled International Flights

Step five adjusts our tentative timetable by taking into account aggregate flows of international passenger and cargo traffic to and from the continental United States (CONUS). Although our modeling focus is the CONUS, we must still account for the additional terminal area traffic, especially important for the international gateway airports, that is generated by flights with only one of two cities in the CONUS. This will be accomplished using a modified top-down approach. All non-CONUS destinations will be associated with growth rates. The rates will then be applied to the number of seats currently being flown to or from those destinations. Applying the growth rate to the number of seats is in line with a passenger demand focus, and also allows for smaller increments of growth.

In 2000, around 26 million passengers traveled to the U.S. from around the world. While a majority (43%) of these passengers originated in Western Europe, the Far East had a respectable 29% share, followed by South America's share of 11%. Almost all of this traffic takes place through 11 gateway airports in the U.S., and therefore, greatly influences the schedule at those airports.

Unlike our O&D model for domestic air travel, here we plan to use regional growth rates from external agencies (e.g., U.S. Department of Commerce International Trade Agency (ITA), FAA) to drive our forecasts of international passenger travel. Using these forecasts and assuming the types of aircraft that are currently flown between these destinations, we can derive the forecast demand for scheduled departures and arrivals. This data will then be added to our domestic schedule.

6. Adding in Non-Scheduled Flights

The last step is to account for unscheduled, or GA traffic. Both the terminal and TRACON handle a large amount of GA traffic. It is estimated that for every scheduled flight, there is another one and half unscheduled operations [7]. There were an estimated 218,000 active GA aircraft in the NAS, which flew almost 40 million operations in 2000. Almost four-fifths of this traffic was in the domain of VFR[5] and thus less likely to crowd the en route air space. However VFR traffic impacts airport towers and TRACONs the same as IFR traffic. Given its significant utilization of

[3] Block time, the time it takes to leave the gate at one airport and arrival at the gate of the destination airport, is largely a function of distance, winds and aircraft speed, but can also be highly influenced by taxi-in and taxi-out times at the arrival and departure airports respectively.

[4] Some time frames will also be unavailable due to constraints on airport operating hours.

[5] VFR stands for Visual Flight Rules. All commercial aircraft are required to fly instrument flight rules (IFR). GA or unscheduled traffic can fly VFR or IFR,

NAS infrastructure, we must include a model of GA traffic in our timetable.[6]

GA traffic is comprised of many different types of operators. Unscheduled business operators tend to file and fly IFR flight plans. Like in the case of O&D travel, particularly the upper end or premium fare travel, IFR flights can be sensitive to economic or financial factors. Specific location and time data for these flights can be derived using the FAA's Enhanced Traffic Management System data.

Data on GA traffic that file and fly VFR flight plans as well as military traffic are not time and location specific, and thus individual operations must be derived using the top-down approach described in the introduction. Based on the historical trends and economic factors, composite growth rates will be applied to both VFR and IFR operations to produce forecasts of activity.

CONCLUSION

This framework is being used to develop a timetable of aircraft operations that will support modeling efforts in evaluating NAS performance. In addition, by changing the inputs, this framework can be used to perform various types of "what if" policy analysis, and thus can stand on its own as a useful analytical tool.

REFERENCES

1. Wieland, F. 1999. "The Detailed Policy Assessment Tool (DPAT)," MITRE Technical Report MTR99W00000012, McLean, VA: The MITRE Corporation.
2. Wieland, F., C. Wanke, B. Niedringhaus, and L. Wojcik. 2002. "Modeling the NAS: A Grand Challenge for the Simulation Community." Proceedings of the Grand Challenges in Modeling and Simulation Conference, San Antonio, TX: 2002 SCS Western Multi-conference.
3. Long, D., D. Lee, E. Gaier, J. Johnson, and P. Kostiuk. 1999. "A Method for Forecasting Commercial Air Traffic Schedule in the Future," NASA/CR-1999-208987, Hampton, VA: NASA/LMI, Langley Research Center.
4. Bhadra, D. 2003. "Demand for Air Travel in the United States: Bottom-Up Econometric Estimation and Implications for Forecasts by O&D Pairs," *Journal of Air Transportation* (forthcoming); paper also presented at the 6th Air Transport Research Society Conference (ATRS 2002) hosted by the Boeing Corporation, July 14 – 16, 2002, Seattle, Wa; and AIAA/ATIO Technical Forum, October 1–3, 2002, Los Angeles, CA.
5. General Accounting Office. 2002. "Commercial Aviation: Air Service Trends of Small Communities Since October 2000, Report to Congressional Reporters," GAO Report No. GAO-02-432, Washington, DC.
6. Oppenheim, N. 1995. *Urban Travel Demand Modeling: From Individual Choices to General Equilibrium.* New York: John Wiley and Sons
7. Federal Aviation Administration. 2003. "Aerospace Forecasts: Fiscal Years 2002-2013," Washington, DC: U.S. Department of Transportation.

[6]Notice that a GA timetable is somewhat *fictitious*. GA traffic, for all practical operational purposes, is unscheduled traffic and hence does not produce timetable on its intent to fly between cities at a given time. It is worth noting here that published schedules are different than the flight plans that IFR GA flights are required to submit. Creation of a schedule, based on their behavior modeled using economic and/or other logic, runs the risk of being truly unreal. Nonetheless, we proceed with this method because of our need to model this entity in simulations of the NAS where they compete with scheduled commercial and non-commercial traffic for scarce air space resources.

Issues and Strategies in VA Health Policy

Chair: Robert Klein, Office of the Actuary, U.S. Department of Veterans Affairs

Introductory Remarks

Robert Klein, U.S. Department of Veterans Affairs

Public Health Care Market Dynamics: The Case of VA Health Care

George Sheldon, Office of the Actuary, U.S. Department of Veterans Affairs

Poverty strongly predicts the use of public health care providers like VA medical centers because means testing regulation directly determines low money prices and indirectly determines high waiting time prices. This paper examines differences observed in the number of low-income veterans found in VA health care market areas throughout the country as measured in 1990 and 2000 decennial census data. We relate these changes to regional changes that have occurred over the 1990s in both patients treated by VA medical centers and VA health care expenditure. We address the marketing question: Is VA investing its capital in markets today where it is most likely to find low-income veteran customers in the future?

VA's Role in U.S. Health Professions Workforce Planning

Dilpreet K. Singh, Gloria J. Holland, Evert M. Melander, Don D. Mickey, and Stephanie H. Pincus
Veterans Health Administration, U.S. Department of Veterans Affairs

In this paper, the Veterans Health Administration's Office of Academic Affiliations (OAA) discusses the 2006 and long-term goals of developing VA's system-wide performance measures for physician residents and other health professions trainees. Also discussed are the impact of the OAA programs, and planning efforts upon the U.S. health care system and upon U.S. health professions workforce planning.

Summary Analysis of Priority 7 Enrollees and Users with Associated Impact on Average Cost Per Enrollee for VA Health Care Services for the Period 1999 to 2002

Surinder S. Gujral, Office of Policy, Planning, and Preparedness, U.S. Department of Veterans Affairs

Decision makers must have access to information on the determinants of demand and supply of VA health care, and how these determinants systematically affect enrollment, utilization, and cost. This is critical for the development of policies related to the provision of uniform access to quality health care services to all veterans at a reasonable cost. A summary analysis of enrollment, utilization, and cost data shows that variability in the growth across VA networks in the numbers of VA enrollees or patients and the associated variability in the changes in costs over time have ramifications for reallocations of resources, for network capacity, and for understanding how supply and demand factors impact cost projections.

Federal and State Medicaid Issues and the Future of VA Health Care

Donald Stockford, Veterans Health Administration, U.S. Department of Veterans Affairs

The reforms mandated by the Veterans Health Care Eligibility and Reform Act of 1996 have revolutionized and modernized VA health care. VA is now a major player in national health care reform discussions and, in particular, in some that focus upon the future of the Medicare and Medicaid programs, which tens of millions of people rely upon or expect to be able to rely upon now and well into the future. This paper focuses on Federal and State Medicaid issues, some of the options being considered, and the current and potential future role of VA as a provider of specialized care and services for at-risk veterans.

SESSION: *ISSUES AND STRATEGIES IN VA HEALTH POLICY*

Introductory Remarks by the Chair
Rob Klein, Office of the Actuary
Department of Veterans Affairs

That healthcare is a critical component of the Department of Veterans Affairs (VA) is only partly indicated by money spent and those who are served. In FY 2002, for example, VA spent nearly 25 billion dollars on health care (including administrative costs), serving 4.7 million veterans.

Behind those numbers are important questions and policy concerns about the kinds of services provided, who gets those services, where and how resources are expended, and what resources and services need to be provided in the future, particularly in light of the VA health care enrollment system, VA's so-called CARES initiative, which aligns resources to demand, and an aging veteran population.

About 10 million veterans 65 or older today make up about 40 percent of the veteran population. Those of advanced age, 85 or older, will increase in number dramatically over the next decade or so. Between 1990 and 2010, for example, there is a projected 8-fold increase of veterans 85 or older, from 163,000 to 1.3 million, reflecting the large World War II cohort and the aging of the Korean Conflict cohort later in the decade. These demographic changes will affect the demand for and mix of VA health care services, notably geriatric health care and long-term care.

The papers today touch on one or more of the broad issues of the demand for services, allocation of resources for VA health care, and the kinds of resources available, now and in the future.

The paper by Singh and her co-authors focuses on VA as a resource for medical education and health care professionals in the U.S. VA is a major player in medical education and will have an impact on the future of medicine and medical education in the U.S.

The Gujral paper provides an analysis of supply and demand factors as they relate to VA health care enrollment, utilization and cost.

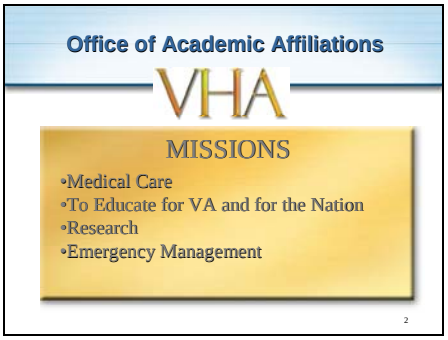
The Sheldon/Gerdes paper looks at strategies for allocating resources across VA market areas based on characteristics of veterans which affect their demand for VA health care services.

And finally, the Stockford paper focuses on prescription drug benefits as a cross-cutting issue for four major health care benefits programs, Medicare, Medicaid, the Defense Department's TRICARE program, and VA health care. A comparison of drug benefits from VA in contrast to the other programs provides insights into VA as a major provider of national health care.

VA's ROLE IN U.S. HEALTH PROFESSIONS WORKFORCE PLANNING

Dilpreet K. Singh, M.S., M.P.A., Gloria J. Holland, Ph.D., Evert H. Melander, M.B.A.,
Don D. Mickey, Ph.D., & Stephanie H. Pincus, M.D., M.B.A.
Office of Academic Affiliations
Department of Veterans Affairs

1. INTRODUCTION



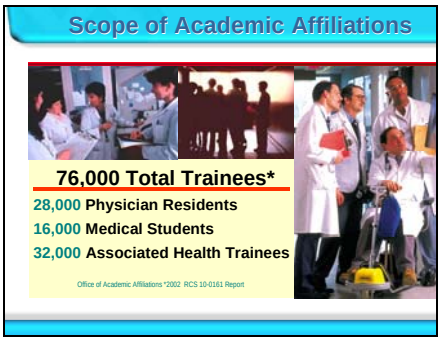
The Department of Veterans Affairs (VA) provides health care for over 4.5 million of the nation's veterans through a network of hospitals, outpatient clinics, and nursing homes. As one of four statutory missions, "To educate for VA and for the Nation," (Slide #2), VA conducts an education and training program for health professions trainees to enhance the quality of care provided to veteran patients. The education and training efforts are accomplished through partnership with affiliated U.S. academic institutions. Affiliations between VA and academic institutions are an invaluable national training resource for students and physician residents.

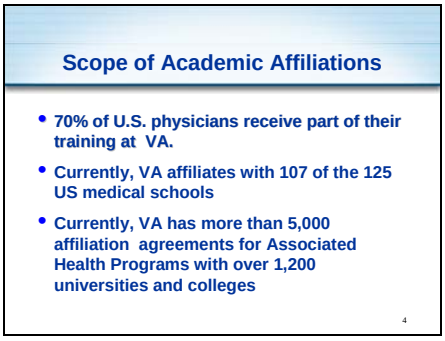
VA is the largest single provider of health professions training in the world. Seventy percent of all physicians and a significant percent of all other health professionals in the United States receive part of their training at VA. As the nation's health care system evolves, VA continues to be on the leading edge with innovative education and training programs that benefit all Americans.

In this paper, the Veterans Health Administration's (VHA) Office of Academic Affiliations (OAA) discusses long-term goals of developing VA's system-wide performance measures for physician residents and other health professions trainees. Also discussed are the impact of the OAA programs and planning efforts upon the U.S. health care system and upon U.S. health professions workforce planning.

2. SCOPE OF CLINICAL TRAINING PROGRAMS (Slides #3 & #4)

Each year, over 76,000 medical and associated health students, physician residents, and fellows receive some or all of their clinical training in VA facilities.





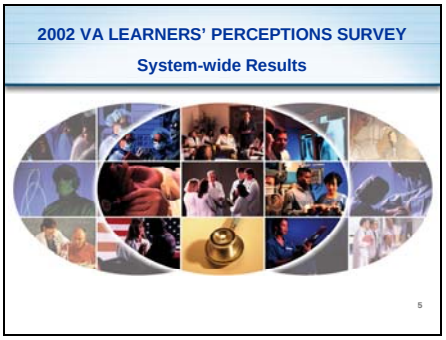
Education of Physicians: VA's medical education program began in the post-war years of World War II. VA's graduate medical education (GME) is conducted through affiliations with university schools of medicine. Currently, 130 VHA medical facilities are affiliated with 107 of the nation's 125 medical schools. Through these partnerships, some 28,000 physician residents and 16,000 medical students receive part of their training in VA every year. VA funding of approximately $ 404 million supports over 8,700 medical resident positions each year. VA physician faculty have joint appointments at the university and VA, see patients at VA, supervise students and physician residents, and conduct research.

Associated Health Education Programs: VA has also been a leader in the training of associated health professionals. Currently, VA has more than 5,000 affiliation agreements for Associated Health Programs with over 1,200 universities and colleges. Through affiliations with individual health professions schools and colleges, clinical traineeships and fellowships are provided to students in more than 40 professions, including nurses, pharmacists, dentists, audiologists, dietitians, social workers, psychologists, physical therapists, optometrists, podiatrists, physician assistants, respiratory therapists, and nurse practitioners. Over 32,000 associated health students receive training in VA facilities each year and provide a valuable recruitment source for new employees. The greatest majority (90%) of associated health trainees receive clinical experiences on a without compensation (WOC) basis. Student funding support of approximately $60 million is provided each year to almost 3,500 trainees.

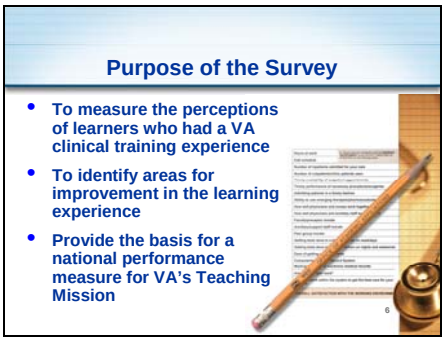
3. NATIONAL PERFORMANCE MEASURE

The Government Performance and Results Act (GPRA) of 1993 required agencies to establish measurable performance goals and develop tools to measure progress toward organizational goals. In support of GPRA, the Office of Academic Affiliations was charged with development of a national performance measure for VHA's teaching mission. The goal was to establish a measure of performance that could be used as a yearly quality indicator to highlight strengths and opportunities for improvement in VA clinical training programs. This paper outlines development, validation, and implementation of a VA system-wide Learners' Perceptions (LP) Survey for all clinical trainees. Information obtained from the survey will help establish performance goals and measure progress toward these goals.

4. LP SURVEY METHODOLOGY (Slides #5 & #6)

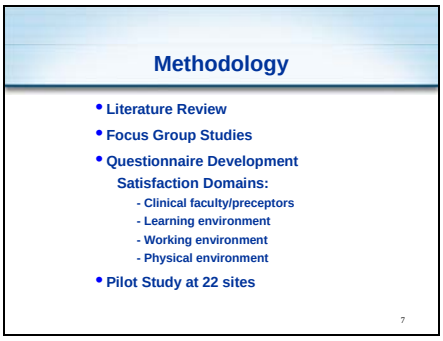


In 1999, a working group was established consisting of the OAA staff, a Steering Committee, and a contractor to oversee this project. (Slide # 5). The working group has representatives with expertise in both multidisciplinary clinical training and survey methodology.



Specific objectives (Slide #6) of the LP Survey are to: 1) measure the perceptions of learners who had a VA clinical training experience; 2) identify areas of excellence and areas for improvement; and 3) provide a basis for a national performance measure for VA's teaching mission.

Literature Review and Focus Group Studies (Slide #7):



To identify items of importance for clinical education, a systematic review of the medical literature from 1975 to the present was conducted in 1999 and updated in January 2002. The search identified 239 articles of which 157 were selected for further review. These studies were graded based on respondent sample size (>50) and response rate. The literature review identified 152 items of importance to clinical training, e.g., workspace, degree of supervision, computer access, etc.

The literature review served as background for 15 focus group sessions held at five VA medical centers during December 1999 and January 2000. Focus group studies were conducted for medical students, physician residents, physician faculty, associated health faculty, nursing students, and graduate and undergraduate associated health trainees. The purpose of the focus group studies was to further explore characteristics of clinical training and validate the themes and items identified through the literature review. VA conducted focus group studies until no new themes emerged.

Questionnaire Development: After the literature review and focus group studies, common and recurrent themes pertaining to attributes of the health care training experience were identified. These themes were collapsed into five conceptually distinct domains, i.e., faculty/preceptors, working, learning, and physical environments, and educational resources. For each domain, a questionnaire was written that asked respondents to rate their satisfaction with the VA training experience using a 5-point Likert scale for most items (very satisfied, somewhat satisfied, neither, somewhat dissatisfied, and very dissatisfied). In addition to collecting demographic information and data about the respondents' programs of study, the questionnaire covered the domains and overall satisfaction with respondents' learning experiences. The complete 75-item questionnaire was developed to rate trainee satisfaction.

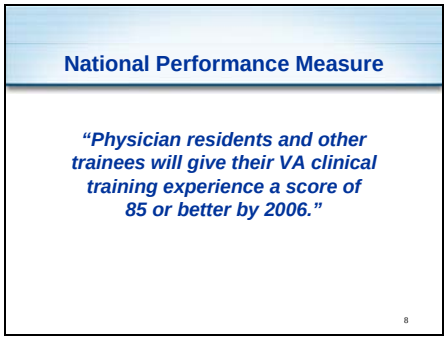
Pilot Study: A pilot test was conducted in 22 geographically diverse VA medical centers to ascertain if the questionnaire could be a valid and reliable tool for measuring the perceptions of clinical trainees. A secondary purpose of the pilot study was to use the results to assist in the development of a system-wide performance measure for clinical training. A total of 1,092 questionnaires were completed and returned. Of these completed questionnaires, 437 (40%) were from residents. The remaining were from other health professional trainees (e.g., nurses, dentists, pharmacists, etc.).

Factor analysis confirmed the grouping of variables into domains and explained the pattern of correlations among the items for each domain. After pilot testing, the five domains were collapsed into four domains and items pertaining to education resources were put into the working environment section. The four domains were: faculty/preceptors, learning, working, and physical environments.

Multiple regression analyses determined which specific items contributed to overall satisfaction for each domain. The 75-item questionnaire was modified to eliminate 18 questions that were demonstrated to be of little value in determining learners' satisfaction. A 57-item questionnaire

was used for nationwide distribution. The final survey was designed to take no more than 15 minutes to complete.

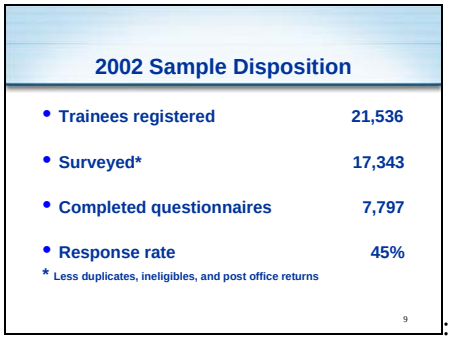
Also, based on the results of the pilot study, a national performance measure was established as a long-term goal for 2006, i.e., "Physician residents and other trainees will give their VA clinical training experience a score of 85 or better by 2006." (Slide #8)

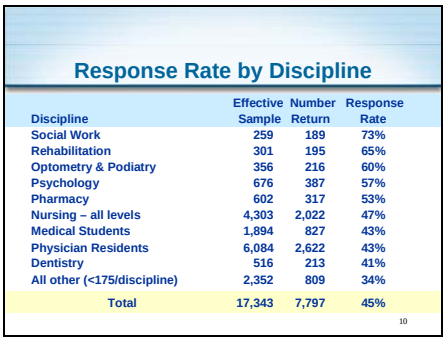


Quantitative assessment of this performance for the past two years has occurred through an annual survey of trainees' perceptions of their clinical training experience. In this paper, focus is on the results of the 2002 LP Survey.

5. 2002 LP SURVEY RESULTS (Slides #9 thru #20)

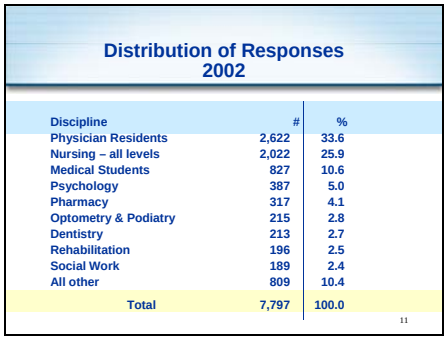
Sample Disposition (Slides #9 & #10)





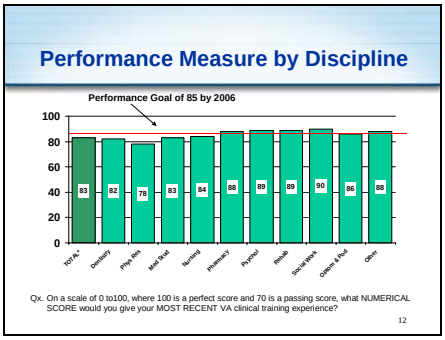
Among all 162 sites that registered clinical trainees, 17,343 trainees were surveyed for the 2002 LP Survey. The survey questionnaire was mailed directly to registered trainees. In order to improve the response rate, non-respondents were sent up to 3 mailings of the questionnaire and 2 mailings of reminder postcards. The questionnaire was also made available on the Internet. The completed survey questionnaires were received from 7,797 registered trainees with a 45% response rate system-wide. Response rates among disciplines such as nursing, dentistry, psychology, etc., varied from 41% to 73%.

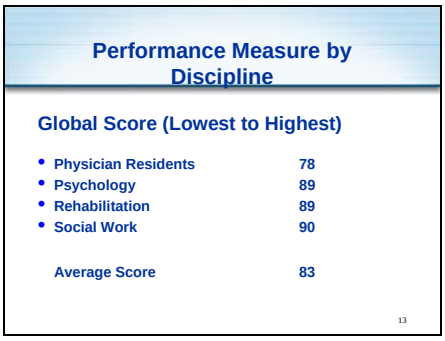
Distribution of Responses (Slide #11):



The distribution of responses varied among disciplines. Of all the respondents, about 34% were physician residents, 26% were nurses, and 11% were medical students. They represented the largest portion of the respondents. For other disciplines, such as social work, rehabilitation, optometry, psychology, etc., the distribution ranged from 2% to 5%. This is a typical representation of the composition of various trainee disciplines in the VA system.

Performance Measure (Slides #12 & #13):

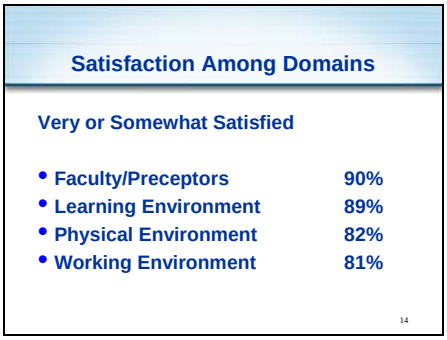


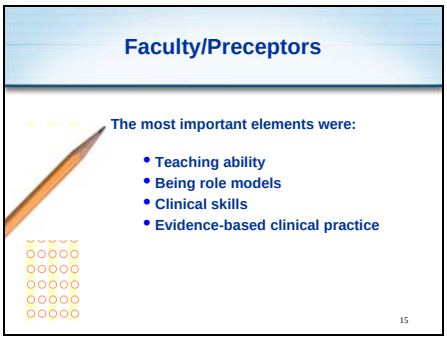


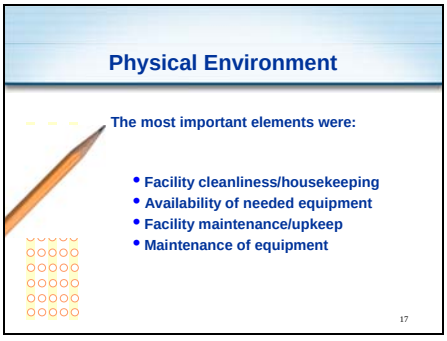
The main outcome measures of the survey were overall satisfaction with the VA clinical training experience and satisfaction with the four domains. Overall satisfaction was measured based on a scale of 0-100 where 100 is a perfect score and 70 is a passing score. For 2002, trainees gave an average overall satisfaction score of 83. A long-term goal of a score of 85 has been established for 2006.

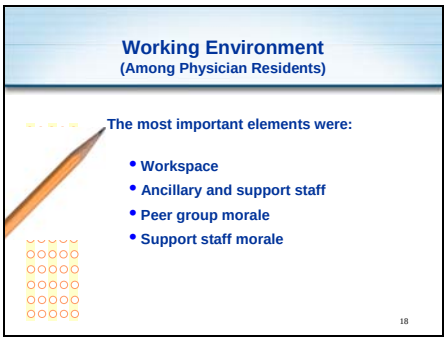
There was variation in scores among disciplines. Among various disciplines of the study, physician residents rated VA clinical training the lowest (78) and psychology (89), rehabilitation (89), and social work (90) the highest.

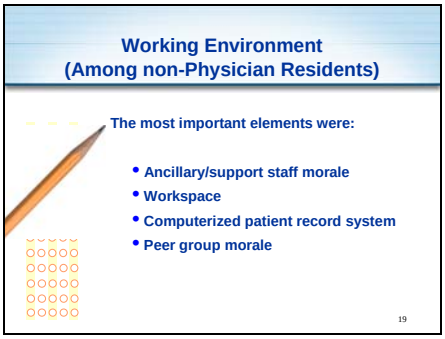
Satisfaction Among Domains (Slides #14 thru #19):







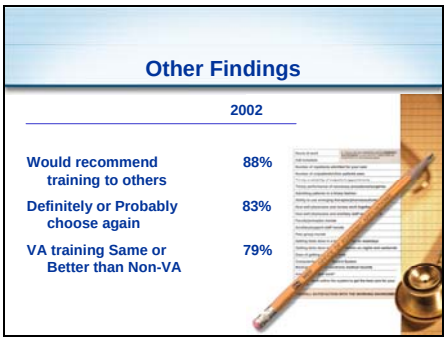




A large majority of trainees were very or somewhat satisfied with the four domains: faculty/preceptors (90%), learning (89%), physical (82%), and working (81%) environments. Statistically, by domain, the most important aspects of training that impacted upon trainee satisfaction were:

- Faculty/Preceptors: teaching ability, being role models, clinical skills, and evidence-based clinical practice.
- Learning Environment: quality of care, preparation for future training, preparation for clinical practice, and time for learning.
- Working Environment (among physician residents): workspace, ancillary and support staff, peer group morale, and support staff morale.
- Working Environment (among non-physician residents): ancillary/support staff morale, workspace, computerized patient record system, and peer group morale.
- Physical Environment: facility cleanliness, availability of needed equipment, facility maintenance, and maintenance of equipment.

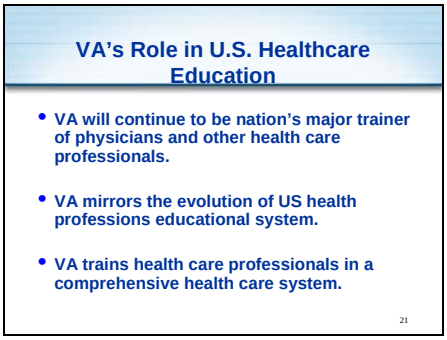
Other Findings (Slide #20):



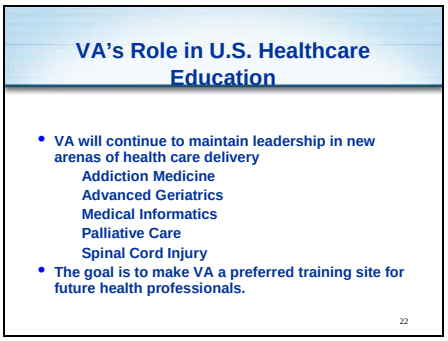
The results of this survey suggest that for all trainees, overall satisfaction with the VA learning experience is high (88% of the trainees indicated that they would recommend VA training experience to peers, and, if given the opportunity, 83% would choose VA training experience again).

Overall, 69% of the trainees rated clinical training as excellent or very good. Comparing VA training experience with non-VA training experience, 79% of the trainees rated VA training as the same or better than non-VA training. For future planning purposes, trainees' ratings will be used as one of the determining factors in allocating resources for training programs at each facility.

6. VA 's Role in U.S. Health Care Education (Slides #21 & #22)



VA will continue to be the nation's major trainer of physicians and other health care professionals. VHA has "partnership" agreements with 107 of the nation's medical schools and 1,200 other colleges and universities offering health professional training programs.



A significant percent of all health professionals and 70 percent of all physicians in the U.S. experience some portion of their training in VHA. More than 76,000 health care professionals receive part of their clinical training in VHA facilities each year.

VA mirrors the evolution of the U.S. health professions educational system and is the second largest financial supporter of education for medical professionals after Medicare.

VHA is much more than a network of medical facilities for sick and disabled veterans. VA health care provides comprehensive care to eligible veterans and trains health care professionals in the total care of the patient.

VA will continue to maintain leadership in new arenas of health care delivery. VHA-based training in addiction psychiatry, pain management, and spinal cord injury medicine is addressing the needs of the nation as well as our veterans. Programs initiated within VHA have led to the development of new medical specialties such as geriatrics, which focuses on care of the elderly.

More than 80% of current trainees highly value their VHA educational experience, and, if given the opportunity, would choose to train in VA again.

The goal of the VA is to make VHA a preferred training site for future health care professionals.

For more information visit Academic Affiliations online at http://vaww.va.gov/oaa/ or http://www.va.gov/oaa/.

SUMMARY ANALYSIS OF PRIORITY 7 ENROLLEES AND USERS WITH ASSOCIATED IMPACT ON AVERAGE COST PER ENROLLEE FOR VA HEALTH CARE SERVICES FOR THE PERIOD 1999 TO 2002[1]

Surinder S. Gujral, Office of the Assistant Secretary for Policy, Planning, and Preparedness
Department of Veterans Affairs

Introduction

From an analytical viewpoint, enrollment for VA care is dependent upon a number of factors such as veteran population, unemployment, health insurance and the availability of alternative care in private sector health care markets. The utilization of VA care by enrolled veterans is similarly dependent upon the age distribution of enrollees, morbidity, health status and the incidence of disease and disabilities. The cost of delivering care, once again, is determined by the mix of resources used by the networks in delivering care. Unless decision makers have access to information on the determinants of demand and supply of VA health care, and how these determinants systematically affect enrollment, utilization and cost, it would be difficult to develop policies that would provide uniform access to quality health care services to all veterans at a reasonable cost.

This summary presents a descriptive analysis of data on enrollment, users and costs across the 21 VISNs for a four-year period from 1999 to 2002. It also analyzes the changes in enrollees, users and cost of care for veterans enrolled in Priority 7 (P7) and Priority 1-6 (P1-6) subgroups during the same period. The findings or conclusions emerging from this analysis are of a gross nature and hence cannot provide satisfactory explanations on factors influencing enrollment, consumption of health care services and cost of care. The analysis however does provide a sufficient basis for additional examination of the inter-relationships among various variables affecting VA health care at the macro and at the micro levels.

An Overview

Veteran enrollees in VA health care increased by 72 percent, from 3.9 million in 1999 to about 6.8 million in 2002. Across the networks, however, the variability in enrollment changes ranged from an increase of 59 percent in VISN 18 to an increase of 89 percent in VISN 23.
All users, on the other hand, increased by 43 percent from 2.9 million in 1999 to 4.2 million in 2002. But the range in variability across networks was greater in users than in enrollees. The users increased by 30 percent in VISN 3 as compared to a 60 percent increase in VISN 23.

The ratio of users to enrollees in all networks decreased from 0.74 in 1999 to 0.62 in 2002. In other words, users per enrollees declined by 16 percent over the same time period. Across networks, the decline in user/enrollee ratio ranged from a 15 percent decline in VISNs 16 and 18 to about a 22 percent decline in VISN 20.

The total cost of VA health care services increased from $15.2 billion in 1999 to $19.6 billion in 2002; about a 29 percent increase over four years, or an average increase of 7 percent

per annum. But the average cost for all enrollees declined about 10 percent during this same period.

How do we explain the changes in enrollment, utilization and cost of VA care during this three-year period, and what deductions can be made about the inter-relationship among these variables? This question cannot be answered in this descriptive analysis. A more sophisticated model is needed to analyze these interactions and to predict the future parameters of supply, demand and cost of VA care.

Priority 7 Veterans' Enrollment and Utilization Relative to That of Priority 1- 6 Veterans

The overview presented above is not very useful in analyzing the impact of the Priority 7 veterans on future enrollment, utilization, and cost of care. Part of the problem is that changes in total enrollment, users and costs pertain to both P7 and P1-6 category veterans, whereas our interest in this analysis is to examine how changes in the enrollment of P7s relative to P1-6s influences the utilization, cost, and availability of VA care.

In examining changes in the enrollment of P7s and their impact on the use and cost of care in relation to P1-6 enrollees, it is important to keep in mind that the two subgroups are different and the constraints that dictate enrollment, utilization, and hence cost, affect these subgroups differently. The enrollment of P7s is drawn from the general veteran population and it is free except for the opportunity cost of the time to the enrollees. Once P7s use VA care and are vested for VA services, the distinction between P7s and P1-6s disappears even though VHA record keeping tracks the P7s separately. Enrollment and utilization of VA care by P7s has a substantial growth potential in the future, as borne out by the data presented below.

The P1-6 subgroup, on the other hand, has to meet certain eligibility requirements for VA care. This is the largest group that was rolled over into enrollment. Barring war and entitlement created by a legislative mandate, this subgroup is not expected to grow beyond its historic growth patterns. This will also be obvious from the data presented below.

The data presented in Table 1 represent a percent change in the ratio of P7s/P1-6s in 2002 to P7/P1-6s in 1999. For example: in VISN 1, the P7/P1-6 ratio in 1999 was 0.243 but in 2002 the ratio was 0.553. The gain in P7s relative to P1-6s in VISN 1 from 1999 to 2002 is simply 0.553/0.243, or a gain of 127.2%.

As is obvious, P7 enrollees increased substantially across all networks and most of the networks experienced increases in P7s relative to P1-6s of more than 75 percent; the largest increases were in VISNs 8 and 16 in the order of 159.7% and 146.5% respectively. The smallest increase was 61.3% in VISN 22.

The data would seem to support the notion that P7s have gained share in the enrollee population in recent years. Unless we understand the factors affecting this change, it is difficult to say if the trend is likely to continue in the future. In the very near future, continuing increases are very likely even if we are not certain about the rate of increase.

The users data, like enrollees, represent an increase in P7 users relative to P1-6 users. There is, incidentally, a high correlation between the growth in users and enrollees.

P7 users relative to P1-6 users also increased from 1999 to 2002 across all networks. The increase ranges from a low of about 70.0% again for VISN 22 to a high of 222.7% for VISN 8. Most (90%) of the networks had an increase of more than 75% over the four years. The increase in users in most VISNs seems to be highly correlated with the increase in enrollees. In VISN 1, the percentage increase in users far exceeds the percentage increase in enrollees.

The range of variability in the growth of users has important ramifications for reallocations of resources and it needs further examination before determining how the reallocations might affect network capacity.

Finally, the average cost per enrollee declines in about half the VISNs from about –2.4% in VISN 22 to –14.7% percent in VISN 12. For VISNs showing an increase in cost, the increase ranges from 0.9% in VISN 7 to 13.9% in VISN 10.

Some of the cost data would seem to be consistent with economic theory. As is shown in Chart 1 below, networks with increasing enrollment and with increasing numbers of users generally show declines in average cost. The

capacity issue becomes relevant when an increase in users leads to an increase in average cost per enrollee.

Questions? Please contact:
Don Stockford
202.273.5112
donald.stockford@hq.med.va.gov

Note

[1] This paper, focusing on FY99-FY02, is an update of an earlier FY99-FY01 version. The FY99-FY01 data and FY02 data updates used in the present version of this paper were obtained by Mike Grindstaff (VA Office of Policy, Planning, and Preparedness). Actual updating of computations, table, chart, and text was done by Don Stockford (VHA Office of the Assistant Deputy Under Secretary for Health).

Table 1

	Rel Chg	Rel Chg	Rel Chg
	P7/P1-6, Es	P7/P1-6, Us	P7/P1-6, Costs
VISN	02/99	02/99	02/99
	Enrollees	Patients	Costs
1	127.2%	186.1%	-13.2%
2	75.5%	90.0%	9.5%
3	62.8%	73.0%	-8.6%
4	99.2%	119.0%	-3.5%
5	125.1%	146.1%	-5.1%
6	72.2%	81.8%	5.3%
7	84.4%	90.2%	0.9%
8	159.7%	222.7%	-16.2%
9	125.6%	119.0%	3.8%
10	80.4%	100.0%	13.9%
11	125.1%	138.8%	-7.2%
12	116.2%	143.9%	-14.7%
15	131.4%	159.7%	-8.1%
16	146.5%	176.5%	-5.2%
17	101.8%	126.3%	4.8%
18	91.0%	106.8%	5.9%
19	79.1%	87.2%	-3.6%
20	95.7%	87.9%	10.4%
21	80.8%	94.5%	10.8%
22	61.3%	70.0%	-2.4%
23	136.9%	149.7%	9.0%
Total	100.2%	121.3%	-3.3%

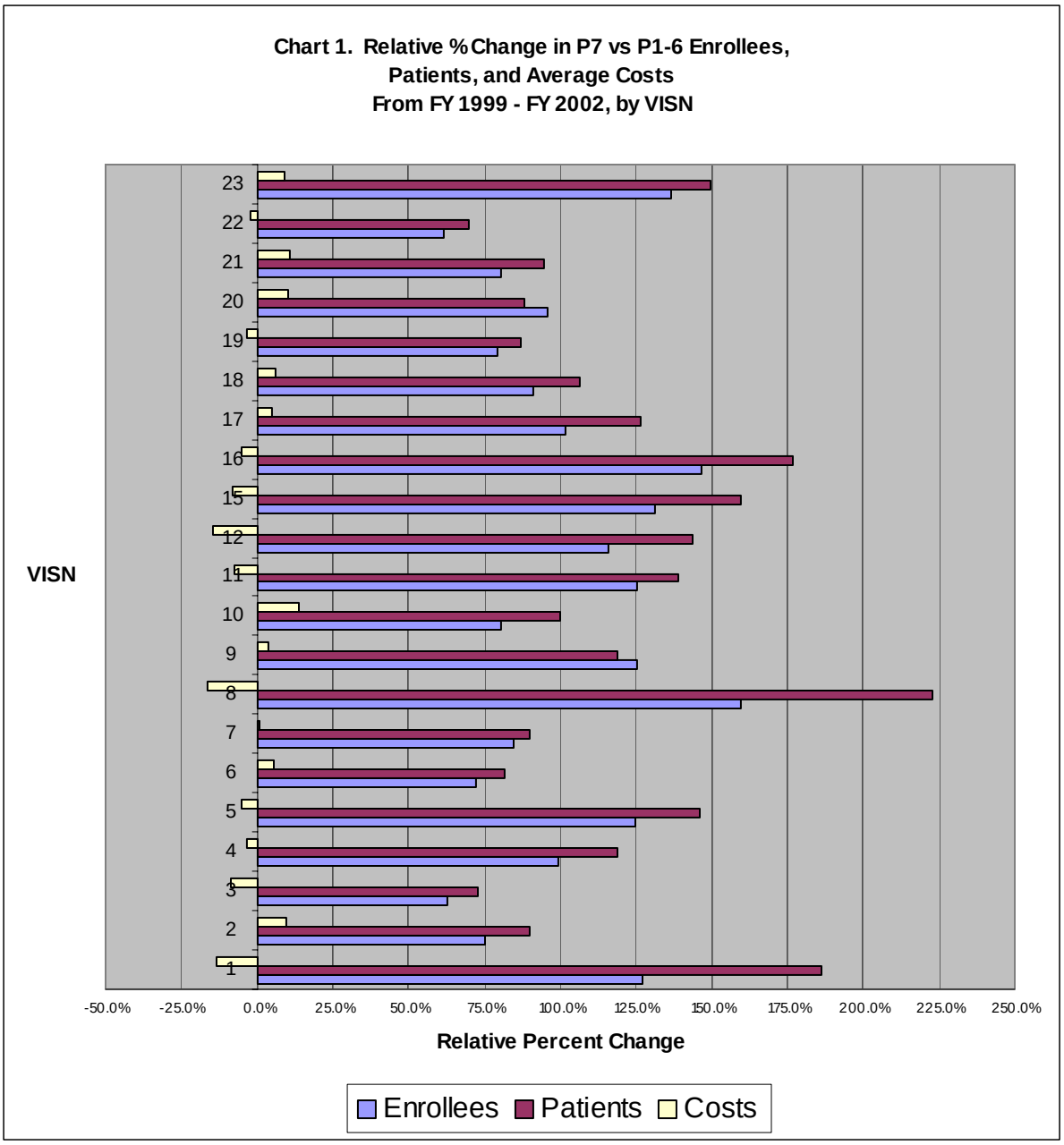


Chart 1. Relative %Change in P7 vs P1-6 Enrollees, Patients, and Average Costs From FY 1999 - FY 2002, by VISN
VISN
23
22
21
20
19
18
17
16
15
12
11
10
9
8
7
6
5
4
3
2
1
-50.0%
-25.0%
0.0%
25.0%
50.0%
75.0%
100.0%
125.0%
150.0%
175.0%
200.0%
225.0%
250.0%
Relative Percent Change
Enrollees
Patients
Costs

FEDERAL AND STATE MEDICAID ISSUES AND VA HEALTH CARE

Donald Stockford, Office of the Assistant Deputy Under Secretary for Health
Department of Veterans Affairs

Multiple Eligibilities

At this moment, House and Senate versions of a Medicare outpatient drug bill (HR1 and S1) are in conference committee. There are many issues which are potential stumbling blocks to any Congressional agreement on Medicare drug legislation, and one of them is the issue of dual Medicare/Medicaid eligibles, particularly who should pay for their outpatient drug needs. There are well over 6 million Americans who are dually Medicare/Medicaid eligible, and Medicaid now pays for their outpatient prescription needs. However, there is a provision in the House bill to the effect that the proposed Medicare drug program will cover dual eligibles; there is no such provision in the Senate bill. This situation marks philosophical differences within Congress in the two Medicare drug bills now in conference committee, but all fifty State governors back the House bill on the issue.

Multiple eligibilities is an issue that Congress finds it must address as it considers Medicare outpatient drug legislation. In fact, multiple eligibilities is an issue in a more general sense, when considering virtually any type of health care service. In particular, there are Americans who are eligible for Medicare and Medicaid and VA care, and/or possibly, as well, TRICARE, the health care program administered by the Department of Defense for active duty and retired military and their survivors or dependents. As important as Medicare is to seniors, premiums, cost sharing, and service gaps such as in regard to drug benefits place significant health cost burdens upon seniors, and helps to point out the significant role VA care plays in helping to fill coverage gaps in other Federal or private health care programs or health insurance. Indeed, VA is filling gaps in Medicare, Medicaid, and TRICARE. This paper focuses on Medicaid and VA care, with pharmacy as an example of a cross-cutting health care service issue.

Veterans Health Care and Enrollment

The veterans health care eligibility and medical benefits reforms of the past decade have revolutionized and modernized VA health care. VA is now a major player in national health care reform discussions and, in particular, in some that focus upon the future of the Medicare and Medicaid programs, which tens of millions of people rely upon or expect to be able to rely upon now and well into the future.

The VHA Health Care Enrollment System, legislatively mandated in 1996 and first implemented as of October 1, 1998, established a 7 category hierarchical system for prioritizing veterans' eligibility for VA health care, with service disabled or low income veterans having higher priority for care than non-service disabled or high income veterans. There are now over 7 million veterans (more than a quarter of the total veteran population of 26 million veterans) enrolled in the VHA Health Care Enrollment System.

VA clientele are largely aged and increasingly so. About half (3.5 million) of all current VHA enrollees are age 65 or older (i.e., Medicare age-eligible); 92% of those are actually on Medicare, while 8% are not. The 50% of all enrollees who are age 65 or older is a higher percentage than the percentage of 65 or older

veterans among all veterans (which is about 38 percent) -- and the percentage of all enrollees who are age 65 or over has been increasing. About 15 million U.S. males age 65 or older are on Medicare, about 10 million (two-thirds) of them are veterans, and about one-third of those are VHA enrolled. Many enrolled veterans (either under or over age 65) are also on Medicaid or have TRICARE eligibility.

Many veterans who use VA are aged, low income, disabled, and/or minorities. As such, many VA enrollees make use of any multiple eligibilities (VA, Medicare, Medicaid, TRICARE-For-Life, IHS, or private insurance, etc.) to cover gaps or perceived gaps in their public or private health insurance coverage. However, many VHA enrolled veterans have no health insurance coverage at all, either pubic or private, and VA care is the only health care option many of them sense they have. Considering the age and other characteristics of veterans and of VA enrollees, it is fair to say that VA provides care and services to a large segment of an elderly U.S. population at–risk of falling through the cracks in the U.S. health care system. In particular, and as just one example, VA has a relatively generous drug benefit for eligible veterans. Without a Medicare outpatient drug benefit for seniors, VA drug benefits effectively fill a major gap in the elderly U.S. population's pharmacy needs.

The fact that many elderly persons make use of multiple eligibilities to maximize their health care resources and reduce costs underscores the significance of eligibility and benefits integration, and the need for coordination of eligibility, benefits, and health care across multiple systems.

Medicaid

Like Medicare, Medicaid (Title XIX of the Social Security Act, 1965) has a 38-year history. Medicaid is a joint State/Federal entitlement program for persons with low incomes and limited resources. Medicaid is an important complement to Medicare, providing coverage for prescription drugs, long-term care, and other services that Medicare largely does not cover. Within broad Federal guidelines, each State administers its own program, establishes its own eligibility criteria, its own breadth and scope of services, and payment rates. As such, there is wide variability across States in Medicaid eligibility, service, and payments. Some States are innovative in regard to the extent that they expand coverage beyond Federal minimum standards and to optional populations, but most are not. See the Urban Institute report, "*States as Innovators in Low-Income Health Coverage*" to see States ranked according to Medicaid "innovation" or "generosity" (www.urban.org).

Not all poor persons are provided Medicaid coverage, only those in special groups who are also tested against State threshold levels for other resources. However, to be eligible for Federal matching funds, states must cover certain mandatory, Categorically Needy (CN), groups, including individuals receiving Temporary Assistance for Needy Families (TANF) benefits [TANF, created by the Welfare Reform Law of 1996 (Personal Responsibility and Work Opportunity Reconciliation Act of 1996, or PRWORA), became effective July 1, 1997 and replaced the "Aid to Families with Dependent Children" (AFDC) program as well as the Job Opportunities and Basic Skills Training (JOBS) program]; children under age 6 and pregnant women in families below 133% of Federal poverty level (FPL); Supplemental Security Income (SSI) recipients in most States (some states are more restrictive due to pre-SSI mandates); certain Medicare beneficiaries; and a few other groups.

The basic Medicaid enrollment groups are: children and their parents, the elderly, and the blind or disabled. Thus, Medicaid is largely for women and children, or for the aged, blind or disabled, and although children may be blind or disabled, most veterans (of whom 94% are males) who have Medicaid eligibility would fall into the aged, blind, or disabled groups. Disabled SSI beneficiaries and elderly SSI beneficiaries are mandatory Medicaid coverage groups. Many veterans with Medicaid coverage would fall into one of these two mandatory groups. As such, veterans' Medicaid eligibility is not likely to be impacted by recent Medicaid reforms, which focus predominantly on trimming State Medicaid optional groups that were added in the 1990s economic boom.

States may also provide Medicaid coverage to certain optional CN groups who share some of the characteristics of the mandatory CN groups and for whom the States may also receive Federal matching funds. The "Medically Needy" (MN) is one such group. The MN would be Medicaid eligible under one of the mandatory or optional groups, except that their incomes/resources are above State thresholds, and States may restrict eligibility and or benefits for them. The MN can qualify immediately or "spend-down" to meet their State's MN income eligibility level by deducting incurred medical expenses from income.

In order to receive Federal matching funds, State Medicaid programs must provide certain basic services, including: inpatient and outpatient hospital services; prenatal care, physician services; nursing facility services for persons age 21 or over; home health care for persons eligible for skilled nursing services; vaccines for children, etc. States may also receive Federal matching funds for the following optional services: diagnostic services; clinic services; rehabilitation and physical therapy services; and home and community-based care to certain persons with chronic impairments, etc.

States determine the amount and duration of Medicaid services, within broad Federal guidelines. States may pay for services either on a FFS basis or through managed care (e.g., HMO) prepayment arrangements. See "*Medicaid Managed Care Enrollment Report*" (www.kff.org), regarding Medicaid managed care market penetration by State. Most Medicaid is now managed care.

Supplemental Security Income (SSI) Summary; and "Medically Needy" and "Spend-Down"

The Supplemental Security Income Program (SSI) is a Social Security program, financed through general tax revenues, that pays monthly benefits to people who are 65 or over, blind, or disabled and who have little or no income or resources. Children as well as adults can receive SSI. Medical requirements and disability determinations are the same under both Social Security Disability Income (SSDI, under which disabled workers receive Medicare benefits) and SSI, but eligibility for SSI is based upon financial need, and persons may be eligible who have never worked nor paid FICA taxes, while eligibility for SSDI is based upon prior work under Social Security.

The amount of income that qualifies one for SSI varies by state. Basic SSI payment amounts are the same nationwide, but many states add to the basic benefit. In 2003, SSI pays $6,624 per year for an individual, or $9,948 for a couple, and many states add to the basic amount.

People on SSI may also get Medicaid, food stamps and other social services. If SSI recipients are eligible for Social Security, they must apply for it, and, if they are disabled, they must accept vocational rehabilitation services if they are offered them. SSI qualifying annual income is about $8,000 nationally.

Pathways to Medicaid for Medicare Beneficiaries

Medicare beneficiaries can obtain Medicaid eligibility through different "eligibility pathways" and the types of Medicaid assistance vary accordingly.

- SSI beneficiaries comprise a mandatory low-income Medicaid eligibility category. (Some states, known as 209(b) states, have more restrictive eligibility standards; SSI beneficiaries do not automatically qualify in 209(b) states.)
- For persons who get Medicare and have low income and few resources, their state may pay their Medicare premiums and, in some cases, other Medicare expenses such as deductibles and coinsurance. The "Medically Needy*" or "Spend-Downs" (both discussed above) get full Medicaid assistance (i.e., "wraparound" Medicaid benefits, Part B premiums, and cost-sharing). Most dual enrollees qualify for SSI or have incurred nursing home costs and get this comprehensive protection. (*MN: States can provide Medicaid coverage to otherwise eligible persons above the income eligibility level set by the state. Persons can qualify immediately or "spend-down" to their state's MN level by incurring health care expenses that they can deduct from income.)
- For Medicare beneficiaries with more income or resources, Medicaid's assistance is more limited, primarily covering premiums. This type of assistance is known as "Medicare Savings Programs" or "Medicare Buy-in Programs", and the beneficiaries are called "Qualified Medicare Beneficiaries (QMB)", "Specified Low-Income Medicare Beneficiaries (SLMB)", and Qualifying Individuals (QI)".
- The QDWI, or Qualified Working Disabled Individuals earning less than or equal to 200% of the FPL comprise another optional Medicaid eligible group; the QDWI generally qualified for Medicare under disability rules before returning to work still disabled.
- SSI beneficiaries generally earn less than or equal to 73% of the FPL; QMBs <=100% FPL; SLMBs, 100% - 120% of FPL; and QIs, between 120% and 175% of the FPL (QIs consist of two groups: QI-1's, 120% - 135% FPL; and QI-2's, 135% - 175% FPL).
- Aside from income tests, there are also asset tests for Medicaid eligibility (depending on program, about $2,000-$4,000 for individuals, and $4,000-$6,000 for couples). The Medically Needy can "spend-down" to state income standard by incurring medical expenses they can deduct from income, but they cannot spend-down or dispose of resources to meet state asset (resources) tests.
- Disabled SSI beneficiaries and elderly SSI beneficiaries are mandatory Medicaid coverage groups, and many veterans with Medicaid coverage would fall into one of these two mandatory groups. SSI is the gateway to Medicaid coverage for many financially needy aged, blind, or disabled veterans. VA care is a safety net in addition to Medicaid for many SSI veterans. SSI beneficiaries under 65 may also qualify for SSDI and, therefore, Medicare. Medicaid covers all prescription costs for those who are dually Medicare/Medicaid eligible. (Note: there are VA/Medicaid/Medicare triple eligibles under SSI).

<u>Question</u>: Who are the "dual (Medicare/Medicaid) eligibles"?

<u>Answer</u>: Medicaid beneficiaries on SSI, or Medicare beneficiaries on SSDI, or those who have exhausted their resources paying for health and long-term care, i.e., the "medically needy" or

"spend-downs". (Note: there are VA/Medicaid/Medicare triple eligibles under SSI/SSDI).

Pharmacy Plus

States can test new approaches to publicly supported health care by obtaining waivers of statutory requirements and limitations from the Department of Health and Human Services. Section 1115 is a research and demonstration authority, which permits States to waive Federal Medicaid statutory and regulatory requirements to extend prescription and over-the-counter pharmacy coverage to certain low-income elderly and disabled individuals who are not otherwise eligible for Medicaid. States can get Federal matching funds for Pharmacy Plus programs. There are presently four states with approved Pharmacy Plus initiatives (SC, WI, IL, FLA), and nine others pending approval (AR, CT, DE, IN, ME, MI, NJ, NC, RI).

State Pharmacy Assistance Programs

State Pharmacy Assistance Programs (SPAPs have been around in one form or another for 25 years. SPAPs are pharmacy programs for low income elderly and disabled individuals who are not on Medicaid.

Medicaid: Some Current Proposals, Issues

Outpatient prescription coverage is an "optional" benefit that all state Medicaid programs now provide, and Medicare does not now offer an outpatient prescription benefit (but covers drugs provided during the course of inpatient treatment). Medicare coverage of prescription drugs could produce major savings for State Medicaid programs, which are jointly funded by the States and the Federal government. In particular, the National Governors Association and the National Conference of State Legislators both want Congress to include a "Dual Eligibles" (Medicare/Medicaid) provision in any final Medicare drug bill to the effect that Medicare will cover drug costs for dual Medicare/Medicaid eligibles (the House version now includes and the Senate version now excludes such a provision). However, the Bush administration continues to advise Congress that it supports the Senate bill's position in terms of drug benefits for Medicare/Medicaid dual eligibles, under which duals would continue to receive drug benefits under Medicaid. The Administration says it would prefer to spend money to cover new people with new benefits rather than substitute Federal dollars for State dollars. The focus on Medicare drug benefits for seniors means that little in the way of national level Medicaid reform will happen soon, and Medicare coverage for prescription drugs for dual eligibles may well be the stumbling block to any Medicare prescription drugs bill this year.

In recent times the State Governors have been urged to back the Bush Administration on Medicaid reform. It has been well-reported that the Administration plans to give States the power to expand, reduce, or eliminate benefits and eligibility for millions of low-income elderly or disabled people. The formal plan, announced Spring 2003, is called "**State Health Care Partnership Allotments",** and under the plan, states can either run Medicaid as they do now or opt for annual allotments, which to some Medicaid experts seem like "block grants", or fixed amounts of money earmarked for a particular purpose. States would be granted new flexibility in the design of their individual Medicaid programs. In return for this flexibility, the States would receive a fixed amount of Medicaid funds, set by statutory formula, over each of the next 10 years. That is, the amount would increase or decrease according to formula, and with medical costs, etc., as well as with levels of future appropriations. Included in the proposed plan are capped federal payments and Maintenance of Effort (MOE) requirements on the part of states. However, under the plan, states would no longer have to apply for Federal waivers to deviate from federal eligibility and benefits standards. Also, states would only have to maintain comprehensive Medicaid coverage for those whose income levels are low enough that the federal government mandates that they be covered.

States currently provide 43% of Medicaid funding and the Federal government 57%. This "Federal Medical Assistance Percentage" (FMAP) is determined annually and based on a comparison of state average per capita income and national average income). Under the new plan, eligibility rules and benefits could change for the two-thirds of Medicaid beneficiaries who are in optional coverage categories, and States would still have to provide coverage to optional and mandatory groups with their annual Federal allotments.

States also want the Federal government share of long term care costs to increase, but under the proposed plan, Federal spending for nursing home care would be capped, and States will be pressured to expand more into home and community-based long term care services. However, costs of doing so are a barrier to States. Home care can easily be more expensive than nursing home care, such as when round the clock nursing assistance is needed, and home care programs require a lot of home care staff.

The Squeeze on the States

Through the early 1990's, many States became innovators in seeking to broaden and expand coverage for the uninsured and at-risk populations. This was, to some extent, reflective of the early 1990s' push towards national health insurance reform, but it was also cotemporaneous with economic good times. In fact, many States became

very creative, fiscally, as well as politically, and became great innovators in terms of expanding Medicaid coverage beyond Federally mandated minimums. The States mostly extended optional groups to cover more women and children, but they also extended long term care services (e.g., nursing home care) to optional groups, such as the elderly or disabled above SSI thresholds, which, although noble, have led to recent budget woes. Long term care accounts for 58% of optional State spending, compared to 10% for drugs, and 32% for acute care services. Alternatives to nursing home care such as home care can be very expensive, too, and the States are now demanding that the Federal government pick up more of the costs of long term care.

Medicaid (and drug) coverage is being scaled back because of worsening state budgets. Many of the frailest and sickest seniors clearly lose. State Medicaid programs must cover the elderly and disabled up to certain income levels, now $6,620 per year (74% of FPL); States with higher ceilings will scale back, so fewer seniors will qualify. Lower/fixed income seniors who qualify for Medicaid will have to use other resources to pay for drugs (such as retirement fund accounts, etc.) States, however, are implementing a variety of reforms (see below). However, veterans and VA would be impacted hardly at all by the Governors' plans to curtail Medicaid coverage or benefits, as most veterans on Medicaid are in mandatory (i.e., SSI or nursing home eligible) Medicaid coverage groups.

State Actions to Control Pharmacy Costs

A recent 50-state survey (Kaiser Commission) indicates that nearly every state faced budget shortfalls in fiscal year 2003 (July 2002 – June 2003) and that a large majority of states were taking and/or planned a variety of actions, including ones aimed at controlling drug costs, including: reducing payments for drug products; subjecting more drugs to prior authorization; implementing or expanding preferred drug lists; mandating the use of generics; imposing new or higher co-payments; imposing new limits on numbers of prescriptions.

Medicaid Disease Management (and Care Coordination) Programs

About 25% of Medicaid beneficiaries have chronic illnesses, but treatments for chronic illnesses account for about 75% of Medicaid spending. Certain chronically ill Medicaid recipients are being monitored under amendments to Section 1915b waivers (to ensure that patients are taking their medications properly and are using preventive strategies) with the goal of reducing expensive ER visits and hospital stays. Florida is a testing ground for new "disease management programs" case management programs in which health plans, State Medicaid agencies, and "Disease Management Organizations (DMOs) manage chronic diseases and keep costs down. There are, however, major criticisms of the role of private companies in disease management. Disease management programs may monitor referrals from state agencies based on diagnosis–related claims data or based on physician referrals. Also, individual pharmacists, or Certified Disease Educators (e.g., Diabetes), etc., may contract to do some of the actual patient monitoring.

Medicaid issue(s): A Medicare drug benefit might be more attractive to Medicaid dual eligibles because there are barriers to Medicaid enrollment and states vary in the breadth and depth of prescription coverage.

VA Suspension of Priority 8 Enrollments

The recent (FY 2003) 7% VA budget increase was obtained basically in order to maintain the current level of VA services. Market place changes that might lead to more veterans coming to VA have not been factored into the increase.

On the other hand, if more veterans do come to VA for any reason, the Secretary has the authority to control demand for care through policy options, such as curtailment of enrollment for particular priority groups. Then again, it is veterans in the lowest priority groups (P7, P8) who would be impacted first.

In fact, on January 17, 2003, VA announced in the Federal Register that VA will enroll all priority groups of veterans, except those veterans in Priority 8 who were not in an enrolled status on January 17, 2003, or who requested disenrollment on or after that date. Priority Group 8 veterans already enrolled will be "grandfathered" and allowed to continue in VA's health care system.

VA has been unable to provide all enrolled veterans with timely access to health care services because of the tremendous growth in the number of veterans seeking VA health care, especially higher income priority 7 and 8 veterans seeking to use VA's relatively generous prescription drug benefit. More than half of all new enrollees have been in Priority Group 8. This demand for VA health care is expected to continue in the future.

If necessary, the Secretary could suspend enrollment of veterans in higher priorities than Priority 8, even to as high as Priority 5 veterans, but this is unlikely, as Priority 5 veterans, who are largely low income, and/or uninsured, and/or in poor health, are part of VA's mission. So other alternatives would be sought, including supplemental appropriations, or other policies concerning co-payments, benefits, etc.

VA + Choice Medicare

Work is underway with the Department of Health and Human Services to explore the possibility of offering Medicare-eligible Priority Group 8 veterans who are no longer eligible to enroll for VA health care the option of receiving their Medicare benefit through VA. The plan calls for VA to participate as a Medicare+Choice provider. VA would receive payments from a private health plan contracting with Medicare that would cover costs. The "VA+Choice Medicare" plan would become effective later this year as details are finalized between VA and the Department of Health and Human Services.

Fiscal Year 2004 Budget

In a Fiscal Year 2004 budget hearing report, it was stated that VA expects to spend about $4.4 billion this year on its pharmaceutical programs. VA's budget for prescription drugs has nearly doubled over the past three years and, at the current rate of growth, will exceed $7 billion by the end of fiscal year 2008. This budget growth is due to three factors: (i) increasing numbers of patients, (ii) intensity of drugs (newer drugs are often branded and more expensive but lead to better outcomes and fewer side effects); and (iii) medical inflation. A variety of initiatives are being considered to help stem the growth in VA's prescription-related expenditures, including a DoD/VA Pharmacy contracting pilot (Joint VA/DoD contracting for pharmaceuticals), and a joint DoD/VA Pharmacy Delivery Service pilot.

CONCURRENT SESSIONS II

Health Care Forecasting

Chair: Kathleen Sorensen, U.S. Department of Veterans Affairs

Research on Factors Important for Projecting Supply, Demand, and Shortages of Physicians

Marilyn Biviano, Tim Dall, Atul Grover, and Steve Tise
Bureau of Health Professions, U.S. Department of Health and Human Services

BHPr's Physician Supply and Physician Demand Models are used to project both the supply of and demand for physicians, by medical specialty, at the national level through 2020. Supply is expected to be driven by a stable number of graduates from U.S. medical schools and, lacking major changes in regulations affecting immigration, a stable number of international medical school graduates. Demand will be driven largely by the aging population, but tempered by the trends in reimbursement. Factors affecting demand that are harder to measure include economic growth and scientific and technological advances.

Projected Supply, Demand, and Shortages of Registered Nurses

Marilyn Biviano, Tim Dall, Atul Grover, Steve Tise, Marshall Fritz, and William Spencer
Bureau of Health Professions, U.S Department of Health and Human Services

In order to identify the extent and distribution of the Registered Nurse (RN) shortage, BHPr's Nursing Supply and Nursing Demand Models were used to project both the supply of and demand for RNs, by State through 2020. In 2000, there was a 6 percent shortage of RNs at the national level. Between 2000 and 2020, demand for full-time equivalent RNs is projected to increase by 37 percent in hospitals, 32 percent in doctors' offices, 66 percent in nursing facilities, 109 percent in home health care, and 18 percent in all other settings. Left unaddressed, the shortage is expected to grow to 29 percent by the year 2020.

Integrating Demand Modeling and Policy Making

Barbara J. Manning, Office of the Assistant Deputy Under Secretary for Health, Department of Veterans Affairs

In January 2003, the Department of Veterans Affairs (VA) reached a decision with the Department of Health and Human Services (HHS) to establish a Medicare+Choice plan where VA will provide a defined cohort of veterans aged 65 or older the option of receiving their Medicare benefit through VA, known as VA+Choice. In order to successfully plan to deliver Medicare benefits to the eligible Medicare population, the Veterans Health Administration is working with the Veterans Health Care Services Demand Model to not only determine potential demand in different geographic areas but also create the appropriate VA+Choice model. This paper will highlight some of the health care policy issues related to the benefit structures, cost projections, and market share to create this new Medicare managed care model.

The Use of Actuarial Data to Develop Policy and Budget in the Veterans Health Administration

Duane Flemming, Veterans Health Administration, U.S. Department of Veterans Affairs

The Veterans Health Care Eligibility and Reform Act of 1996 requires the Secretary to review the upcoming budget and determine if VA will be able to enroll and care for all veterans. Enrollment has increased from 4.2 million in FY1999 to 6.9 million today. Rapid enrollment growth and an aging veteran population present many challenges. In order to prepare for future demands, VA uses its Veterans Health Care Services Demand Model to project the number of veteran enrollees and their expected health care demand. This paper will highlight some of the trends identified in the projections and the impacts of several policies on future demand for health care.

RESEARCH ON FACTORS IMPORTANT FOR PROJECTING SUPPLY, DEMAND, AND SHORTAGES OF PHYSICIANS

Marilyn Biviano, Tim Dall, Atul Grover, and Steve Tise
Bureau of Health Professions
U.S. Department of Health and Human Services

ABSTRACT

Bureau of Health Professions' Supply and Physician Demand Models are used to project both the supply of and demand for physicians by medical specialty, at the national level through 2020. Supply is expected to be driven by a stable number of graduates from U.S. medical schools and, lacking major changes in regulations affecting immigration, a stable number of international medical school graduates. Demand will be driven largely by the aging population, but tempered by the trends in reimbursement. Factors affecting demand that are harder to measure include economic growth and scientific and technological advances.

Slide 1

Physician Supply and Demand Models

National Center for Health Workforce Analysis:
Atul Grover

Federal Forecasters Conference
October 27, 2003

Health Resources and Services Administration
HRSA
Bureau of Health Professions
National Center for Health Workforce Information and Analysis

Slide 2

Demand for Health Professionals Will Grow at Twice the Rate of All Occupations Between 2000-2010

	2000 (000's)	2010 (000's)	Percent Change
Total U.S Employment	145,594	167,754	15%
Total Health Occupations	10,984	14,186	29%
Physicians	598	705	18%
Dentists	152	161	6%
Pharmacists	217	270	24%
Registered Nurses	2,194	2,755	26%
Mental and Behavioral Health Occupations	518	657	27%
Therapists	479	639	33%
Public and Environmental Health	241	302	25%
Health Technicians and Technologists	2,459	3,090	26%
Health Service Occupations	3,197	4,264	33%

Source: Dept. of Labor, Bureau of Labor Statistics, Occupational Employment Statistics

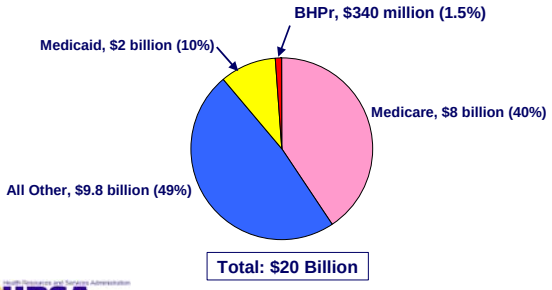
Slide 3

Physician Supply and Demand: Why Numbers Matter

- In a typical year, physician services account for one-quarter to a third of health care costs
- Physicians control 80% of health care dollars (Stoline and Weiner, 1993)
- Many projections of impending physician workforce surplus made in the 1990's and some individuals now claiming impending shortage

Slide 4

Support for GME Financing



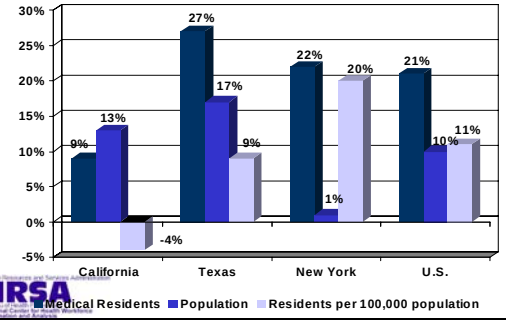
National Conference of State Legislatures

Slide 5

Regional Center for Health Workforce Studies--SUNY Albany

Change in Number of Medical Residents, Population and Residents per 100,000 Population (1988 to 1997)

"The Physician Workforce in TX, CA, and NY" Source: AMA; U.S. Bureau of the Census



Slide 6

Major Factors Impacting on Future Demand for Physicians

- Aging of population and overall growth
- Growing wealth of the nation
- Public expectations
- New medical interventions
- Evolution of managed care
- Cost containment efforts
- Growth in the number of non-physician clinicians

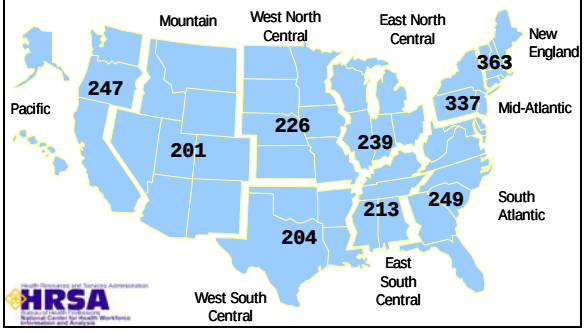
Slide 7

Factors Impacting on the Future Effective Supply of Physicians and Physician Services

- Medical school production
- Policies toward IMGs
- Aging of physician workforce/retirement patterns
- Increasing women in medicine
- Changing practice patterns
- Productivity changes

Slide 8

Active Physicians per 100,000



Slide 9

Physician Distribution–A Type of Shortage!

- There are 3,168 HPSAs*.
- Most are predominantly rural counties.
- 56 million people live in HPSAs; 33 million are underserved.
- 15,000 primary care physicians are needed to alleviate the maldistribution.

Non-Financial Barriers to Access: Inadequate Capacity

Shortage Designation
Whole County

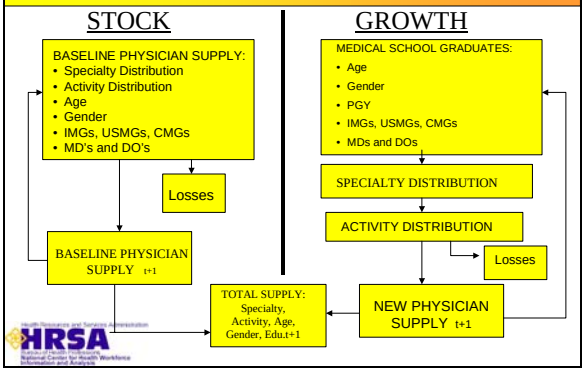
* Health Professions Shortage Areas as of March 31st 2002.

Slide 10

BHPr Supply Model

Slide 11

BHPr Physician Supply Model



Slide 12

Advantages of NCHWA Physician Supply Model

- Projects physician supply by:
  - Specialty and activity.
  - Country of medical education (USMG, IMG, CMG).
  - MDs and DOs
- Can estimate impacts on supply resulting from:
  - Aging workforce, retirement age, productivity, and women participation in the workforce.

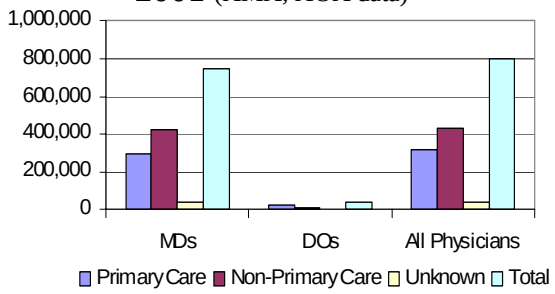
Slide 13

Limitations and Assumptions in the PSM

- Is based on historic trends in specialty choice
- Assumes constant separation rates for each age, gender, and country of medical education category
- Does not include specialty specific separation rates
- Assumes a constant distribution of physicians by activity
- Estimates of hours worked by gender and age is at 1994 level when converting to FTE

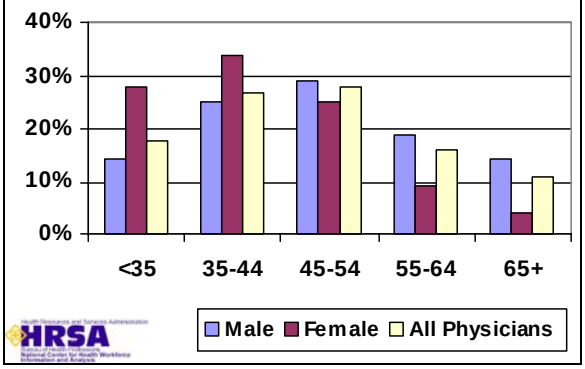
Slide 14

Primary Care vs. Specialty Supply 2001 (AMA, AOA data)



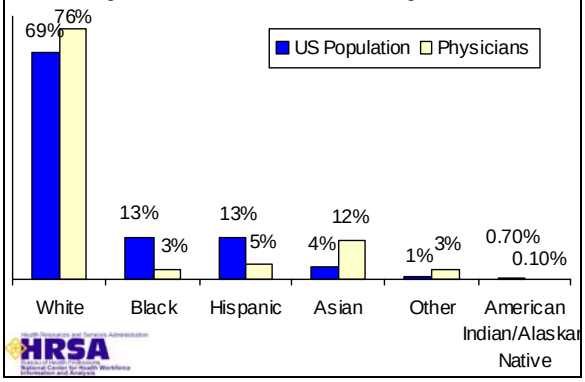
Slide 15

Physician Age Distribution, 2001



Slide 16

Physician Race/Ethnicity, 2001



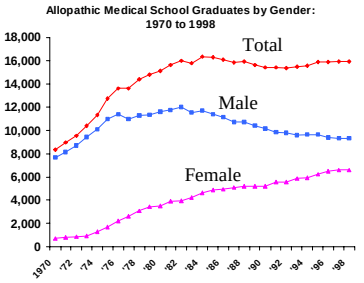
Slide 17

Research Questions the Supply Model can Help Answer

Slide 18

How does the increased participation of women in the physician workforce affect effective supply?

- Women accounted for 8% of graduates in 1970 and 41% in 1998.
- Women now account for 23% of the physician workforce.
- Women physicians, however, work 5 hours less per week and see about 20 fewer patients than male physicians.1/

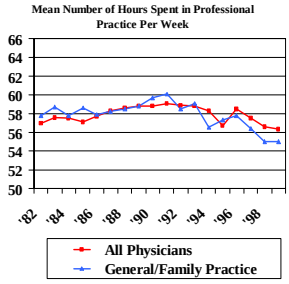


1/ COGME Fifth Report: Women and Medicine. 1995

Slide 19

What would be the impact of a continued decline in the number of hours physicians worked?

- Between 1982 and 1998, the average number of hours "All Physicians" spent in professional practice declined by 2%.
- For GP/FP physicians the decline was 5%.



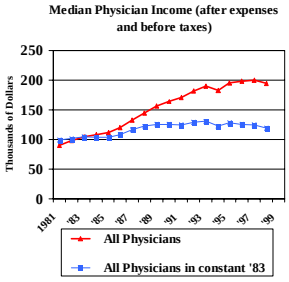
HRSA

Physician Socioeconomic Statistics and Characteristics of Medical Practice, various years.

Slide 20

What do changes in salaries say about the adequacy of physician supply?

- Changes in salaries are an indication of shortages and supply excess.
- While actual physician salaries have grown since 1981, when adjusted for inflation they have remained relatively flat since the mid-1990's.

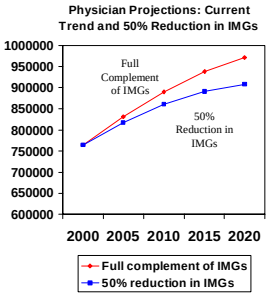


HRSA

Slide 21

What would be the impact of a decline in the number of IMGs on Supply?

- Reducing the number of IMGs entering the U.S. physician workforce by half would result in 62,000 fewer physicians in 2020 than is currently projected.
- This represents 7% fewer physicians than is currently projected for 2020.

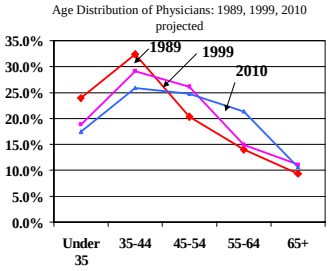


HRSA

Slide 22

What will be the effect of an aging physician workforce on supply?

- An increasing proportion of physicians are approaching retirement age.
- In 1989 23% of physicians were 55 + compared to 26% in 1999.
- By 2010, 32% of physicians will be 55 +.

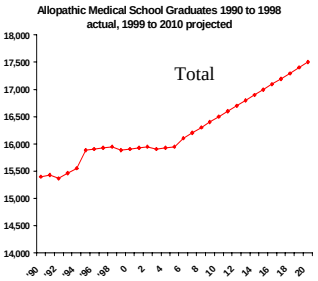


HRSA

Slide 23

What would be the effect of expanded medical school enrollments on supply?

- If medical schools enrollments increased so that graduates grew by 100 per year through 2020, then there would be 1,500 more graduates in 2020 than in 1996.
- Holding the number of IMGs constant, this would result in a 4% increase in first year residents in 2020 over the 2000 level.
- The cumulative effect would be an additional 1,500 physicians in the workforce



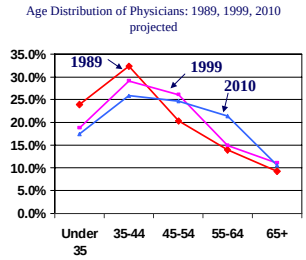
1/ COGME, Fifth Report: Women and Medicine. 1995

HRSA

Slide 24

Assessing the Trend in the Retirement Plans of Physicians

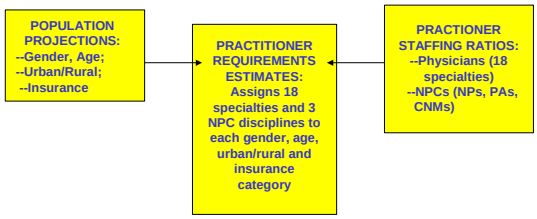
- An increasing proportion of physicians are approaching retirement age, yet little is know about when they actually retire.
- By 2010, 32% of physicians will be 55 +.
- Information on retirement plans by such factors as age, gender, and specialty will be used to help develop the next generation of physician supply projections.



HRSA

Slide 25

Physician Demand Model (PDM): Methodology



Slide 26

PDM Methodology

- Population Projections
  - Population by age, gender, urban/rural location
  - Insurance distributions by age, urban/rural location
- Practitioner Staffing Ratios
  - Staffing ratios for each population group--age, gender, insurance setting, urban/rural location
- Practitioner Requirements
  - For all population groups: type of insurance coverage and urban/rural location

Slide 27

Population Data Characteristics in PDM

Description	Categories	
Sex	Male, Female	
Age	1-4, 5-17, 18-24, 25-44, 45-64, 65-74, 75-84, 85+	
Location	Urban, Rural	
Insurance Type	Staff HMO IPA HMO FFS Medicaid Staff HMO	Medicaid IPA HMO Medicaid FFS No Insurance Medicare FFS, Staff HMO, IPA

Slide 28

Specialties in the PDM

General/Family Practice
General Internal Medicine
Pediatrics
Internal Medicine Subspecialties
Cardiovascular Diseases
General Surgery
Obstetrics/Gynecology
Otolaryngology
Orthopedic Surgery
Urology
Ophthalmology
Other Surgical Specialties
Psychiatry
Anesthesiology
Emergency Medicine
Radiology
Pathology
Other Specialties

Slide 29

Non-Physician Primary Care Clinicians in the PDM

Nurse Practitioner
Physician Assistant
Certified Nurse Midwives

Slide 30

Advantages of NCHWA's Physician Demand Model

- Provides estimates of the impact on physician demand resulting from:
  - Non-physician provider substitution.
  - Changing population demographics (age and gender)
  - Alternative insurance options.
  - Population migration (urban and rural)
- Provides estimates of demand for three primary care specialties and 15 non-primary specialties.
- Results are reproducible
- Model is easily shared and adapted to researchers studying local jurisdictions.

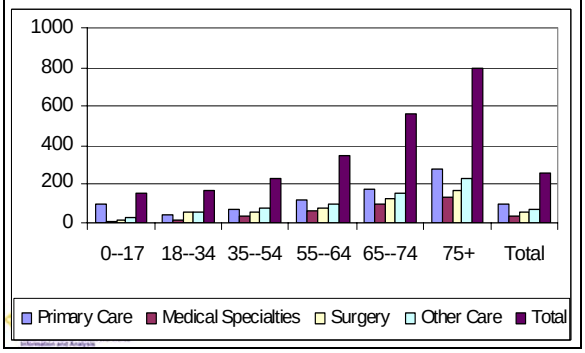
Slide 31

Limitations and Assumptions in NCHWA Physician Demand Model

- Staffing ratios are static.
- Physician demand are not affected by economic variables.
- Technology is static.

Slide 32

PDM: Physician Demand by Age, 2000



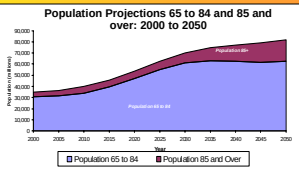
Slide 33

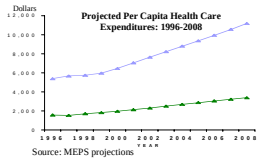
Research Questions that the PDM can Help Answer

Slide 34

What are the impacts on primary care and specialist physician requirements resulting from the aging US population?

- Over the next 25 years, the population over 65 will grow at a rate 5 times that of those under 65.
- Health care expenditures more than double by 2008 to **$ 2,000,000,000,000** (trillion dollars).
- The elderly will account for 13% of the population but a **third of the dollars.**



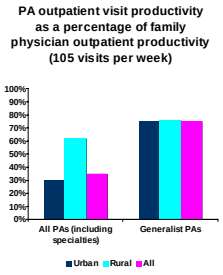


Source: MEPS projections

Slide 35

What is the impact of non-physician clinicians (NPCs) substitution on primary care demand?

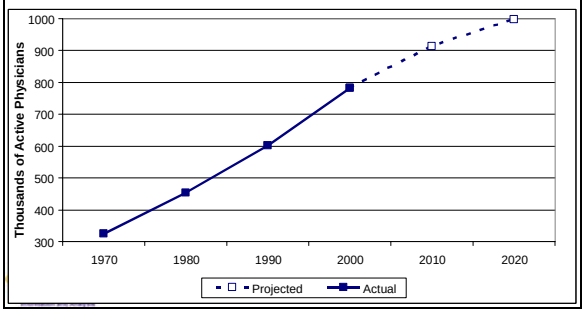
- PA productivity (measured in outpatient visits) ranges from 0.35 to 0.75 of an FTE family physician, depending on specialty and location.
- Substitution rates of PAs can be introduced in the model by specialty and rural vs. urban location.



Larson, Eric H., L. Gary Hart and Ruth Ballweg. National Estimates of Physician Assistant Productivity. 2000(January). WWAMI Center for Health Workforce Studies/Rural Health Research Center Working Paper #57.

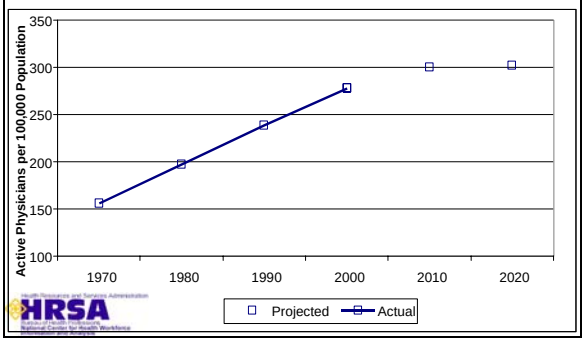
Slide 36

PSM Supply Projections: Total Physicians



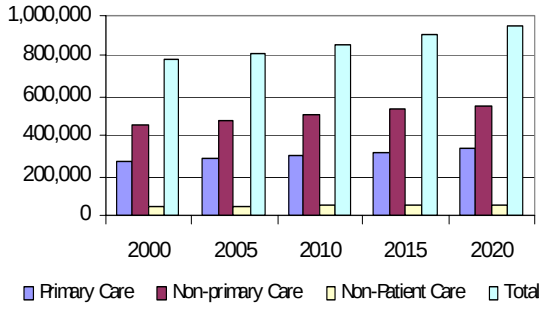
Slide 37

PSM Per Capita Projections: Physician Supply, 1970-2020



Slide 38

PDM: Projected Physician Demand 2000-2020



Slide 39

Rising Physician Workforce Issues

- Limits on resident work hours & subsidies
- Ability to fill call rosters for emergency services
- Regional specialty shortages
- Changing activities of physicians
- How much growth in health expenditures can be sustained?

Slide 40

Atul Grover

Chief Medical Officer

Email: agrover@hrsa.gov
Ph: (301) 443-1070
Fax: (301) 443-8003

http://bhpr.hrsa.gov/healthworkforce

PROJECTED SUPPLY, DEMAND, AND SHORTAGES OF REGISTERED NURSES

Marilyn Biviano, Tim Dall, Atul Grover, Steve Tise,
Marshall Fritz, and William Spencer
Bureau of Health Professions
U.S. Department of Health and Human Services

Abstract

An adequate supply of nurses is essential to achieving the nation's goals of ensuring access to affordable, high-quality health care. Models developed by the National Center for Health Workforce Analysis to project the future supply of and demand for registered nurses suggests that if current trends continue the current shortage of approximately 100,000 nurses will grow ten-fold by 2020. In this paper, we describe the data, methods and assumptions used to develop the Center's Nursing Supply Model (NSM) and Nursing Demand Model (NDM) and we present findings from the models.

Slide 1

Projected Supply, Demand, and Shortages of Registered Nurses

Federal Forecasters Conference
October 27, 2003

National Center for Health Workforce Analysis:
Marilyn Biviano
http://bhpr.hrsa.gov/healthworkforce/

Slide 2

Overview of Presentation

- National Center for Health Workforce Analysis—who we are
- Nursing Supply Model
- Nursing Demand Model
- Projected Nursing Supply, Demand, and Shortages

Slide 3

National Center for Health Workforce Analysis

WHO WE ARE
Shortage Designation Branch
Workforce Analysis Branch

Slide 4

Health Workforce Analysis Mission and Strategic Plan Highlights

Mission
Collect, analyze, and disseminate health workforce information and facilitate national, State, and local workforce planning efforts.

Strategic Plan Functions

- Collect heath professions related data.
- Assist State and local workforce planning efforts.
- Develop tools and conduct research on health workforce.
- Conduct issues related analyses.

Slide 5

National Center for Health Workforce Analysis

http://bhpr.hrsa.gov/healthworkforce

- **National Center for Health Workforce Analysis**: Only Federal (or non-Federal) effort that focuses on health workforce supply, demand, and issues. (http://bhpr.hrsa.gov/healthworkforce/).
  - **Regional Centers for Health Workforce Studies**
    - University of California/SanFrancisco (UCSF) (http://futurehealth.ucsf.edu/cchws.html)
    - State University of NY/Albany (SUNY/Albany) (http://chws.albany.edu)
    - University of Illinois/Chicago (UIC) (http://www.uic.edu/sph/ichws)
    - University of Washington/Seattle (UW) (http://www.fammed.washington.edu/CHWS/index.html)
    - University of Texas Health Sciences Center at San Antonio (Not Yet Available)

Slide 6

Nursing Supply Model (NSM)

Slide 7

NSM: Overview

- Definitions
- Structure and data
- Model limitations and assumptions

Slide 8

NSM: Definitions

- The NSM projects three measures of RNs available to provide services
  - RN population: the number of licensed RNs
  - RN supply: the number of licensed RNs in the nurse workforce (i.e., employed in nursing or seeking employment in nursing)
  - RN FTE supply: the number of FTE RNs employed in nursing
- Full Time Equivalent (FTE):
  - Employed full time during entire year = 1 FTE
  - Employed part time or for part of year = ½ FTE

Slide 9

NSM: Structure and Data

- Model dimensions
  - 50 States plus the District of Columbia
  - Three education levels
    - Diploma and associates degree
    - Baccalaureate degree
    - Masters or higher degree
  - RN age
  - Years 2000 to 2020

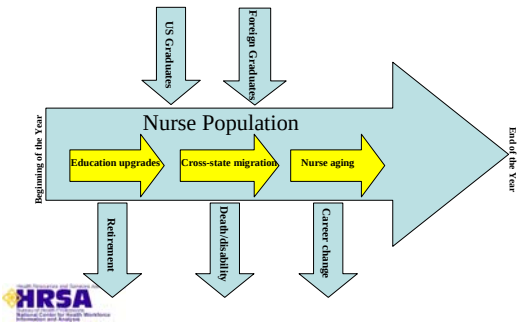
Slide 10

NSM: Structure and Data

- The NSM projects the RN population based on projections of
  - The number of graduates from U.S. nursing schools
  - Immigration of RNs from outside the U.S.
  - Change in educational attainment
  - Cross-state migration
  - Attrition from the current RN population (e.g., due to career changes, retirement, death and disability)

Slide 11

NSM: Structure and Data



Slide 12

NSM: Structure and Data

- The NSM projects the RN supply based on projections of
  - The RN population
  - RN activity rates (i.e., the ratio of RNs in the workforce to total licensed RNs), by RN age and education level
- The NSM projects the RN FTE supply based on projections of
  - The RN population
  - RN FTE activity rates (i.e., the ratio of FTE RNs in the workforce to total licensed RNs), by RN age and education level

Slide 13

NSM: Structure and Data

- To project new graduates from nursing programs
  - Used State-level data from multiple years
  - Estimated multiple regression model to estimate the relationship between number of RN graduates and
    - Characteristics of the current RN population
    - Teachers' average annual salary
    - Nurses' average annual salary
    - Number of females age 20 to 34
    - Per capita income
    - HMO saturation rates
    - Size of the elderly population

Slide 14

NSM: Structure and Data

- Primary data sources
  - 2000 Sample Survey of RNs
    - To estimate base year RN population by age, education level and State
    - To estimate labor force activity rates by RN age and education level
    - To estimate cross-state migration patterns
  - Current Population Survey
    - To estimate RN retirement and disability rates
  - Other
    - To estimate the relationship between graduates from nursing programs and economic/demographic factors

Slide 15

NSM: Limitations and Assumptions

- National projections are more reliable than State projections
- Parts of model are static
  - Retirement, education upgrades, and labor force participation patterns constant across States and over time, but differ by RN age and education level
  - Cross-state migration patterns constant over time, but vary by State and RN age and education level

Slide 16

NSM: Limitations and Assumptions

- NSM not easily adaptable to make intra-state supply projections
- RN supply projections are independent of
  - RN demand projections
  - Projected supply of other health workers (e.g., LPNs)
- NSM does not project supply by RN specialty (e.g., APNs)

Slide 17

Nursing Demand Model (NDM)

Slide 18

NDM: Overview

- Definitions
- Structure and data
- Model limitations and assumptions

Slide 19

NDM: Definitions

- Nurse Demand: is defined as the number of full time equivalent (FTE) nurses that employers are willing to hire given population needs, economic considerations, the healthcare operating environment, and other factors.
- Nurse Aides: refers to all paraprofessional nursing staff working in hospitals, nursing homes, or home health care.

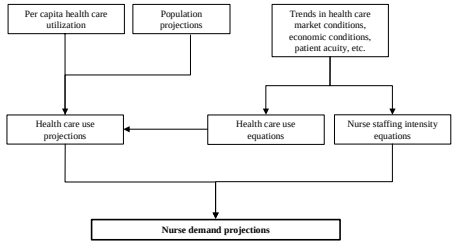
Slide 20

NDM: Structure and Data

- Model dimensions
  - Demand for RNs in 12 settings, and demand for LPNs and NAs in 5 settings
  - 50 States plus the District of Columbia
  - Years 2000 to 2020

Slide 21

NDM: Structure and Data



Slide 22

NDM: Structure and Data

- Health care use and nurse staffing equations estimated using multiple regression analysis
- Determinants include
  - Demographics and geographic location of population
  - Economic factors
    - Nurse wages (e.g., ratio of RN to LPN hourly wages)
    - Medicare and Medicaid reimbursement rates
    - Per capita personal income
  - Characteristics of the health care system
    - Shift from hospital inpatient to outpatient services
    - Percent uninsured
    - HMO saturation rate

Slide 23

NDM: Limitations and Assumptions

- NDM is a simplified model of the complex health care system
  - Contains limited number of explanatory variables
  - NDM has limited ability to model effect on nurse demand of substitution between nurse types
  - Technology is static

Slide 24

NDM: Limitations and Assumptions

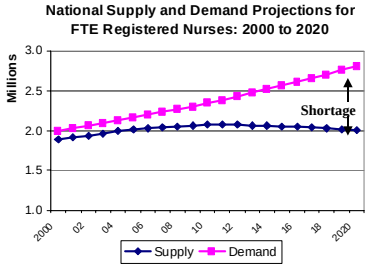
- Data limitations
  - NDM uses state-level data to estimate relationships that occur at the local level: e.g.,
    - The relationship between demand for health care services and its determinants
    - The relationship between nurse staffing rates and their determinants
  - Small sample size in some surveys (e.g., RN Survey) reduces reliability of projections
    - For smaller States
    - For settings that employ fewer nurses

Slide 25

Projected Nursing Supply, Demand, and Shortages

Slide 26

Projections of Supply and Requirements for Full-time Equivalent RNs: 2000-2020



Source: HRSA, BHPr, *Projected Supply, Demand and Shortages of Registered Nurses: 2000-2020*

Slide 27

Between 2000 and 2020, State wide shortages of RNs will increase.

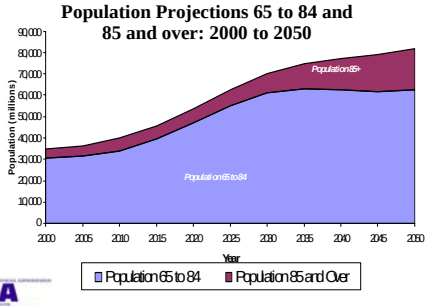
In 2000, 30 States were estimated to have shortages.

By 2020, the number of States estimated to have shortages will grow to 46.

Source: HRSA, BHPr, *Projected Supply, Demand and Shortages of Registered Nurses: 2000-2020*

Slide 28

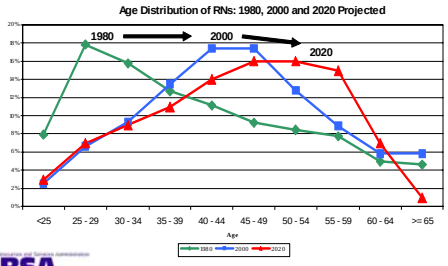
The aging of the population will have a major impact on the demand for RNs.



Source: U.S. Dept. of Commerce, Bureau of the Census

Slide 29

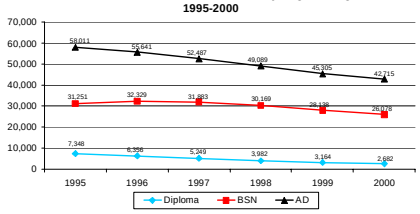
The RN Workforce has a Declining Proportion of Young Nurses and an Increasing Proportion of Nurses Nearing Retirement Age



Source: HRSA, BHPr, *The National Sample Survey of Registered Nurses 1980-2000* and *Projected Supply, Demand and Shortages of Registered Nurses: 2000-2020*

Slide 30

RN Supply Trends: U.S. Graduates

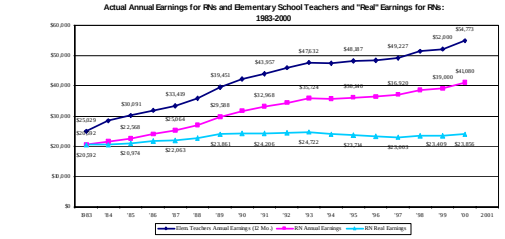


- Since 1995, there has been a significant reduction in the number of U.S. RN graduates: Associate Degree (30%), Diploma (66%), BSN (23%), and Total (31%).

Source: Unpublished Data from the National League of Nursing

Slide 31

Alternative careers offering higher salaries, reduce the number of students choosing nursing.

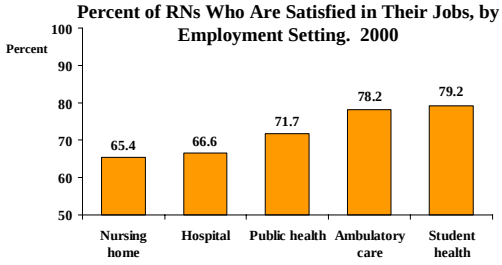


- Salaries for elementary school teachers are an attractive alternative to being a registered nurse.

Source: Bureau of Labor Statistics, Current Population Survey

Slide 32

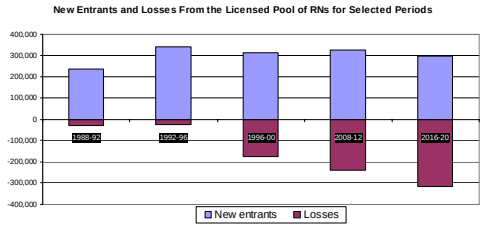
RNs in nursing homes and hospitals report the least job satisfaction



Source: HRSA, BHPR, *National Sample Survey of Registered Nurses: 2000*

Slide 33

If current trends continue, the number of RNs giving up their license will outnumber the number of new entrants.



Source: HRSA, BHPr, *The National Sample Survey of Registered Nurses 1980-2000* and *Projected Supply, Demand and Shortages of Registered Nurses: 2000-2020*

Slide 34

Questions?

Contact:
Marilyn Biviano
mbiviano@hrsa.gov

INTEGRATING DEMAND MODELING AND POLICY MAKING

Barbara J. Manning, Office of the Assistant Deputy Under Secretary for Health
Department of Veterans Affairs

ABSTRACT

In January 2003, the Department of Veterans Affairs (VA) reached a decision with the Department of Health and Human Services (HHS) to establish a Medicare+Choice plan where VA will provide a defined cohort of veterans aged 65 or older the option of receiving their Medicare benefit through VA, known as VA+Choice. In order to successfully plan to deliver Medicare benefits to the eligible Medicare population, the Veterans Health Administration is working with the Veterans Health Care Services Demand Model to not only determine potential demand in different geographic areas but also create the appropriate VA+Choice model. This paper will highlight some of the health care policy issues related to the benefit structures, cost projections, and market share to create this new Medicare managed care model.

Slide 1

Integrating Demand Modeling and Policy Making

Barbara J. Manning
Enrollment and Forecasting
Veterans Health Administration

Slide 2

VHA Health Care Demand Model Now Supports

- Budget formulation
  - Establishes health care resource requirements
  - Identifies funding gaps
  - Used to formulate budget policy options
- Policy development
  - VA+Choice
  - Regulations and legislation
- VHA's capital planning process
  - Provides enrollment and workload projections out 20 years

Slide 3

Integration Driven by…

- Projected health care costs that exceed available resources
- Need for fast, accurate costing of policy options to address funding gaps
- Recognition of the value of cost estimates developed by an independent entity

Slide 4

Proposed Policies Modeled

- Suspend enrollment
- Disenroll segments of the enrollee population
- Increase enrollee cost sharing
  - Deductible
  - Enrollment fee
  - Increase co-payments
- Explore alternative means of providing access to health care – VA+Choice

Slide 5

Communication and Education Key to Integration

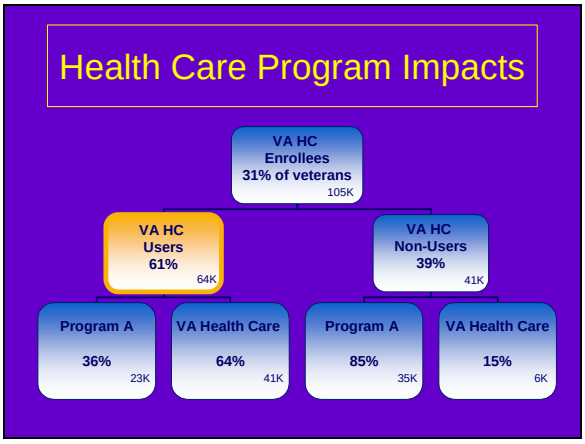
- Educate decision makers and staff about methodology and results of the modeling
- Translate volumes of data into useable information that supports decision making

Slide 6

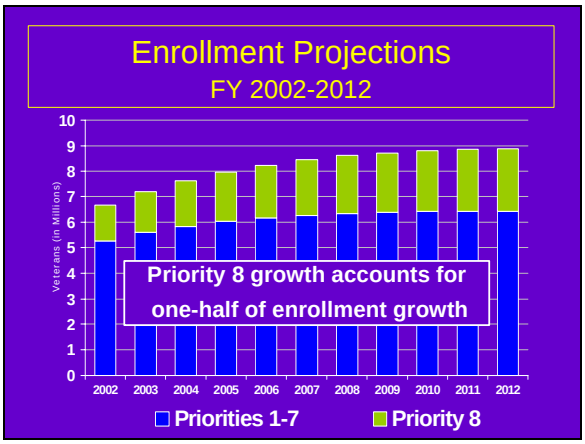
Lessons Learned

- Quickly provide big picture
- Provide only facts relevant to policy decisions
- Rewrite technical descriptions of methodology into short, clear explanations
- Use graphics to illustrate key points

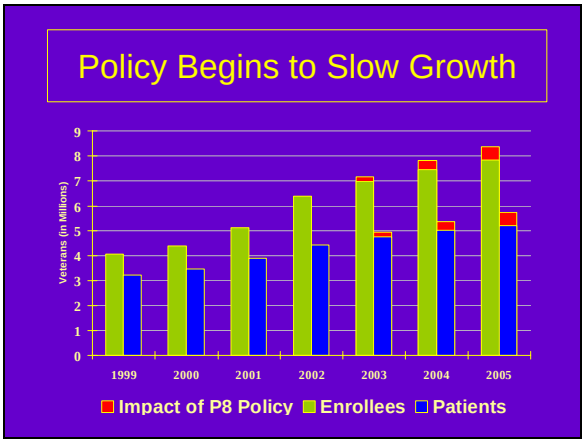
Slide 7



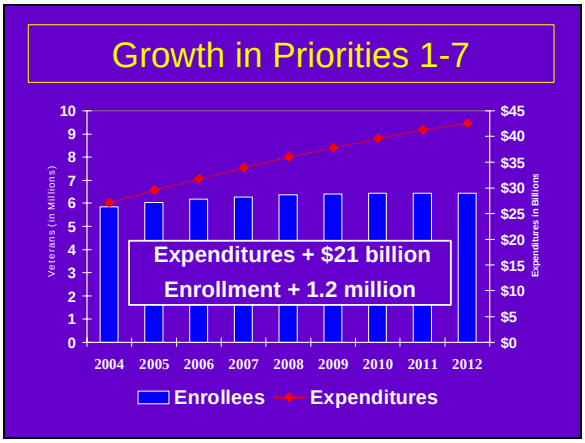
Slide 8



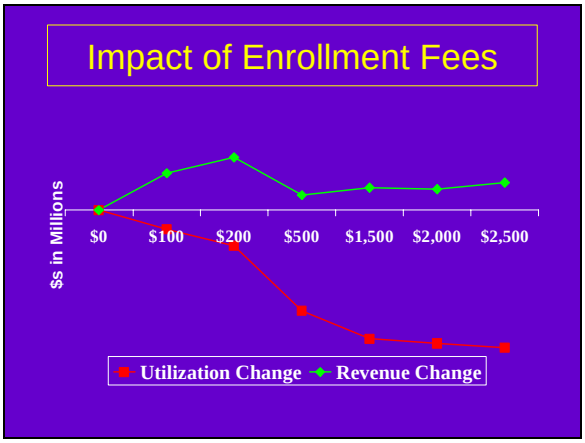
Slide 9



Slide 10



Slide 11



Slide 12

Modeling for VA+Choice

- Enrollment, workload, and expenditures at facility, market, and national level
  - Run with VA costs or Medicare Allowable Charges
- Average payment rate from Medicare
  - Based on the projected veteran enrollees
- PMPM
- Profit/Loss Ratio
  - Reflected all costs including program administration, M+C provider profit, and risk contingency

Slide 13

Modeling VA+Choice

- Assess the effect of changes in benefit and cost sharing on enrollment and expenditures
- Explore VA's ability to offer a national versus regional benefit structure
- Highlighted need to enhance the model to reflect variable and marginal costs

Slide 14

Barbara J. Manning
202.273.6097
barbara.manning@hq.med.va.gov

THE USE OF ACTUARIAL DATA TO DEVELOP POLICY AND BUDGET IN THE VETERANS HEALTH ADMINISTRATION

Duane Flemming, Veterans Health Administration
Office of the Assistant Deputy Under Secretary for Health

The Veterans Health Care Eligibility and Reform Act of 1996 (Public Law 14-262) requires the Secretary of the Department of Veterans Affairs to review the budget allocated to VA for the upcoming fiscal year and determine whether VA will be able to continue to enroll and care for all veterans. Enrollment in VA has increased from 4.2 million in Fiscal Year 1999 to over 7.1 million today. Rapid enrollment growth and an aging veteran population present many challenges to VA in the 21st century. In order to prepare and position itself to meet the future demands of veterans, VA uses its Veterans Health Care Services Demand Model to project the number of future veteran enrollees and their expected demand for health care. This paper will highlight some of the trends identified in veteran enrollment projections and the impacts of several policies on future demand for health care.

Background

The Veterans Administration was established in 1930 when Congress authorized the President to "consolidate and coordinate Government activities affecting war veterans". It has responsibility for providing federal benefits to veterans and their dependents. On March 15, 1989, the Department of Veterans Affairs (VA) was established as a Cabinet-level position. The Veterans Health Administration (VHA) is one of the three administrations; the others are the Veterans Benefits Administration (VBA) and the National Cemetery Administration (NCA). The four missions of VHA are: health care, health professional training, research and emergency preparedness.

The Veterans Health Administration (VHA) is the largest integrated health care system in the United States. The VA health care system has grown from 54 hospitals in 1930 to include 162 hospitals, 133 nursing homes, 43 domiciliaries and more than 650 community based outpatient clinics (CBOCs). VA has been providing health care to veterans for over sixty years. In 2003, VA provided medical care to more than 4.4 million veterans. VHA is also the largest single provider of health professional training in the world, has one of the largest and most productive research organizations in the country, and provides backup to both the Department of Defense and National Disaster Medical System.

VA's budget for health care is a discretionary program and subject to the annual appropriation process by Congress; unlike Medicare, it is not an entitlement that must be funded every fiscal year. Years ago, the VA health care system was primarily an inpatient-based system with little outpatient care. Complex eligibility rules determined where and how which veterans could be treated for which conditions. These rules were difficult for clinicians and administrators to understand and often were not applied uniformly throughout the system.

VA has traditionally been a primary provider of health care for veterans with service-connected disabilities, VA pensioners, veteran populations with special rehabilitation needs and low-income veterans lacking other health care coverage. In October 1996, Congress enacted the Veterans' Health Care Eligibility Reform Act of 1996, Public Law 104-262, permitting VA, for the first time ever, to offer a comprehensive medical benefits package to all veterans who enrolled in the VA health care system. VA began promoting preventive and primary care, striving to keep veterans healthy and avoiding preventable resource intensive hospital admissions. While this Act (P.L. 104-262) simplified the system, it required VA to implement a priority based enrollment system. Since the beginning new priority and subpriorities have been created in response to changing requirements and providing the VA with flexibility to address future budgetary issues. Table 1 shows the FY 2003 Priority Levels and their description.

Table 1. VA Enrollment Priorities, FY 2003

Priority	Description
1	Veterans with service-connected disabilities rated 50% or more disabling
2	Veterans with service-connected disabilities rated 30% or 40% disabling
3	Veterans who are former POWs Veterans awarded the Purple Heart Veterans whose discharge was for a disability that was incurred or aggravated in the line of duty Veterans with service-connected disabilities rated 10% or 20% disabling Veterans awarded special eligibility classification under Title 38, U.S.C., Section 1151, "benefits for individuals disabled by treatment or vocational rehabilitation"
4	Veterans who are receiving aid and attendance or housebound benefits Veterans who have been determined by VA to be catastrophically disabled
5	Nonservice-connected veterans and noncompensable service-connected veterans rated 0% disabled whose annual income and net worth are below the established VA Means Test thresholds Veterans receiving VA pension benefits Veterans eligible for Medicaid benefits
6	Compensable 0% service-connected veterans World War I veterans Mexican Border War veterans Veterans solely seeking care for disorders associated with: • Exposure to herbicides while serving in Vietnam; or • Exposure to ionizing radiation during atmospheric testing or during the occupation of Hiroshima and Nagasaki; or • For disorders associated with service in the Gulf War; or • For any illness associated with service in combat in a war after the Gulf War or during a period of hostility after November 11, 1998.
7	Veterans who agree to pay specified copayments with income and/or net worth above the VA Means Test threshold and income below the HUD geographic index • Subpriority a: Noncompensable 0% service-connected veterans who were enrolled in the VA Health Care system on a specified date and who have remained enrolled since that date • Subpriority c: Nonservice-connected veterans who were enrolled in the VA Health Care System on a specified date and who have remained enrolled since that date • Subpriority e: Noncompensable 0% service-connected veterans not included in Subpriority a above • Subpriority g: Nonservice-connected veterans not included in Subpriority c above
8	Veterans who agree to pay specified copayments with income and/or net worth above the VA Means Test threshold and the HUD geographic index • Subpriority a: Noncompensable 0% service-connected veterans enrolled as of January 16, 2003 and who have remained enrolled since that date • Subpriority c: Nonservice-connected veterans enrolled as of January 16, 2003 and who have remained enrolled since that date • Subpriority e: Noncompensable 0% service-connected veterans applying for enrollment after January 16, 2003 • Subpriority g: Nonservice-connected veterans applying for enrollment after January 16, 2003

The number of priority levels VHA will be able to deliver care to is in direct relationship to medical care funding appropriated to VA and enrollees' projected demand for health care services. In response to passage of the "Veterans' Health Care Reform Act of 1996", The Office of the Assistant Deputy Under Secretary for Health contracted with Condor Technology Solutions, Inc, in partnership with Milliman, USA, a well-respected actuarial firm, to develop an actuarial health care services demand projection model. Now, in its sixth year, this model has been used to make enrollment-related projections and analyses.

The VA Enrollee Health Care Projection Model has revolutionized VHA's planning, budgeting, and policy-making processes. Before the development of this model, VHA budgets (like most other federal budgets) were based on historical expenditures that were adjusted for inflation and then increased based on proposed new initiatives. Using this model, VHA developed its FY 2003, 2004 and most recently 2005, budgets based on actuarial forecasts of projected expenditures. This transition from a historical to an actuarial-based model as the basis of budget formulation represents not only a significant innovation for VHA, but for the federal government.

The model has also become a key component of VHA's planning process and VA's policy development. During development of the FY 2003 and 2004 budgets, VHA compared actuarial expenditure projections with expected resources and identified significant gaps between veteran demand for VA health care and the resources to pay for that care. VHA then used the model to predict the impact of proposed policy options, such as requiring copayments and limiting enrollment, on expenditures and revenue. Data generated from the model were also used to estimate the impact of these policies on veteran access to care and VHA's performance indicators. The proposed health care policies in the VA's FY 2004 President's Budget were developed through this process.[1]

General Approach

The VA Enrollee Health Care Projection Model was developed through a public-private partnership and is based on private sector benchmarks that have been risk adjusted for the characteristics of the VHA enrollee population and on actual VHA unit costs. The health care utilization benchmarks developed by the actuary are based on their private sector averages and adjusted to reflect the VA enrollee population by age, gender, morbidity and reliance upon VA for health care services. It produces projections for veteran enrollment, utilization and expenditures including detailed projections for 50 health care service categories for the upcoming fiscal year as well as future years requested by Office of the Assistant Deputy Under Secretary for Health and VA. In addition to enrollment projections, VA has also requested patient (unique user) projections as a way of identifying how many veterans VA may expect to request health care services. Actual enrollment experience is tracked by VHA and reported on a monthly basis. A master enrollment file of every veteran and all of the events about the veteran's enrollment and health care utilization is created; VHA provides actual VA enrollment, utilization and unit cost data for the last complete fiscal year (i.e. FY02 data was provided for development of FY04 projections) to the contractor. Utilization is also adjusted by the degree of community management within the VA compared to community private sector's degree of management. Projected enrollee expenditures are calculated by multiplying VA unit costs by the adjusted private sector utilization norms for VA enrollees. Unique patients are also projected based upon the enrollee and utilization projections. [2]

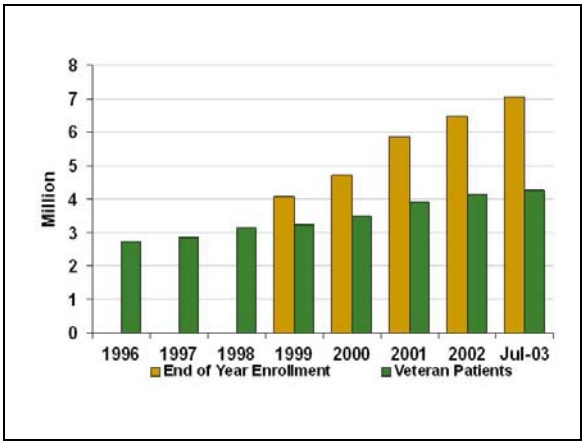
Since enrollment began in FY 1999 with passage of the "Veterans' Health Care Reform Act of 1996", VHA has seen unprecedented growth in veteran enrollment and patients. Although the veteran population is declining, VA has enrolled over 3.7 million new veterans since FY 1999. As of July 2003, over 7.1 million veterans have enrolled with VHA, an increase of 114% since 1999. The number of patients whom we provided health care grew by 54 percent between 1996 and 2002. Figure 1 shows the growth in enrollees between fiscal years 1999 and 2002 and patients between fiscal years 1996 and 2002. Several factors have contributed to this remarkable increase in demand for health care services by veterans from VHA. VA has earned a national reputation as a leader in the delivery of quality health care services through advances in quality and patient safety. Access to health care has

[1] VHA Vision 2020, Veterans Health Administration, U.S. Department of Veterans Affairs

[2] The Department of Veterans Affairs Health Care Enrollment Projections, FFC 2002, Gregg A. Pane, MD, MPA, Mary E. (Beth) Martindale, Dr PH, Randall J. Remmel, PhD, MBA, Don Stockford, MA

improved tremendously since the mid 1990's with the opening of hundreds of community based outpatient clinics. VA also has a very favorable pharmacy benefit as part of its medical benefits package compared to other health care providers, especially Medicare, thus attracting many older or sicker veterans to VHA.

Figure 1. Veterans Enrolled and Treated by VHA, Fiscal Years 1996 – 2002



This growth in enrollment has exhausted VA's marginal capacity to provide care. Since FY 2001 VHA has seen an increase in the number of patients on waiting lists for outpatient appointments and longer periods between the time a veteran makes an appointment and the time he is seen by a clinician.

Our core veteran population is defined as veterans who are in enrollment Priority Groups 1 – 6. As of June 2003, 68% of our enrollees and 73% of the veteran patients are in these priority groups. Nearly 35% of our enrollees and patients are veterans who are nonservice-connected whose annual income and net worth are below the established VA Means Test threshold (Priority Group 5).

The veteran population is projected to decrease 12.2% from 25.6 million in FY 2002 to 22.5 million in FY 2009 as shown below in Table 2. The core veteran population is projected to decline 8 percent by FY 2009, from 10.1 million to 9.3 million compared with the P7-8 veteran population which is projected to decline by 15 percent from 15.5 million to 13.2 million. The under age 65 veteran will decline 13.9 percent from 16.3 million to 14 million and the ages 65 and over will decline by only 9.3 percent to 8.4 million veterans. However, the number of veterans enrolling with VHA for health care continues to increase. More importantly, those enrollees who utilize VHA for some or all of their health care needs has also significantly increased. The VA Enrollee Health Care Projection Model shows the number of P1-6 enrollees is projected to increase 25.5 percent from 4.4 million in FY 2002 to 5.5 million in FY 2009 and the number of P1-6 veteran patients is projected to increase from 3.3 million in FY 2002 to 4.1 million by FY 2009, an increase of 24 percent. Our patients are older, sicker, have less income and less insurance than the general population. There are 25.2 million living veterans who are eligible for health care. As of July 2003, over 7.1 million have enrolled in VHA with 4.2 million veterans treated so far in FY 2003.

Table 2. Projected Changes in the Veteran Population by Priority and Age Groups FY 2002 - 2009

Priority	Under Age 65	Age 65 and Over	All Ages
P1-6	-7.9%	-8.1%	-8.0%
P7-8	-17.1%	-0.1%	-15.0%
All Priorities	-13.9%	-9.3%	-12.2%

Based on our survey of enrollees we know that our patients are older, sicker and have less income and less insurance than the general U.S. population (Table 3). The veteran population, while declining, is an aging population. The average age of VA enrollees is 63 years. Two out of every three males age 65 and over are veterans.

Table 3. Comparison of VA Patients and the General Population

	General Population	VA Patients (Users)
Ages 65 and Over	12%	48%
Income <$26,000	56%	28%
No health insurance	27%	15%

The Model in Brief

The VA Enrollee Health Care Projection Model has been of tremendous value for senior managers in terms of developing Department budgets and policy initiatives for veterans' health care. The model, first developed in FY 1999 continues to evolve. The predictive performance of the model has been refined, through on-going reviews of the model and incorporating additional health care utilization data by veterans to measure their reliance on VA, self-reported income and health care insurance information from annual surveys and a match of survey information with data from the Centers for Medicare and Medicaid Services (CMS) for enrollees ages 65 and over. An example of a recent model enhancement is the addition of pharmacy data

including medications dispensed and complete unit cost information. Even with the volume discounts VA receives from pharmaceutical suppliers, VA spent over $3 billion in FY 2002 on medications and the budget for drugs is expected to exceed $7 billion by FY 2009.

The demand model is now used for many activities within the Department including the following:

- Secretary's annual enrollment decision
- Budget formulation
- VA+Choice
- Capital Asset Realignment for Enhanced Services (CARES)
- Strategic Planning

VHA is beginning to develop a front-end user interface that will provide limited ability for modeling "what-if" scenarios and real-time interaction. This model is a powerful tool we need to be sure that it is used appropriately.

The model's assumptions and methodologies have been subject to rigorous review by the Office of Management and Budget (OMB), General Accounting Office (GAO), and the VA Office of the Actuary. As VA and VHA identify more applications for model projections, the complexity of the model has increased in order to refine its predictive capability. New data sources are recognized and incorporated into the model each year.

Modeled projections are continually reviewed with actual experience data to determine how well the model performed and identify model enhancement that may result in more accurate projections. Table 4 shows a comparison of projected versus actual for FY 2002. The model projected average enrollment for the fiscal year of 6,375,154 which exceeded actual average enrollment by 0.09% or 5,738 enrollees. The number of live enrollees as of the end of FY 2002 was under projected slightly at 6,665,271. Total unique patients were over projected by 0.19% or 8,391 individuals. Obligations were over projected by 2.6% and VHA is continuing to investigate this variance and understand the impact on model projections for FY 2004 and 2005. One possible reason is that VHA achieved greater improvements in efficiency than the model projected.

Table 4. Comparison of Projections to Actual Experience in FY 2002

Average enrollees	0.09%
Live-end-of-year enrollees	-0.34%
Unique patients *	0.19%
Obligations *	2.60%

* Adjusted for new enrollees on primary care wait list as of September 2002

In 2002 and 2003, VA conducted CARES – Capital Asset Realignment for Enhanced Services – following a pilot study that was completed in 2001 for one VHA health care network. The CARES process is a strategic planning process to address the future infrastructure, i.e., beds, outpatient capacity, and other services, to meet the health care needs of veterans in the future. During the CARES process, hundreds of VA staff used the model's projections. One significant outcome of CARES was a comprehensive review of the model by planners and local health care managers. Their questions and comments identified a number of areas for further investigation and opportunities for improving the model. Eight advisory groups were formed with representation from both VA and VHA that provide subject matter experts as we explore more than a dozen model enhancements for the next version of the model that will be used for the FY 2005 budget and strategic planning process. Table 5 provides a listing of areas within the VA Enrollee Health Care Model that have been identified for further study and possible improvement.

Table 5. FY 2004 VA Enrollee Health Care Projection Model Improvements

Enrollment rates	Model validation study
Veteran enrollee mortality study	Actual to expected analysis
Enrollee migration	VHA unit costs
Special disability population projections	Non-medical benefit projections
Service line projections	Prosthetics utilization
Patient projections	Enhanced ability to modify copay and covered benefits
Front-end user interface	Impact of enrollment fees

VHA and VA continue to make progress improving the model by working in the areas identified in Table 5. For example, more recent, more historical data has been analyzed and VHA has concluded that constant enrollment rates that vary by sector, age group, priority level and enrollee type are appropriate in modeling future enrollment for budgeting purposes. This may not be the case for long-term strategic

planning purposes and work on this issue continues. The model also projects for the migration of veterans between priority levels and movement of their primary residence. We have learned that a percentage of veterans will move to a higher priority once enrolled, particularly during their first twelve months of enrollment. While this is the case for many veterans, we have observed some veterans moving to a lower priority. This is important to the Department since the Secretary's annual enrollment decision is based on our projections and many of factors in the model vary by priority level. A number of services provided by VHA are not considered typical of a private sector medical benefits package and VHA has developed some internal models to project future demand. These include services such as inpatient care for Spinal Cord Injury and Blind Rehabilitation, just two areas where VA excels as a leading provider of health care and rehabilitative services. Another major enhancement to the model is the introduction of VA prosthetics utilization and cost data for durable medical equipment, hearing aids and eyeglasses. Another model enhancement is a front-end user interface, which was previously discussed in this paper.

In light of budgetary constraints and recognition of VHA's budget as a discretionary program, VHA identifies alternative policies to control the demand for health care services. These may include changing the copayments for certain services required of some priority levels of veterans for their nonservice-connected conditions, requiring an annual enrollment fee (rather that monthly premiums charged by private sector health plans), suspension of enrollment for a priority level of veterans or disenrolling enrolled veterans with the lowest eligibility for health care.

The model has tangible benefits for the Department such as enabling VA to develop a health care budget that justifies its base funding level. It also brings the credibility of a respected actuarial firm. The model combines the knowledge and capabilities of VA and the private sector to make a powerful team and enables a sophisticated level of analysis and reporting that VA could not produce cost effectively in house.

The FY04 Budget

The Secretary of Veterans Affairs suspended enrollment for Priority 8 veterans who are nonservice-connected and not enrolled prior to January 17, 2003. This decision impacted a projected 174,000 veterans in FY03. The decision was based on several factors including the developing conflict in Iraq, long waiting times for access to health care for new enrollees and the projected funding shortfalls in appropriations for VA.

The President's 2004 budget proposal was submitted to Congress earlier this year requesting $27.5 billion for medical care, including over $2 billion in collections. This represents a 7.7 percent increase from the 2003 budget. Included in the President's budget were several proposed legislative and regulatory changes in addition to a request for increased resources. There were four major changes in the President's budget:

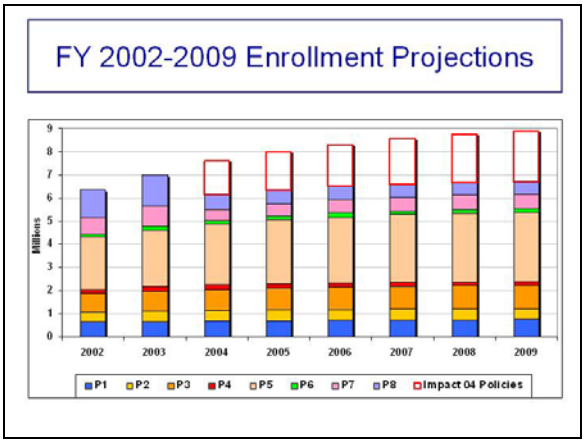
- assess an annual enrollment fee of $250 for nonservice-connected Priority 7 veterans and all Priority 8 veterans
- increase copayments for nonservice-connected Priority 7 and all Priority 8 veterans for outpatient primary care from $15 to $20 and for pharmacy benefits from $7 to $15
- eliminate the pharmacy copayment for Priority 2-5 veterans who income is below the pension aid and attendance level of $16,169
- expand non-institutional long-term care with reductions in institution care in recognition of patient preferences and the improved quality of life possible in non-institution settings.

As the budget process continues, VHA has been reviewing its projections, and assessing the impact of the President's budget proposals that require legislation or regulatory changes. Legislative items that are outside the control of the VA are the $250 annual enrollment fee and increasing the pharmacy copayment from $7 to $15. What is the impact on VHA's budget if these legislative initiatives are not enacted? The VA Enrollee Health Care Projection Model provides VA with the capability to assess and understand the impact of these two proposals on enrollment and utilization.

For FY04, the President's budget proposal included continued suspension of Priority 8 veterans as well as the policies described above. Results from the VA Enrollee Health Care Projection model show that while increasing the pharmacy copay has a marginal impact on utilization; it is not expected to impact a veteran's decision to enroll in VHA. From our model and in consultation with our actuaries, VA knows that establishing an annual enrollment fee, on the other hand, has significant impact on the number of Priority 7 and 8 veterans that will enroll or remain enrolled with VHA. The model projects that over 1.4

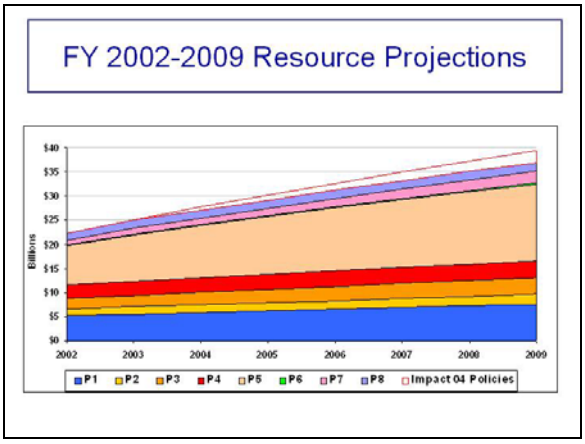
million or approximately 50 percent of Priority 7 and 8 enrollees would leave VHA if they were required to pay a $250 annual enrollment fee. Figure 2 shows the impact of the enrollment fee and copay changes on enrollment for fiscal years 204 through 2009. Average enrollment for Priority Groups 7 and 8 decrease 42 percent from 7.004 million in FY03 to 6.684 million by FY 2009; Priority Groups 1-6 increase 25 percent during this same period. This is because many veterans, particularly those who are nonservice-connected, enroll with the VA but may not receive any health care from VA. Many choose to enroll because it costs them nothing; they view VHA as a safety net for themselves. The number of Priority 7 and 8 users (patients) expected to leave VHA is smaller for the reasons previously stated. The impact on users would reduce the number of Priority 7 and 8 users by approximately 25% or just over 309,000. Priority 7 and 8 patients who used to VA on an infrequent basis are expected to leave the system.

Figure 2. Average Enrollment - FY03 VA Enrollee Health Care Projection Model



The next chart (Figure 3) shows the impact of suspension of enrollment for Priority 8 veterans, establishment of a $250 annual enrollment fee and copay changes for primary care and medications on resource requirements for VHA. The chart shows projected impact of these proposals will result in cost avoidance of $704 million in FY 2004. In FY 2009, VHA resource requirements are reduced by $2.4 billion if these proposals are implemented. Without these policies, VHA's budget requirements will grow from $22 billion in FY02 to nearly $40 billion in just seven years.

Figure 3. Expenditures - FY03 VA Enrollee Health Care Projection Model



If the enrollment fee and copay changes proposed in the President's FY 2004 aren't considered by Congress, VA's health care system will face many challenges in the future. The impact of such policy decisions on a health care system as large as the VHA are marginal in the first years of implementation but their impact in the out years is substantial. Suspension of Priority 8 enrollment impacts 174,000 veterans in FY 2003, 520,000 in FY 2005 and 1.1 million by FY 2012. In terms of dollars, these policies result in cost avoidances of $800,000 in FY 2004 and $2.8 billion by FY 2012.

Veteran demand for VA health care has outpaced the resources appropriated and available to VHA. In response, VHA continues to identify other opportunities to close the gap between demand and resource availability. Strategic policies that may close this gap may be grouped into five basic areas, some of which were discussed in this paper. They are: enrollment actions, services provided, cost sharing proposals, efficiencies (administrative and clinical) and resources. The model projects the impact of suspending enrollment or disenrollment, as well as limiting the services offered as part of the medical benefits package to all enrollees or selected priority groups. Cost sharing proposals range from charging an annual enrollment fee, a first-use fee, an annual deductible for health care as well as copayments for nonservice-connected conditions. Clinical and administrative/operational efficiencies are also modeled as the VHA improves its health care delivery system.

VHA has demonstrated that this model has delivered many tangible benefits for the agency. With the VA Enrollee Health Care Projection Model, VA has developed a health care budget that justifies its base funding requirements. These projections have been accepted by OMB and have been used to develop the President's budget for the past two years. The analyses carry the credibility of a respected actuarial firm; combining the knowledge and capabilities of VA and the private sector. VA must continue to constantly assess and validate its demand and utilization projections to ensure that its resource needs are clearly articulated to the White House and to Congress. Additional management actions must be considered to ensure that VHA provides veterans with a first class health care system.

Transportation and Energy Forecasting

Chair: Brian W. Sloboda, Bureau of Transportation Statistics, U.S. Department of Transportation

U.S. Greenhouse Gas Models: An Evaluation From A User's Perspective

David Chien, Bureau of Transportation Statistics, U.S. Department of Transportation

Since the Kyoto Protocol in 1997, greenhouse gas forecasting has been very prevalent in the news media and in the literature. Several forecasts in the past have been made by the U.S. Government at the request of the President and Congress to estimate the future greenhouse gas emissions from the transportation sector. Essential to the debate on greenhouse gases are the models and the data that drive the results. This presentation will review some of the data available to measure greenhouse gas emissions from the United States, and the statistical models that the U.S. Federal Government uses to evaluate potential greenhouse gas policies and generate greenhouse gas forecasts.

Issues in Developing a Transportation Infrastructure Index

Brian W. Sloboda, Bureau of Transportation Statistics, U.S. Department of Transportation
Herman Stekler, Department of Economics, The George Washington University

Lahiri et. al. (2002) presented the theoretical development, selection, and the testing of the Index of Output of Transportation Services. This Index serves as a coincident indicator of economic activity in the services sector of the transportation industry. This monthly index of transportation output covers the period of 1980:1-2002:12, and measures the economic activity for the transportation modes of air, rail, water, truck, transit, and pipelines. However, this Index only measures the activity of the transportation services sector while the transportation industry also includes the transportation equipment and transportation infrastructure sectors. This paper will explore in detail the data and classification problems that are involved in developing a measure of economic activity in the infrastructure sector of the transportation industry.

Business Cycle Analysis for the U.S. Transportation Sector

Kajal Lahiri and Wenxiong Yao, Department of Economics, SUNY-Albany
Peg Young, Bureau of Transportation Statistics, U.S. Department of Transportation

Since most of final and intermediate goods in an economy are moved by the transportation sector, indicators of this sector may have strong forecasting value for the overall economy. We have developed monthly coincident and leading indicators for the transportation sector. Four coincident indicators—transportation output, employment, payroll, and personal consumption expenditure—and a number of leading indicators were selected. A composite coincident index was constructed using both the conventional National Bureau of Economic Research approach and a regime-switch state space model using Gibbs-sampling methodology. We identified the business cycle chronology for the U.S. transportation sector, selected leading indicators, and a composite leading index was constructed. The growth cycles in the U.S. transportation sector were compared with those of the overall economy.

U.S. GREENHOUSE GAS MODELS: AN EVALUATION FROM A USER'S PERSPECTIVE[1]

David Chien
U.S. Department of Transportation,
Bureau of Transportation Statistics

Introduction

In December 1997, approximately 160 nations met in Kyoto Japan and developed the Kyoto Protocol, which would limit developing nations to 1990 greenhouse gas emissions (GHG) levels while the U.S. agreed to reducing GHG emissions levels to 7 percent below 1990 levels from 2008 to 2012. Therefore, it is vitally critical to review those models , which the U.S. federal government uses to estimate GHG emissions under several scenarios. Although the scenarios and data inputs used in many analyses will not be evaluated, the contents of the models will be reviewed. Although we are not evaluating the input data, the maxim "garbage in and garbage out" still applies. Therefore, the author would highly recommend reading the model documentation and visiting the model websites, which are at the beginning of each model description section (in the footnotes) in order to gain more detailed information on each model.

Many of the data sources used to develop these models resides at the federal agencies that have developed and currently maintain the models. Among those that are used most frequently for estimates of historical GHG emissions are EPA[2] (Environmental Protection Agency), and EIA[3] (Energy Information Administration). Transportation and energy related data can be found at the Transportation Energy Databook website run by Oak Ridge National Laboratory (ORNL) and USDOE[4] (U.S. Department of Energy), the USDOT (U.S. Department of Transportation) BTS website[5] and the BTS TRANSTATS website.[6] The TRANSTATS website contains the NTS (National Transportation Statistics) data, originating from many BTS surveys (Office of Airline Information databases, NHTS (National Household Transportation Survey), CFS (Commodity Flow Survey), and many more).

Although this article is designed to provide potential model users with brief descriptions of the GHG models used by the U.S. federal government, it also evaluates the models based on key operational factors that are often overlooked by potential users. Examples would include topics such as: the size of the data inputs and source code of the models , the hardware and software platform and requirements, the run time or amount of time associated with execution of the models, the resources needed to develop and maintain the models, and examples of studies which have extensively used the models. Detailed coverage of the models with respect to the transportation sector will also be evaluated.

1 Although this article was written by the author, there was a heavy reliance on the following more detailed document: U.S. Department of Transportation (USDOT), US DOT Center for Climate Change & Environmental Forecasting, prepared by Kevin Greene of the Volpe National Transportation Systems Center, *Transportation Greenhouse Gas Emissions Data & Models: Review and Recommendations*, March 2003, Cambridge, Mass.

2 http://yosemite.epa.gov/oar/globalwarming.nsf/UniqueKeyLookup/RAMR5CZKVE/$File/ghgbrochure.pdf

3 http://www.eia.doe.gov/env/ghg.html

4 http://www-cta.ornl.gov/cta/data/Index.html

5 http://www.bts.gov/

6 http://www.transtats.bts.gov/

National Energy Modeling System (NEMS) Model[7]

Energy Information Administration, U.S. Department of Energy

"NEMS is a computer-based energy-economy modeling system of the U.S. energy markets for the midterm period through 2025. NEMS annually projects the production, imports, conversion, consumption and prices of energy, subject to assumptions on macroeconomic and financial factors, world energy markets, resource availability and costs, behavioral and technological choice criteria, cost and performance characteristics of energy technologies, and demographics. The purpose of NEMS is to project energy, economic, environmental, and security impacts on the United States of alternative energy policies and of different assumptions about energy markets."[8] Congress and other federal agencies have extensively used NEMS to evaluate energy and transportation policies. The model has the advantage of extensive peer review by the U.S. transportation community including USDOE, USDOT, EPA, Office of Management and Budget (OMB), General Accounting Office (GAO), and the National Academy of the Sciences.

The modeling structure of NEMS consists of integrated models representing all sectors of the economy (residential, commercial, industrial, transportation) including a macro-economic component, and energy supply sources (crude oil supply, oil refinery, oil distribution module, natural gas model including exploration and drilling and distribution, electricity model including nuclear, coal, natural gas, residual fuel, and small generators like wind and solar, coal model, and renewable fuels).

The following discussion will be on the NEMS Transportation Model (TRAN) and the degree to which it covers the transportation sector. TRAN has a wide coverage of the aggregate transportation system including the following modules: LDV (light-duty vehicle) Module (for cars and light-trucks), Aviation Module (wide and narrow-body, GA (general aviation) for passengers, and freight), Freight Truck Module (medium and heavy-duty trucks for freight), Rail Module (passenger and freight), Waterborne Module (passenger and freight), Miscellaneous (Military, Mass Transit, Recreational Boats, etc.), and Emissions Module.[9]

The LDV module has 6 sub-modules: Fuel Economy Module (6 car and 6 light truck EPA size classes across 63 fuel savings technologies), Regional Sales Model (9 Census Divisions), Alternative-Fuel Vehicle Module (12 types of alternative-fuel vehicles), Light-Duty Vehicle Stock Module (vehicle retirement curves and capital stocks by 20 vintages and vehicle type),Vehicle-Miles Traveled (VMT) Module (by car and light truck as a function of income per capita, cost of driving per mile, female to male annual VMT ratio, age distribution of population, population growth), which currently uses National Household Transportation Survey (NHTS)/National Personal Transportation Survey (NPTS)/American Travel Survey (ATS) data, and the LDV Fleet Module for business, government, and utility fleets (as part of the Energy Policy Act or EPACT).

The Air Travel Demand Model forecasts revenue passenger-miles (RPM) for business and personal travel, international and domestic travel, revenue ton-miles (RTM) for freight, and seat-miles demanded (SMD). The Aircraft Fleet Efficiency Model has several fuel-saving technologies for both narrow and wide-body aircraft, and contains over 25 vintages of aircraft with aircraft survival curves and stock model representation. There are 6 advanced fuel saving technologies: ultra-high bypass, propfan, improved thermodynamics, hybrid laminar flow, improved aerodynamics, and weight reduction.

The Freight Truck Module uses macro-economic gross outputs by STCC (Standard Transportation Commodity Classifications) industrial code in determining VMT (vehicle-miles traveled). The Commodity Flow Survey (CFS) and the Truck Inventory and Use Survey (TIUS) are

[7] NEMS Model contact: Mary Hutzler, USDOE, EIA; NEMS Transportation Model contact: John Maples, USDOE, EIA; http://www.eia.doe.gov/oiaf or http://www.eia.doe.gov/bookshelf/docs.html

[8] U.S. Department of Energy, Energy Information Administration, *The National Energy Modeling System: An Overview 2003*, DOE/EIA-0581(2003), March 2003, Washington, D.C.

[9] U.S. Department of Energy, Energy Information Administration, *The Transportation Sector Model of the National Energy Modeling System: Model Documentation Report*, DOE/EIA-M070(2003), February 2003, Washington, D.C.

extensively used to establish the connection between commodities and mode of travel. The Truck Stock Model uses capital stocks by truck size and age, which allows the modeler to bring in new higher efficiency truck technologies. Technology choice is based on commercial availability, fuel prices, capital cost, and other cost-effectiveness criteria such as discount rates and payback period. There are over nine future advanced fuel saving technologies and numerous current truck fuel saving technologies.[10] Gasoline, diesel, natural gas, and liquid petroleum gas (LPG) are the fuels and fuel truck technologies represented in the Freight Truck Module.

Rail and Waterborne Modules also use ton-miles traveled estimated equations based on industrial output by STCC code. Energy efficiency for new and stock of old vehicles is estimated. A major drawback of the model is the lack of capital stocks and vintaging by age. Therefore, the growth rates of efficiency improvements must be made exogenously based on trends rather than an explicit endogenous calculation of the model. Specific technology representation and turnover cannot be endogenously determined, which limits the effect of advanced technologies over time, unless of course the modeler pre-determines this in the exogenous input file. Overall, this section of TRAN has no sensitivity to fuel prices or the cost of travel in either travel or efficiency forecasts.

The Mass Transit Module includes three types of passenger rail (transit, commuter, and intercity). Passenger buses (transit, intercity, and school) are also included, bringing the total modes of travel to six for the Mass Transit Module. Travel is estimated for all 6 transit modes as a function of the relative historical growth rate passenger-miles traveled relative to light-duty vehicle passenger-miles. Growth rates of efficiency improvements are calculated based on the growth rates of similar technology modes. This assumes that technology advancements will parallel those in modes using the same vehicles. For example, mass transit rail efficiencies would then be assumed to grow at the same rate as Class I freight rail. Therefore, the same caveats from the Rail and Waterborne models apply to the Mass Transit Module since both lack explicit model responsiveness to fuel prices and travel costs.

TRAN also has an emissions module, which can forecast emissions of the criteria pollutants SO_x, NO_x, HC, CO, and also CO_2. Most recently, TRAN has incorporated the EPA Mobile 6.0 model, which is used by EPA and several state governments to calculate regional emissions.

The Macro-Economic Module currently consists of the Global Insight (formerly DRI/WEFA) model of the U.S. Economy, Industry Model, Employment Model, and Regional Model. One issue in using the input-output (I-O) National Accounts data is that it undercounts the effects of the transportation system upon the economy due to the exclusion of almost all private commercial businesses, which have their own private transportation, and are currently counted under commercial operations. Potential improvement to the model would be to adjust the I-O data with the BTS Transportation Satellite Accounts (TSA), which attempts to measure and adjust the I-O accounts with the private transportation associated with commercial operations. Despite these issues, the Macro-Economic Module is a key part of measuring the impacts of potential GHG strategies upon the economy. This component of NEMS is one of the most important parts of NEMS because it is essential to the convergence process, and it fully integrates the economy with the modeling process, which many of the other GHG models reviewed in this article do not possess. Reaching equilibrium in a large model of this size is of paramount importance, especially because feedback effects of prices upon transportation services has a tendency to be dampened significantly when macro-economic feedback with the rest of the model is turned on. What does this tell us? The conclusion is that models, which do not have this capability, have a tendency to overstate the effects of any given policy that may be implemented, because they do not account for economic changes and responses to those changes. Reaching an equilibrium solution is critical to the accuracy in the measurement of costs and benefits of any policy or program.

One of the drawbacks to using the model is also one of NEMS' greatest strengths, the size of the whole NEMS model is very large, requiring over

[10] Argonne National Laboratory, prepared for the U.S. Department of Energy, Energy Information Administration, *Heavy-and Medium-Duty Truck Fuel Economy and Market Penetration Analysis for the NEMS Transportation Sector Model*, August 1999, Washington, D.C.

10-15 megabytes of storage just for the "restart file," which contains the starting values for the model each year. In order to run a "standalone " run, which consists of running only one model and keeping the others at reference case levels, would require 100 megabytes of storage space. Although NEMS can be installed on an individual PC, the storage requirements are substantial. Hardware should consist of 512 megabytes of RAM and a 486 or Pentium processor. The model operates in Compaq Visual FORTRAN, and requires the EViews software. If the user wanted to also run the supply models, then OML, a linear programming software, is also necessary. When running in "standalone" mode with only one model endogenously turned on or active, the model will return a solution within a few minutes. However, submitting a fully integrated run with all of the modules turned on or active would take about 2-4 hours depending on how many changes were made to the model. The current NEMS model at EIA employs approximately 40 full-time employees and 20 contractors. Therefore, enhancing, updating, and maintaining the model requires significant resources. However several agencies and National Laboratories maintain versions of the NEMS models and they usually employ about 2-4 people to operate and maintain the model. These NEMS Model clones require receiving the updates to the models from EIA annually.

Energy Markal-Macro Model[11],

Brookhaven National Laboratory and U.S. Department of Energy

The Energy Markal-Macro Model at the U.S. Department of Energy (DOE) is a dynamic linear programming model system of two models, Markal and Macro. Markal is the "bottom-up" technological model of energy and environment, which includes depletable and renewable natural resources, processing of energy resources, and end-user technologies for all sectors. Macro is the "top-down" macro-economic growth model that links Markal to the economy and maximizes utility (discounted sum of consumption). Markal-Macro finds the least-cost dynamic equilibrium under specific market and policy assumptions. At the Department of Energy, the Energy Markal-Macro Model is calibrated to the NEMS model outputs annually.

The Markal-Macro Model is used by over 35 countries and was developed by Brookhaven National Laboratory and then further developed by 18 OECD countries.

Markal-Macro optimizes the mix of fuels and technologies based on the consumer discount rate, technology characteristics, and fuel prices. Marginal costs for technologies and applications are used to determine the most efficient level of energy inputs along with technology costs and energy efficiencies. Emission sources and levels are forecast for CO_2, SO_X, and NO_X. The value of carbon rights (marginal cost of emissions) is one of the important outputs of the model. Outputs are solved in five year intervals through 2050. Transportation coverage includes passenger cars, light trucks, heavy trucks, buses, airplanes, shipping, passenger rail, and freight rail.

The model can output a business as usual energy and carbon emission profile. Identification of dynamic technology paths to meet emissions growth targets is one of the more common uses of the model outputs. Costs of alternative approaches to reducing carbon emissions has been studied frequently by many countries using the Markal-Macro Model. Policy options would include fuel switching, substitution of capital and/or labor for energy services, demand reduction, etc. Markal-Macro can also identify opportunities for reducing carbon emissions through supply and demand technologies. Based on the technologies chosen, the model can calculate the cost of carbon emission reductions.

U.S. Department of Energy has used the Energy Markal-Macro model to analyze the Energy Policy Act of 1992. The Energy Information Administration has also built an international version of the Markal-Macro model called SAGE. U.S. EPA (Environmental Protection Agency) is developing a national Markal database and scoping out a regional Markal representation of the U.S. economy. Markal-Macro is used by over 35 countries to support environmental planning. The International Energy Agency also has a version of the Markal-Macro Model, which they use for energy technology scenarios. Most recently the model has focused on externalities measurement,

[11] USDOE Energy Markal-Macro Model contact: Philip Tseng, USDOE, Energy Efficiency and Renewable Energy Office; http://www.etsap.org

hydrogen economy development, cost-competitive life cycle analysis, oil market response, technology learning, and country analysis.

There are a few limitations of Markal-Macro, such as it does not cover all sectors as the NEMS Model. However, it can provide an alternative and complimentary approach (projection of renewable fuel penetration and reducing carbon dioxide emissions). The model is a very aggregate model that forecasts energy demand based on housing stocks, commercial floor space, industrial production index, and vehicle-miles traveled.

The data inputs to the model use about 7-20 megabytes of storage space, and the sourcecode is approximately 7-10 megabytes. The model can be run on a Pentium IV processor with a 2 GHz processor speed and 256 MB of RAM. Model execution is fairly quick at around 5 minutes. The model is quite complicated and requires special skills to run, similar to the NEMS model, but with many less people. The Department of Energy has about 2 National Laboratory analysts using and maintaining the model.

Mini-Cam Model[12]

Pacific Northwest National Laboratory (PNL)

The Mini-Cam Model forecasts carbon dioxide and other GHG's emissions, and estimates the impacts on GHG atmospheric concentrations, climate, and the environment. Although the model is a "top-down" energy-economy model, it contains "bottom-up" assumptions about end-use energy efficiency. Projections are made through 2100 and therefore, the model has more advanced technologies than NEMS. The model outputs forecasts in fifteen yearly increments. Projections cover the entire planet in 14 global regions: U.S., Canada, Western Europe, Australia and New Zealand, Japan, former Soviet Union, Eastern Europe, China, Southeast Asia, Middle-East, Africa, Latin America, South Korea, and India.

[12] Mini-Cam Model contact: Son H Kim, Pacific Northwest National Laboratory; http://www.pnl.gov/aisu/pubs/chinmod2.pdf and http://sedac.ciesin.org/mva/minicam/MCHP.html

Mini-Cam is comprised of two larger models: Edmonds-Reilly-Barns Model (ERB) and the Model for the Assessment of GHG Induced Climate Change (MAGICC). ERB represents the Energy/Economy/Emissions system, including supply and demand of energy, the energy balance, GHG emissions, and long-term trends in economic output. MAGICC models the atmospheric/climate/sea-level system, which includes a Gas Cycle, climate, and sea-level model. MAGICC outputs atmospheric composition, radiative forcing, global mean temperature change and sea-level rise.

Energy supply and demand are calculated in the model. Energy demand is a function of population, labor productivity, economic activity, technological change, energy prices, and energy taxes and tariffs. Energy supply of renewable and non-renewable sources are dependent upon resource constraints, behavioral assumptions, and energy prices by region. Transportation is one of 3 sectors (residential/commercial and industrial) and includes passenger and freight technologies and modes. Model inputs consist of total service, service cost, energy intensity, load factor, price and income elasticities, technical change, percent of population licensed to drive, and average speeds. The transportation system coverage includes automobiles, light trucks, buses, rail, air, motorcycles for passenger modes; and trucks, rail, air, ship, pipeline, and motorcycles for freight modes. Six major energy sources are modeled including oil, gas, solids (coal and biomass), resource-constrained renewables, nuclear, and solar.

The current Mini-Cam Model has an executable file size of about 1 megabyte and the data input files are about the same size. The source code is approximately 903 KB. Run time is approximately 30 seconds on a Pentium 4 with 1.7 GHz and is also dependent upon the number of scenarios run at one time. Mini-Cam can operate on a Pentium III or higher speed processor. FORTRAN is the modeling language using a MS Visual Studio compiler. However there is a GUI (graphics user interface) front-end to the model if desired, which requires MS Acesss and MS Excel software. With the GUI, the user can run multiple scenarios at once, and query, view and chart results. Currently two people use and maintain MiniCam at PNL.

TRANSIMS Model[13]

Los Alamos National Laboratory (LANL)

"TRANSIMS creates a virtual metropolitan region with a comprehensive representation of its population, the population's activities, and the transportation infrastructure. Building upon these factors, TRANSIMS simulates the movement of individuals across the transportation network, including their second-by-second use of vehicles."

TRANSIMS is a network-based metropolitan area traffic simulation modeling system. At the early stages of development, TRANSIMS will be used to model Portland, Oregon, which is the first city to use the model. A commercialized version of TRANSIMS is being developed by Pricewaterhouse Coopers. A TRANSIMS-like national freight model is in the early stages of development (NTNAC or National Transportation Network Analysis Capability). TRANSIMS uses local factors such as land use, street design, and transportation infrastructure.

TRANSIMS reports the amounts of fuel used within the transportation network based on vehicle-by-vehicle fuel burn rates and modes of operation on a second-by-second basis. Outputs include: trip forecasts, criteria pollution, and fuel usage by type. Trip forecasts require details about the mode of travel, number of passengers per vehicle, distance, and duration. Outputs can be at the vehicle level, metropolitan area, neighborhood, road segment, or a subgroup of the population by time of day. Subgroups can be defined by population characteristics, types of vehicles owned or by types of activities of the population.

TRANSIMS can be combined with other emissions models to calculate atmospheric conditions, local emissions transport and dispersion. The model can handle policy issues such as congestion pricing, traffic flow, infrastructural changes upon traffic, transportation control measures, Intelligent Transportation Systems (ITS), and motor vehicle emissions. Micro-level detail of transportation networks is required, including traffic signals, merging and turning lanes, pedestrian impediments, and topography, which can allow detailed analysis of time of day usage of the network. This makes TRANSIMS a very good model to use to evaluate traffic flows and patterns relative to potential policies to increase traffic flows or redirect traffic patterns.

As with all very large models, their strengths of detail and coverage or almost always offset by their sheer size, hardware requirements, detailed knowledge and expertise, run times, maintenance, and resource allocation. TRANSIMS requires multiple super computers to run the model. "The Portland Metro linux cluster is 30 or 32 compute nodes, each node having two 1.2 GHz processors that share 2 GB of memory. There are also four 1.2 GHz processor servers with 6 GB of shared memory. Data storage is in the 500 GB range. The TRANSIMS-LANL source code size is about 23 MB and consists of several modules and utilities. The network input data is about 108 MB."[14] Run-time can exceed a couple of days. Reviewing the statistics of the model requires much more time in order to make adjustments, such as rerouting traffic. The model requires specialized knowledge and skills to operate the model. TRANSIMS uses extremely detailed data and enormous data inputs and outputs, which are fed from one simulation generator to another. Updating data may require significant resources, as well as re-calibration and adjustments related to data updates. Relevancy of data inputs and model equations over time periods exceeding perhaps months may be questionable, which may necessitate data updates to handle long-term projections. Seasonal and monthly variations would be expected to vary significantly especially with regards to time of day travel activities.

GREET (Greenhouse Gases, Regulated Emissions, and Energy Use in Transportation) Model[15],

Argonne National Laboratory (ANL)

[13] TRANSIMS Model contact: LaRon Smith, Los Alamos National Laboratory; http://www.lanl.gov or http://transims.tsasa.lanl.gov/ or see Nagel, K, Beckman, R.J., Barrett, C.L., *TRANSIMS for urban planning*, LA-UR-98-4389, Los Alamos, New Mexico, 1998.

[14] Based on conversations with LaRon Smith at LANL.

[15] GREET Model contact: Michael Quanlu Wang, Argonne National Laboratory; http://www.transportation.anl.gov/pdfs/TA/153.pdf.

"The GREET model is intended to serve as an analytical tool for use by researchers and practitioners in estimating fuel-cycle energy use and emissions associated with alternative transportation fuels and advanced vehicle technologies."

GREET provides full fuel cycle emissions analysis from wells to wheels, which represents emissions from all phases of production to arrival at a gas station. However the model does not include more tertiary sources such as vehicle production, disposal and recycling. The strength of GREET is that it analytically compares emissions from vehicle technologies matched with several fuels, especially very advanced alternative-fuels.

The beauty of GREET is that it has a substantial combination of vehicle technologies and fuel types. GREET contains the following powertrains: conventional, direct injection, spark ignition, compression ignition, hybrid electric vehicles which can be grid connector or not, electric vehicles, and fuel cell vehicles. Fuel types are also numerous: gasoline which comes reformulated or non-reformulated, diesel and low sulfur diesel, , CNG, LPG, LNG, Dimethyl Ether, FT (Fischer-Tropsch) Diesel, gaseous and liquid hydrogen, methanol, ethanol, biodiesel, and electricity. These powertrains and fuel types can be produced from several feedstocks: petroleum, natural gas, flared gas, landfill gas, corn, cellulosic biomass, soybeans, and electricity. GREET is excellent as an emissions model to determine individual vehicle emissions, and would be valuable in assisting to set or meet emissions standards. EPA has decided to include GREET within their air emissions model Mobile 6.

The hardware requirements to run and operate GREET are: GREETGUI (GREET with a GUI interface or front end) works on PCs with Microsoft's Windows 95 or later, but Windows 98 or greater is best. Minimum hardware requirements are a Pentium III processor at 166 MHz or higher, at least 64 MB RAM; and at least 30 MB of free space on the hard drive. Recommended hardware profile: Pentium processor at 400 MHz or higher, 128MB or more of RAM, 100MB of free hard disk space or more.[16]

GREET can also run on a spreadsheet model, which takes about 5 MB on an EXCEL spreadsheet. GREET recently added a Monte Carlo simulation module, which stochastically generates a distribution rather than a point estimate. Running the model would normally be almost instantaneous, but with the simulation, run times may be approximately 3 ½ hours. Four people developed and are currently maintaining and running GREET at ANL.

GREET only applies to light-duty vehicles. However, this does not preclude it from being done on other vehicle types in the future perhaps. GREET does not include a vehicle choice model to forecast what people might purchase based on consumer preferences. However, the model may be used in combination with policy options to reduce emissions and set emissions standards to achieve a goal.

TAFV (Transitional Alternative-Fuels and Vehicles) Model[17]

Oak Ridge National Laboratory and the University of Maine

TAFV represents economic decisions among auto manufacturers, vehicle purchasers, and fuel suppliers, including distribution to the end users. The model simulates decisions during a transition from current fuels to alternative-fuels, and traditional vehicles to advanced technology vehicles. Limited availability of alternative-fuels, including refueling infrastructure, and availability of alternative-fuel vehicle technologies are inter-dependent. TAFV assumes retail alternative fuel providers will maximize profits, and spread capital costs across outlets to increase availability.

TAFV also contains a model for predicting choice of alternative fuel and alternative vehicle technologies for light-duty motor vehicles. The

[16] Argonne National Laboratory, *Development and Use of GREET 1.6 Fuel Cycle Model For Transportation Fuels and Vehicle Technologies*, ANL/ESD/TM163, Center for Transportation Research, Energy Systems Division, June 2001, Argonne, Illinois.

[17] TAFV Model contacts: Paul Leiby and David Greene, Oak Ridge National Laboratory, and Jonathan Rubin, University of Maine; http://pzl1.ed.ornl.gov/altfuels.htm

nested multinomial logit (MNL) mathematical framework is used to estimate vehicle choice among technologies and fuel type combinations based on consumer preferences and vehicle attributes. Vehicle choice is dependent upon prices, luggage space, fuel availability, refueling time, vehicle performance, cargo space, and vehicle offerings. AFV's (alternative-fuel vehicles) have 3 costs on vehicle manufacturers: capital costs, variable costs, and costs associated with diverse vehicle offerings. Calibration of the model through some key parameters such as the value of time and discount rates is based on existing literature. A spreadsheet model has been developed for calibration and preliminary testing of the model.

Limitations of TAFV include: a) TAFV only includes light-duty vehicles, b) growth rates in transportation demand, and oil and gas prices are exogenous, c) it is not clear if TAFV includes federal mandates for vehicle acquisitions (i.e. policies such as the Low Emission Vehicle Program, and the Energy Policy Act).

The model is actually quite small at 208K but data inputs could be a few megabytes of spreadsheet data). The main program is written in the GAMS (Generalize Algebraic Modeling Language) Language. TAFV uses the MINOS5 and CONOPT2 nonlinear optimization solvers. The source code is about 111K, in GAMS language, but the model requires many (>100) megabytes to execute. A model run takes approximately 30-60 minutes on a Pentium III 1000 MHz PC. Work files that are generated during a run can approach 1 GB. It is recommended that users have 128 MB or more memory. TAFV can be run on Windows, Linux, Unix, depending on which platform the licensed GAMS software resides on.
Development required a team of 5 for 3 years. Maintenance currently involves a team of 2. However, plans in the future are for a team of 5 over the next 2 years.

References:

1) Argonne National Laboratory, prepared for the U.S. Department of Energy, Energy Information Administration, *Heavy-and Medium-Duty Truck Fuel Economy and Market Penetration Analysis for the NEMS Transportation Sector Model*, August 1999, Washington, D.C.
2) Argonne National Laboratory, *Development and Use of GREET 1.6 Fuel Cycle Model For Transportation Fuels and Vehicle Technologies*, ANL/ESD/TM163, Center for Transportation Research, Energy Systems Division, June 2001, Argonne, Illinois.
3) Energy Technology Systems Analysis Program, *ETSAP Newsletter*, Volume 8, no. 2, August 2003; http://www.etsap.org
4) Leiby, Paul, and Jonathan Rubin, The Alternative Fuel Transition: Results from the TAFV Model of Alternative Fuel Use in Light-Duty Vehicles 1996-2010 (Final Report, TAFV Version 1), September 17, 2000.
5) Nagel, K., Beckman, R.J., Barrett, C.L., *TRANSIMS for urban planning*, LA-UR-98-4389, Los Alamos, New Mexico, 1998.
6) Pacific Northwest National Laboratory, Advanced International Studies Group, *China-Korea-U.S. Economic Environmental Workshop Conference Proceedings*, May 23-25, 2001; http://www.pnl.gov/aisu/pubs/chinmod2.pdf and http://sedac.ciesin.org/mva/minicam/MCHP.html
7) U.S. Department of Energy, Energy Information Administration, *The National Energy Modeling System: An Overview 2003*, DOE/EIA-0581(2003), March 2003, Washington, D.C.
8) U.S. Department of Energy, Energy Information Administration, *The Transportation Sector Model of the National Energy Modeling System: Model Documentation Report*, DOE/EIA-M070(2003), February 2003, Washington, D.C.
9) U.S. Department of Energy, prepared by Stacy Davis and Susan Diegel of the Oak Ridge National Laboratory, *Transportation Energy Databook: Edition 22*, ORNL 69-67, Oak Ridge, Tennessee, September 2002.
10) U.S. Department of Energy, Energy Information Administration, *Emissions of Greenhouse Gases in the United States 2001*, DOE/EIA-0573(2002), December 2002, Washington, D.C.
11) U.S. Department of Transportation (USDOT), US DOT Center for Climate

Change & Environmental Forecasting, prepared by Kevin Greene of the Volpe National Transportation Systems Center, *Transportation Greenhouse Gas Emissions Data & Models: Review and Recommendations*, March 2003, Cambridge, Mass.

12) U.S. Environmental Protection Agency, *The U.S. Greenhouse Gas Inventory: In Brief*, EPA 430-F-02-008, April 2002, Washington, D.C. 20460

Appendix: Model Uses

NEMS Model

General Topics of Energy Related NEMS Studies

- Impacts of existing and proposed energy tax policies on the U.S. economy and energy system
- Impacts on energy prices, energy consumption, and electricity generation in response to carbon mitigation policies such as carbon fees, limits on carbon emissions, or permit trading systems
- Responses of the energy and economic systems to changes in world oil market conditions as a result of changing levels of foreign production and demand in the developing countries
- Impacts of new technologies on consumption and production patterns and emissions
- Effects of specific policies, such as mandatory appliance efficiency and building shell standards or renewable tax credits, on energy consumption
- Impacts of fuel-use restrictions, for example, required use of oxygenated and reformulated gasoline or mandated use of alternative-fueled vehicles, on emissions and energy supply and prices
- Impacts on the production and price of crude oil and natural gas resulting from improvements in exploration and production technologies
- Impacts on the price of coal resulting from improvements in productivity
- Numerous energy related studies for Congress or other federal agencies:
    - Energy Information Administration, *Measuring Changes in Energy Efficiency for the Annual Energy Outlook 2002*, (Washington, DC, 2002).
    - Energy Information Administration, *Analysis of Corporate Average Fuel Economy (CAFÉ) Standards for Light Trucks and Increased Alternative Fuel Use*, SR/OIAF/2002-05, (Washington, DC, March 2002).
    - Energy Information Administration, *Analysis of Efficiency Standards for Air Conditioners, Heat Pumps, and Other Products*, SR/OIAF/2002-01, (Washington, DC, February 2002).
    - Energy Information Administration, *Strategies for Reducing Multiple Emissions from Electric Power Plants With Advanced Technology Scenarios*, SR/OIAF/2001-05, (Washington, DC, October 2001).
    - Energy Information Administration, *Impact of Renewable Fuels Standard/ MTBE Provisions of S. 1766*, SR/OIAF/2002-06, (Washington, DC, March 2002).
    - Energy Information Administration, *Impact of Renewable Fuel Standard/ MTBE Provisions of S. 517*, SR/OIAF/2002-06 Addendum, (Washington, DC, April 2002).
    - Energy Information Administration, *Analysis of Strategies for Reducing Multiple Emissions from Power Plants: Sulfur Dioxide, Nitrogen Oxides, and Carbon Dioxide*, SR/OIAF2000-05, , (Washington, DC, December 2002).
    - Energy Information Administration, *Impacts of a 10-Percent Renewable Portfolio Standard, SR/OIAF/2002-03*,

(Washington, DC, February 2002).
    - Energy Information Administration, *Impacts of the Kyoto Protocol on U.S. Energy Markets & Economic Activity*, SR/OIAF/98-03, (Washington, DC, October 2002).
    - Energy Information Administration, *Reducing Emissions of Sulfur Dioxide, Nitrogen Oxides and Mercury from Electric Power Plants*, SR/OIAF/2001-04, (Washington, DC, September 2001).
- Carbon and Vehicle Emissions Modeling for Congress, EPA, and DOE
    - Energy Information Administration, *Impacts of the Kyoto Protocol on U.S. Energy Markets and Economic Activity*, Prepared for the U.S. House of Representatives Committee on Science, SR/OIAF/98-03, October 1998.
    - Energy Information Administration, *Service Report: Analysis of Carbon Stabilization Cases*, Prepared for the U.S. Dept. of Energy Office of Policy and International Affairs, SR-OIAF/97-01, October, 1997.
    - Energy Information Administration, *Analysis of the Impacts of an Early Start for Compliance with the Kyoto Protocol*, Prepared for the U.S. House of Representatives Committee on Science, SR/OIAF/99-02, July 1999.
    - Energy Information Administration, *Analysis of The Climate Change Technology Initiative*, Prepared for the U.S. House of Representatives Committee on Science, SR/OIAF/99-01, April 1999.
    - Energy Information Administration, *Analysis of The Climate Change Technology Initiative: Fiscal Year 2001*, Prepared for the U.S. House of Representatives Committee on Science, SR/OIAF/2000-01, April 2000.
    - Interlaboratory Working Group on Energy-Efficient and Low-Carbon Technologies, *Scenarios of U.S. Carbon Reductions: Potential Impacts of Energy Efficient and Low Carbon Technologies by 2010 and Beyond*, (Oak Ridge National Laboratory, Lawrence Berkeley National Laboratory, Pacific Northwest National Laboratory, National Renewable Energy Laboratory, and Argonne National Laboratory), September, 1997.
    - Interlaboratory Working Group on Energy-Efficient and Low-Carbon Technologies, *Scenarios for a Clean Energy Future*, (Oak Ridge National Laboratory, Lawrence Berkeley National Laboratory, Pacific Northwest National Laboratory, National Renewable Energy Laboratory, and Argonne National Laboratory), ORNL/CON-476 and LBNL-44029, November, 2000.
- Transportation Specific Model Runs for White House and Other Governmental Agencies
    - **Transportation Gasoline Tax Model Runs for the White House**, 1996. These model runs lead to the 3 cent tax on gasoline implemented by the Administration in 1996.
    - EIA, *The Impacts of Increased Diesel Penetration in the Transportation Sector*, Prepared by the Office of Integrated Analysis and Forecasting, August, 1998. These model runs and scenarios were developed for the Office of Transportation Technologies within the Dept. of Energy.
    - Request from EPA on travel and emissions associated with various Heavy-duty truck

emissions standards levels for criteria pollutants.

- Request from GAO to estimate future alternative fuels penetration levels
- Request from GAO to estimate alternative fuel vehicle sales and stocks effect of the Energy Policy Act (EPACT)
- EIA, *Analysis of Corporate Average Fuel Economy Standards for Light Trucks and Increased Alternative Fuel Use*, SR/OIAF/2002-05, March 2002. This service report assesses the impacts of more stringent corporate average fuel economy standards on energy supply, demand, and prices, macroeconomic variables where feasible, import dependence, and emissions. This study addresses the provisions of H.R. 4, S. 804, and S. 517 that pertain to light vehicle fuel economy in the transportation sector. A qualitative discussion is provided for the alternative fuels provisions included in S. 1766 and H.R. 4 **at the request of Senate Committee on Energy and Natural Resources.**
- EIA, *The Transition to Ultra-Low-Sulfur Diesel Fuel: Effects on Prices and Supply*, SR/OIAF/2001-01, May 2001. This study is an evaluation of EPA's Ultra-Low-Sulfur Diesel Fuel regulations for Heavy-duty Trucks **at the request of the House Science Committee.**
- NEMS vehicle travel equations were used to develop a DOT FHWA vehicle-miles traveled (VMT) model. **The proposed VMT model development was an inter-agency effort between EIA, EPA, and FHWA.**

Markal-Macro Model

http://www.etsap.org/annex5/main.html#3.1

Markal-Macro was used for a project on "Policies and Measures for Common Action" that was conducted by the Annex I Expert Group on the UN Framework Convention on Climate Change

As part of a study by the OECD Secretariat of the environmental implications of energy and transport subsidies, the Italian participant used an "elastic" version of MARKAL to evaluate the impact of removing financial subsidies from the electric sector in Italy. The many ways in which financial interventions affect the electric supply industry were searched out, and MARKAL was used to assess their effect on electric and energy system costs and CO2 emissions.

The Energy Technology Systems Analysis Programme (ETSAP) of the International Energy Agency continues to provide a multinational capability to determine the most cost-effective national choices to limit future emissions of greenhouse gases, using consistent methodology that offers a basis for international agreement on abatement measures. The basic MARKAL model continues to serve national interests, as illustrated by its use for a major national RDD&D appraisal in the UK, its use to help develop the national least-cost energy strategy in the USA, and its acceptance by a wider international community. Outside ETSAP, MARKAL was used in Taiwan and (in the form of MENSA) in Australia to inform the debate on response strategies under the UN Framework Convention on Climate Change.

With the cooperation of the participants from Italy, Japan, UK and USA, ETSAP contributed to the International Energy Agency study, "Electricity and the Environment." Detailed descriptions were provided of technologies available for electricity supply and demand in the short and medium term. The information included technical performance and engineering costs. Specific data were drawn from the MARKAL databases of the four cooperating countries.

Although a common set of runs among the ETSAP participants was delayed, four countries participated in CHALLENGE, a cooperative international project on energy and environment systems analysis. CHALLENGE consists of a

network of scientists from East and West European countries. The project is intended to facilitate international negotiations and cooperation by providing a scientific basis for decisions on response strategies to reduce environmental stresses and climate risks due to energy use.

During Annex V, some participating countries provided inputs to major international studies by the International Energy Agency, Organization for Economic Cooperation and Development, and the Annex I Expert Group on the UN Framework Convention on Climate Change.

ETSAP originated as an International Energy Agency program to help establish energy technology R&D priorities on the basis of the needs of all the IEA countries. A common methodology and comparable databases have been the touchstone of the program since its very beginnings. The standard MARKAL model has continued to be the focus of the group's analyses, and recurring efforts have been made to assure reasonable consistency in the national databases.

Mini-Cam Model

Edmonds, Wise, and MacCracken. 1994. *Advanced Energy Technologies and Climate Change: An Analysis Using the Global Change Assessment Model (GCAM)*. PNL-9798. Pacific Northwest Laboratory. Richmond, Wash.

Richels, R., and J. Edmonds. 1994. "The Economics of Stabilizing Atmospheric CO2 Concentrations." In *Energy Policy*. Forthcoming.

Edmonds, J.A., J.M. Reilly, R.H. Gardner, and A. Brenkert. 1986. "Uncertainty in Future Global Energy Use and Fossil Fuel CO2 Emissions 1975 to 2075." TR036, DO3/NBB-0081 Dist. Category UC-11. U.S. Department of Commerce. Springfield, Va.: National Technical Information Service.

Edmonds, J., and J. Reilly. 1985. *Global Energy: Assessing the Future*. New York: Oxford University Press.

GREET Model

The major applications of the GREET Model consists of the following:

1) Energy and GHG emission effects of fuel ethanol (for the State of IL, DOE, USDA, and EPA). Posted two reports from this effort to GREET website.
2) Energy and emission effects of natural gas-based transportation fuels for DOE. Posted a report to GREET website.
3) Well-to-wheels analysis of energy and GHG emissions of advanced vehicle technologies and transportation fuels for GM (the three volume report is posted at the GREET website.
4) Fuel-cycle energy and emission effects of the fuels petitioned to DOE under the Energy Policy Act for DOE.
5) Working with EPA to integrate GREET into EPA's next generation of motor vehicle emission model (called the MOVES).

TAFV Model

Publications

Leiby, Paul N. and Jonathan Rubin, 2003. Transitions in Light-Duty Vehicle Transportation: Alternative Fuel and Hybrid Vehicles and Learning, forthcoming in *Transportation Research Record*, March 30.

Paul N. Leiby, Jonathan Rubin, and David Bowman, 2002. "Efficacy of Policies to Promote New Vehicle Technologies: Alternative Fuel Vehicles and Hybrid Vehicles," *Proceedings of the 25th Annual IAEE International Conference*, June 26-29, Aberdeen, Scotland.

"Flexible Greenhouse Gas Emission Banking Systems," DRAFT, Final Technical Report, Integrated Assessment of Global Climate Change Research Program, Notice 98-15, Principal Investigators Jonathan Rubin (Margaret Chase Smith Center for Public Policy, University of Maine) and Paul Leiby (Energy Division, Oak Ridge National Laboratory), March 31, 2001(Report to DOE Office of Science).

"Effectiveness and Efficiency of Policies to Promote Alternative Fuel Vehicles," Paul Leiby and Jonathan Rubin, *Transportation Research Record*, Vol. 1750, pp. 84-91, 2001.

The Alternative Fuel Transition: Results from the TAFV Model of Alternative Fuel Use in Light-Duty Vehicles 19962010 (Final Report, TAFV Version 1), ORNL/TM2000/168, September 17, 2000, Paul Leiby and Jonathan Rubin.

"An Analysis of Alternative Fuel Credit Provisions of US Automotive Fuel Economy Standards," Jonathan Rubin and Paul Leiby, *Energy Policy*, 28(9):589- 602, (July, 2000).

"Sustainable Transportation: Analyzing the Transition to Alternative Fuel Vehicles," Paul Leiby and Jonathan Rubin, *Transportation Research Board Circular, Transportation, Energy, and Environment*, No. 492:54-82, August 1999

"A Dynamic Analysis of Achievable Potential and Costs for Alternative Fuel Vehicles," Paul Leiby, Oak Ridge National Laboratory. Invited presentation at the International Energy Agency, International Workshop on Technologies to Reduce Greenhouse Gas Emissions: Engineering-Economic Analyses of Conserved Energy and Carbon, 57 May 1999, Washington DC, USA

Leiby, Paul and Jonathan Rubin, 1997 "The Transitional Alternative Fuels and Vehicles Model," Paul Leiby and Jonathan Rubin. *Transportation Research Record,* 1587:1018.

Example Reports available online: (http://pzl1.ed.ornl.gov/altfuels.htm)
* Effectiveness and Efficiency of Policies to Promote Alternative Fuel Vehicles, Paul Leiby and Jonathan Rubin, November 17, 2000 (Revised). Presented to the Transportation Research Board, 80th Annual Conference, January 7-11, 2001. Forthcoming in Transportation Research Record.
* The Alternative Fuel Transition: Results from the TAFV Model of Alternative Fuel Use in Light-Duty Vehicles 1996-2010 (Final Report, TAFV Version 1), September 17, 2000, Paul Leiby and Jonathan Rubin (Adobe Acrobat format, 387K).
* Analyzing the Transition to Alternative Fuel Vehicles, Project Progress Briefing, December 16, 1998, Paul Leiby and Jonathan Rubin.
* The Production of Alternative Fuel Vehicles for CAFE Credits, Jonathan Rubin and Paul Leiby, Presentation at the Transportation Research Board Workshop on Air Quality Impacts of Conventional and Alternative Fuel Vehicles, A1F03/A1F06 Joint Summer Meeting, Ann Arbor, MI, August 2-4, 1998.
* Analyzing the Transition to Alternative Fuel Vehicles, Paul Leiby and Jonathan Rubin Presentation to the Society of Automotive Engineers, 1998 SAE Government/Industry Meeting, Washington, D.C., April 20-22, 1998.
* The Transitional Alternative Fuels and Vehicles Model, 1997, (Transportation Research Record 1587) Paul Leiby and Jonathan Rubin
* Sustainable Transportation: Analyzing the Transition to Alternative Fuels, March 1999, (Asilomar Conference, August 1997) Paul Leiby and Jonathan Rubin (Adobe Acrobat format, 236K).

ISSUES IN DEVELOPING A TRANSPORTATION INFRASTRUCTURE INDEX

Herman O. Stekler
The George Washington University
Department of Economics

Brian W. Sloboda
U.S. Department of Transportation
Bureau of Transportation Statistics

The Bureau of Transportation Statistics of the U.S. Department of Transportation has recently undertaken several initiatives that involve monitoring and estimating economic activity in the US transportation industry. The procedure developed by Ord, Young, and Crutcher (2001) for monitoring developments in the industry involved forecasting key transportation indicators and then testing whether the outcomes reflected in the actual data were significantly different from the predictions. If the differences were significantly different, this was considered a signal (of a problem) that warranted further analysis.

In order to estimate the monthly level of economic activity in the transportation services sector of the industry (also known as the transportation services index (TSOI)), an index of output of that sector was developed Lahiri et al (2004). This index was estimated for the period 1980-2002. BTS is now in the process of assessing this index in preparation of using it to estimate and forecast the output of freight and passenger services on a monthly basis. Additionally, Lahiri et al (2003) highlight the close nonlinear relationship between freight transportation, inventory cycles and industrial production. Their empirical tests based on Granger-causality also confirm strong feedback relationships between transportation output, input inventories, and alternative measures of aggregate economic activity.

Besides the transportation services sector, the transportation industry also contains two other sectors: transportation equipment and transportation infrastructure. The Federal Reserve Board=s Index of Industrial Production already estimates the output of transportation equipment. Consequently, BTS can use that information and does not need to devote any further efforts to estimating activity for the transportation equipment sector. However, there is no existing source that currently estimates and forecasts the level of economic activity in the infrastructure sector of the transportation sector. This paper surveys the problems and the general methodology that are involved in estimating and eventually forecasting economic activity in the transportation infrastructure sector.

I. Definition of the Infrastructure Sector

The transportation infrastructure consists of: public roads, bridges, airports, railroad tracks, urban transit tracks, and ocean and inland water-ports. An investment in transportation infrastructure occurs whenever there is activity involved in constructing or repairing any of these facilities. Within the North American Industry Classification System (NAICS), the construction of transportation infrastructure was classified within NAICS code 237 (Heavy Construction).

2002 NAICS Classification System for Transportation Infrastructure

NAICS code	Description
237	Heavy and Civil Engineering Construction
2373	Highway, street, and bridge construction
23710	Highway, street, and bridge construction[1]
2371	Highway, Bridge, and Street Construction
2379	Other heavy and civil engineering construction
237990	Other heavy and civil engineering construction[2]
23712	Oil and Gas Pipeline and Related Structures Construction
23799	Other heavy and civil engineering construction (includes tunnel construction)

Unfortunately, the data that are available are not collected on a NAICS basis. We now turn to the data sources to determine which data would be available to construct an index that measures economic activity in the transportation infrastructure sector.

II. Data Sources

The data that measure spending on transportation infrastructure are contained in the time series that measure construction activity, and there are a number of such data series. The Census Bureau publishes three such series: Value of Construction Put in Place (VIP), the Census of Construction, and the Annual Capital Expenditures Survey (ACES). Only the VIP series is published monthly, but the other data might be useful for our purposes if they were to enable us to benchmark those categories of transportation infrastructure expenditures that are not available on a monthly basis.

VIP is a measure of the value of construction installed or erected at a site during a given month. VIP is the sum of the value of work done on construction projects underway during the month regardless of the actual start date of the project or when payment is made to the contractors. However, VIP excludes expenditures such as the value of maintenance and repairs. Thus for example, repairing roads, which involves the transportation infrastructure yields an underestimate of economic activity in that sector.[3] More specifically, VIP contains the following construction expenditures:

- New buildings and structures
- Additions, alterations, major replacements etc to existing buildings and

1 This level of detail includes airport runway construction and airport runway line painting.
2 This level of details contains railroad construction i.e., interlocker, roadbed, signal; railway roadbed construction; light-rail construction, port facility construction; subway construction; and harbor construction.
[3]However, if repairs are a relatively constant percentage of total outlays, this need not be a problem.

structures4

- Installed mechanical and electrical equipment
- Installed industrial equipment
- Site preparation and outside construction such as streets, sidewalks, parking lots etc.
- Cost of labor and materials (including owners supplied)
- Cost of construction equipment rental
- Profit and overhead costs
- Costs of architectural and engineering work
- Any miscellaneous costs related to the project

VIP is disaggregated into two major categories: private and public construction. These categories are then further disaggregated into the various types of construction. Table 1 presents the various categories of construction that are contained in the VIP series. Table 1 also shows where the information about the transportation infrastructure is contained within the VIP series. Information about railroads, pipelines, and highway and streets is reported directly, but data on the other portions of the infrastructure are contained within the miscellaneous and all other categories. This means that these data will have to be obtained separately from the Census Bureau or will have to be estimated. This still remains to be addressed.

The Census will not be of assistance in constructing our measure of economic activity involving the transportation infrastructure because it does not report on the categories of construction-put-in-place but rather measures the characteristics of establishments performing construction work. On the other hand, the ACES might provide some information about private infrastructure spending since ACES collects data on fixed assets and depreciation, sales and receipts, and capital expenditures for new and used structures and equipment. Additionally, ACES provides detailed statistics on actual business spending by domestic, private, non-farm businesses operating in the US.

Table 1
Components of the Value of Construction Put in Place

Type of Construction	Description of its Relation to Transportation Infrastructure
Private Construction	
Residential Buildings	
New Housing Units	
1 unit	
2 units or more	
Improvements	
Nonresidential Buildings	

4 VIP excludes several types of expenditures such as the value of maintenance and repairs to existing structures and land acquisition.

Industrial	
Office	
Hotels, motels	
Other commercial	
Religious	
Educational	
Hospital and Institutional	
Miscellaneous	This category includes bus and airline terminals
Farm, nonresidential	
Public Utilities	
Telecommunications	
Other Public Utilities	
Railroads	This will also count as transportation infrastructure
Electric light and power	
Gas	
Petroleum pipelines	This will also count as transportation infrastructure
All other private	Includes privately owned streets and bridges, airfields, and other forms of transportation infrastructure
Public Construction	
Buildings	
Housing and redevelopment	
Industrial	
Educational	
Hospital	
Other	This category contains estimates on passenger terminals
Highways and Streets	This will also count as transportation infrastructure
Military facilities	
Conservation and development	
Sewer systems	
Water supply facilities	
Miscellaneous	Includes airfields, transit systems, airfields and other modes for transportation infrastructure

III. General Methodology

Once we have obtained the data for all of the categories, the remainder of the project involves the following steps:

A. Seasonally adjust all of the time series
B. Calculate an index (C_i) that measures the economic activity of that category
C. Construct a weighted index that measures economic activity across all categories. We would use chained value-added weights.

The subsequent discussion focuses on some details for the aforementioned steps to prepare the Transportation Infrastructure Index (TII). The first step in constructing the index, the data will be adjusted accordingly and the data will be seasonally adjusted using the X-11 program. [5]

The total output for transportation infrastructure is an aggregate of real output generated by each of the categories of transportation infrastructure. The data from the categories were used to construct the Transportation Infrastructure Index (TII). Each of these categories represents the output quantity for each category of transportation infrastructure. Therefore, each of these categories was converted into index number form with 1996 =100. In order to construct an index for the transportation infrastructure sector, assigning weights to each of the categories combined the indices of each of the categories. Also these weights measure the relative importance in the base year, 1996, of each transportation infrastructure component to the overhaul transportation infrastructure sector. Thus, the index for transportation infrastructure could be aggregated using the formulization of the linked-Laspeyres quantity index.

The use of fixed-weighted measures of quantity index, such as the Laspeyres quantity index may result in a "substitution bias" which results in an overstatement of output growth for periods after the base year and an understatement of growth for periods before the base year Landefeld and Parker (1995). The tendency of "substitution bias" reflects the fact that those commodities for which output grows rapidly tend to be those for which prices change less proportionately. Although this bias may be small enough to be safely ignored for shorter sample periods, the output measures derived from a fixed-weighted index can become increasingly subject to "weighting effects" as the time between weighting period and the current period lengthens. A similar but opposite problem occurs with another type of fixed-weighted index, the Paasche quantity index, which uses current period prices as weights.

In applied economics, the method to rectify the problem of the Laspeyres and Paasche indices is the Fisher Ideal Quantity Index, and this index is the geometric mean of the Laspeyres and Paassche indices. Put in another way, the Fisher index registers changes that fall between those from Laspeyres and Paasche indices. Also this is known as a chain index. In fact, the Bureau of Economic Analysis has been publishing the National Income and Product Accounts (NIPAs) estimates using this approach since the 1990's. Moreover, the Board of Governors of Federal Reserve Board has also adopted the Fisher-ideal formula in constructing the Industrial Production Index since 1996. Conceptually, the transportation infrastructure index measure is similar to the transportation services output index and both of these indices are conceptually similar to the Industrial Production Index as produced by the Federal Reserve.

5 If the data series were measured in real quantities, no price deflation would be applied.

References

Executive Office of the President, Office of Management and Budget. (2002). North American Industry Classification System, United States, 2002, Springfield, Virginia: National Technical Information Service.

Lahiri, K., Stekler, H.O., Yao, W. and Young, P. (2004) "Monthly Output Index for the US Transportation Sector," *Journal of Transportation and Statistic*s, Forthcoming.

Lahiri, K. and Yao, W. (2003) Transportation Activity and Economic Growth: the Causal Link, paper prepared for presentation at the Transportation Research Board (TRB) 83 rd Annual Meeting in January 2004, Washington D.C., USA.

Landefeld, J.S. and Parker, R.P.(1995). "Preview of the Comprehensive Revision of the National Income and Product Accounts: BEA's New Featured Measures of Output and Prices," *Survey of Current Busines*s.

Ord, K., P. Young, and B. Crutcher. (2001). "Creating a Monthly Monitoring System for Transportation Indicators," a paper presented at the International Symposium on Forecasting, Pine Mountain, Georgia, June 2001.

Business Cycle Analysis for the U.S. Transportation Sector

Kajal Lahiri and Wenxiong Yao, University at Albany-SUNY

Peg Young, Bureau of Transportation Statistics

1. INTRODUCTION

Most of the developed economies have increasingly become more service intensive in the postwar period. For instance, in the U.S. during 1953:I-2003:II, the share of goods in the GDP has declined from 54% to 35%, compared to an increase in the share of services from 34% to 56%. The relative change in shares of employees providing these two products in total non-farm employment is even wider. The information from the service sector has also become essential in understanding fluctuations in a contemporary economy. Moore (1987) points out that the ability of the service sectors to create jobs has differentiated business cycles since 1980s from their earlier counterparts, and has made economy-wide recessions to be shorter and less severe. Layton and Moore (1989) argue that two factors can account for less severity in service sector recessions. One is the increase in importance of non-manufacturing (non-Mfn) labor market relative to that of the manufacturing sector. The other is the non-storability of services and thus inventories. Since inventory movement is the dominant feature of business cycles, we can appreciate why service sectors were not paid much attention in the past for business cycle analysis. This could also be one of the reasons for the absence of service sector indicators in NBER Committee's deliberations in dating business cycles.

However, transportation services sector is different. Besides its sizable part in the U.S. economy,[1] transportation plays a crucial role in facilitating economic activity between sectors and across regions like the flow of blood in a human body. NBER scholars, from the very beginning of their study of business cycles, had noticed the pervasive influence of transportation on all aspects of economy, and paid adequate attention in observing the recurrent feature of business cycles from the perspective of transportation.[2] In addition, transportation equipment and infrastructure have been one of the major contributors to both total investment and corporate bond issuances from the beginning till today. Using this argument, Dixon (1924) proposed that regulation of the railways be part of stabilization policies. Unfortunately, further efforts to study the role of transportation in monitoring modern business cycles were hindered largely due to the discontinuation, in the 1960's, of many of the monthly transportation indicators used by early NBER scholars.

Lahiri and Yao (2003a) provide a schematic illustration of the stage-of-fabrication production process employed by a typical firm to transform input inventories (purchased materials-supplies and work-in-progress) into output inventories (finished goods), as depicted in Figure 1. The middle and lower parts illustrate that freight transportation is closely related with input inventories in the overall economy, which account for 65% of the total manufacturing (Mfn) inventories by its value and 67% by variance (Allen, 1995; Blinder and Maccini, 1991). This diagram clearly shows that transportation sector is connected with different stages of fabrication in the aggregate economy. In particular, for-hire freight transportation is closely related to inventory cycles, which is considered as the dominant feature of economic fluctuations in GDP since Abramovitz (1950). For a more recent study on inventories, see Humphreys *et al.* (2001).

These arguments motivated us to study the classical and growth cycles characteristics of this sector with economic indicator analysis (EIA) approach. This study can also be considered as an extension of the work by Layton and Moore (1989) on the general service sector, and also as a continuation of NBER scholars' pioneering work many decades ago.

The remaining text is structured into four sections after the Introduction. Section 2 studies the current state of transportation sector through the transportation composite coincident index (CCI) using with both NBER non-parametric method and parametric dynamic factor models. Section 3 constructs a transportation composite leading index (CLI) to predict its CCI. In both sections, selection of indicators is done with rigorous statistical tests besides the standard EIA criteria. The last section summaries the conclusion of the paper.

[1] Using different concepts about the scope of the transportation industry would yield different measures of its importance, varying anywhere from 3.09% (Transportation GDP) to 16.50% (Transportation-driven GDP).

[2] Burns and Mitchell (1946, p. 373) and Hultgren (1948) found that the cyclical movements in railroads coincided with the prosperities and depressions of the economy. Moore (1961, volume I, pp. 48-50), based on updated data through 1958, found that railway freight car loading, while still being coincident at troughs, showed longer leads at peaks after the 1937-1938 recession.

2. INDEX OF COINCIDENT INDICATORS FOR TRANSPORTATION

2.1. Comovement among Four Coincident Indicators

Burns and Mitchell's (1946) definition of business cycles has two key features. The first is the comovement or concurrence among individual economic indicators; the other is that business cycle is governed by a switching process between different regimes or phases. Extracting the comovement among coincident indicators leads to the creation of CCI, which is the basis to define the current state of the aggregate economy or of a single sector.

Following the NBER tradition and Layton and Moore (1989), we use four conventional coincident indicators to define the current state of U.S. transportation sector. They are: transportation services output index or TSOI (Y_{1t}) – a newly constructed monthly series developed in a research project sponsored by U.S. Bureau of Transportation Statistics,[3] real aggregate payrolls (Y_{2t}), real personal consumption expenditure (Y_{3t}), and employment (Y_{4t}) of this sector. These indicators, plotted in Figure 1, reflect information on output, income, sales, and labor usage in the transportation sector. Given these four available data series, the existence of comovement among them should be tested for their statistical significance. That is, how well are they synchronized with each other in terms of directional change? This topic has been the subject of considerable research in recent years because the economic costs associated with forecast errors during business cycle turning points and other times are considerably different; see Pesaran and Timmermann (2003).

This concept of comovement can be illustrated with four outcomes adapted from Granger and Pesaran (2000). With a similar contingency table, various χ^2 tests were designed mainly based on the value of P_1+P_2 from the main diagonal to test the statistical relevance between two events or series, see Henriksson and Merton (1981), Schnader and Stekler (1990), and Pesaran and Timmermann (1994) for further discussions. Using the information, Harding and Pagan (2002) propose an index of concordance for two series x_t and y_t with sample size T:

$$\hat{I} = \frac{1}{T}\{\sum_{t=1}^{T} S_{xt}S_{yt} + \sum_{t=1}^{T}(1-S_{xt})(1-S_{yt})\}. \quad (1)$$

S_{xt} and S_{yt} are the underlying states (0 or 1) of each series based on turning points defined using the NBER procedures. The degree of concordance defined in (1) between two variables is quantified by the fraction of time that both series are simultaneously in the same state of expansion ($S_t = 1$) or contraction ($S_t = 0$). In other words, they measure the ratio of sum of Hits and Correct Rejections relative to the sample size (T) such that the value $\hat{I}$ ranges between 0 and 1.

The NBER procedures of dating turning points were formalized and documented in Bry and Boschan (1971), namely, the BB algorithm. In practice, the BB algorithm is supplemented by censoring procedures to distinguish the real peaks and troughs from spurious ones, *e.g.*, a movement from a peak to a trough (phase) cannot be shorter than six months and a complete cycle must be at least fifteen months long. The resulting turning points define the "specific cycle" of each component series. Likewise, a "reference cycle" can be defined based on the CCI (namely, NBER index), which is constructed from four coincident indicators using the Conference Board methodology (2001). They are listed in Table 1.

The synchronization of cycles among coincident indicators can be measured and tested based on the index of concordance between four specific cycles and the reference cycle. All the pairs of transportation coincident indicators have positive correlations ranging between 0.5 ~ 0.7 and concordance indexes between 0.8 ~ 0.9. With the reference cycle, both numbers are even higher. These statistics suggest strong evidence of synchronization of cycles between them. None of the series is dominated by either of the states. Hence the high concordance indexes are significantly associated with the high correlations. Statistics can also be developed to test if synchronization of cycles is significant between indicators and the reference cycle. A simple way to do so is the t-test for $\boldsymbol{H_0}$: $\rho_S = 0$. $\hat{\rho}_S$ is obtained from the regression

$$\frac{S_{yt}}{\hat{\sigma}_{S_y}} = a_1 + \rho_S \frac{S_{xt}}{\hat{\sigma}_{S_x}} + u_t. \quad (2)$$

Standard t-statistics is based on OLS regression. We use Newey-West heteroskedasticity and autocorrelation consistent standard errors and covariance to account for serial correlation. All these statistics significantly reject H_0, producing strong evidence for the existence of a common cycle among four transportation coincident indicators. Thus they are qualified coincident indicator for this sector.

2.2 Transportation CCIs

The NBER CCI is constructed non-parametrically by assigning fixed standardization factors as weights to each of the components. An alternative would be using techniques of modern time-series analysis to develop dynamic factor models with regime switching (Kim-Nelson) or without regime switching (Stock-Watson).

[3] For details on construction of this index and discussion on its characteristics, see Lahiri *et al.* (2004) and Lahiri and Yao (2003a).

The resulting single indexes would represent the underlying state of its constituent time series. Thus dating turning points could be based on the probabilities of the recessionary regime implied by the model.

Given a set of coincident indicators Y_{it}, their growth rates can be explained by an unobserved common factor ΔC_t, interpreted as growth in CCI, and some idiosyncratic dynamics. This defines the measurement equation for each component:

$$\Delta Y_{it} = \gamma_i \Delta C_t + e_{it}, \quad (3)$$

where ΔY_{it} is logged first difference in Y_{it}. In the state-space representation, ΔC_t itself is to be estimated. In the transition equations, both the index ΔC_t and e_{it} are processes with AR representations driven by noise terms w_t and ε_{it} respectively.

$$\Phi(L)\,(\Delta C_t - \mu_{st} - \delta) = w_t, \quad (4)$$

$$\Psi(L)\, e_{it} = \varepsilon_{it}. \quad (5)$$

These two noise terms are assumed to be independent of each other. The transitions of different regimes (μ_{st}), incorporated in (2), are governed by a Markov process:

$$\mu_{st} = \mu_0 + \mu_1 S_t,\ S_t = \{0, 1\},\ \mu_1 > 0, \quad (6)$$

$$Prob\,(S_t=1|S_{t-1}=1)=p,\ Prob(S_t=0|S_{t-1}=0)=q, \quad (7)$$

Equations (3) ~ (5) define the Stock-Watson model while the Kim-Nelson model includes all five equations. To implement the Kim-Nelson model, we used priors from the estimated Stock-Watson model. Priors for regime switching parameters were obtained from sample information of the NBER index. Both models were estimated using computer routines described in Kim and Nelson (1998). Unlike the Stock-Watson (1991) model specification for the aggregate economy, personal consumption expenditure and employment in transportation appear to be somewhat lagging to the current state of transportation.

The final specification and parameter estimates from Stock-Watson and Kim-Nelson models are reported in Table 2. The two sets of estimates are close except that the sum of the AR coefficients for the state variable in the Stock-Watson model is significantly higher, implying more state dependence in the resulting index. This difference is complemented by a much larger role that employment plays in the Kim-Nelson model. The latter model also distinguishes between two clear-cut regimes of positive and negative growth rates. Based on transitional probabilities (P_{00} and P_{11}), expected durations of recessions and expansions are calculated as $(1- P_{00})^{-1}$ and $(1- P_{11})^{-1}$ respectively. This would give us 13.5 and 66.7 months on average of recessions and expansion in the transportation sector in compared to the actual durations of 13 and 68 months.

The estimated transportation CCIs from these two models are plotted against the NBER index in Figure 2. Compared to Kim-Nelson, the Stock-Watson index agrees more closely with the NBER index throughout the period. The NBER index picks up some of the details of the cyclical movements better than the two alternative indexes (e.g., the delineation between 1980 and 1981 recessions). Despite differences in their model formulations and in minor details, their cyclical movements appear to be very similar to one another and synchronized well with the NBER-defined recessions for the economy (the shaded areas).

The turning points of NBER CCI were identified in Table 1. Together with specific cycles of four individual indicators, the chronology of business cycles in the U.S. transportation sector is defined for the period since January 1979. There are clearly four major recessions: 3/79–8/80, 1/81–2/83, 5/90–6/91, and 11/00-12/01. Overall, there is one-to-one correspondence between business cycles of the transportation sector and those of overall economy. The comparison between these two is reported in Table 3. Transportation cycles have a slight lead at peaks of the economy-wide business cycles while being roughly coincident at troughs, *i.e.*, the duration of transportation recessions is slightly longer. Interestingly, these findings are very similar to those in Moore (1961), who used only railway freight data for his conclusion. More discussions on the economic theory behind these results are given in Lahiri *et al.* (2004).

3. INDEX OF LEADING INDICATORS FOR TRANSPORTATION

Based on seven selected leading indicators, a leading index was constructed using the conventional NBER approach (see Lahiri *et al*., 2003b for details). Standardization factors of leading indicators used for constructing a NBER index are the inverse of the standard deviation of each series, as reported in Table 3. Following the Conference Board (2001), the constructed transportation CLI is a weighted average of their transformed symmetric month-to-month change then converted back to a level index. It is plotted in Figure 3.

The exact lead-lag relation of the transportation CLI relative to transportation business cycle chronologies is also reported in Table 4. For the latest recession that started in November 2000, the leading index led the transportation coincident index by 20 months. As the trough of transportation sector has been determined in December 2001, the CLI has clearly reached its trough three monthly earlier. Overall, the leading index of U.S. transportation sector leads its CCI by 10 months at peaks and 6 months at troughs on the average. The CLI also gives two short false signals in 2/95-2/96 and 5/98-7/98. However, these extra turns are very short and mild and could be easily ignored using the censoring rule in the BB algorithm. The extra turn in 1995 is associated with a growth cycle recession instead of full-fledged recession in transportation

sector; see Lahiri *et al.* (2004). The other one might be caused by a sector-wide temporary shock, as seen in most of transportation indicators.

We should, however, point out that the lead-time analysis presented above does not take into account either the lag involved in obtaining the data necessary to construct the series or the necessity of employing a non-parametric filter rule that by its very nature involves a delay in identifying a turn. After all, a leading indicator is only as good as the filter rule (*e.g.,* three consecutive decline rule for signaling a downturn) that interprets its movements. These rules typically involve trade-offs of accuracy for timeliness and miss signals for false alarms, see Lahiri and Wang (1994).

4. CONCLUSION

This paper studies both the classical business cycles and growth cycles of the U.S. transportation sector since 1979. Based on four coincident indicators are selected to measure labor inputs, production, income, and spending in this sector, CCI was created using both NBER non-parametric method and parametric model estimation methods developed in the past two decades. Recessions in the transportation have a one-one-one correspondence with those in the aggregate economy. Then seven leading indicators were selected from an initial list to 20 variables. The constructed transportation CLI works well in predicting the future cycles of this sector. Our study suggests that transportation, as an important service sector, has both its unique business cycle characteristics, and some features that are common to the general service sectors of the economy.

References

Abramovitz, M. (1950), *Inventories and Business Cycles, with special reference to manufacturers' inventories*. New York: NBER.

Blinder, A.S., and Maccini, L.J. (1991), "Taking Stock: A Critical Assessment of Recent Research on Inventories," *Journal of Economic Perspectives*, Vol. 5, No. 1, Winter 1991, pp. 73 – 96.

Bry, G., and Boschan, C. (1971), "Cyclical Analysis of Time Series: Selected Procedures and Computer Programs," *NBER Technical Paper 20*.

Burns, A.F., and Mitchell, W.C. (1946), *Measuring Business Cycles*, New York: National Bureau of Economic Research.

Conference Board (2001), Calculating the Composite Indexes, www.conference-board.org, revised 01/01.

Dixon, F.H. (1924), "Transportation and Business Cycle," in Edie, L.D. (ed.), *The Stabilization of Business*, New York: The Macmillan Company, pp. 113-163.

Granger, C.W.J., and Pesaran, M.H. (2000), "Economic and Statistical Measures of Forecast Accuracy," *Journal of Forecasting*, Vol. 19, pp. 537-560.

Harding, D, and Pagan, A. (2002), "Dissecting the cycle: a methodological investigation," *Journal of Monetary Economics*, 49, pp. 365-381.

Hultgren, T. (1948), *American Transportation in Prosperity ad Depression*, New York: National Bureau of Economic Research.

Humphreys, B.R., Maccini, L.J., and Schuh, S. (2001), "Input and Output Inventories," *Journal of Monetary Economics*, Vol. 47, pp. 347-375.

Henriksson, R.D., and Merton, R.C. (1981), "On Market Timing and Investment Performance 2: Statistical Procedures for Evaluating Forecast Skills," *Journal of Business*, Vol. 54, pp. 513-533.

Kim, C.J., and Nelson, C.R. (1998), "Business Cycle Turning Points, A New Coincident Index, and Tests of Duration Dependence Based on A Dynamic Factor Model with Regime Switching," *The Review of Economics and Statistics*, Vol. 80, No. 2, pp. 188-201.

Lahiri, K., Stekler, H.O., Yao, W., and Young, P. (2004), "Monthly Output Index for the US Transportation Sector," Forthcoming, *Journal of Transportation and Statistics*.

Lahiri, K., and Yao, W. (2003a), "Transportation Activity and Economic Growth: the Causal Link," mimeo, University at Albany-SUNY, 2003.

Lahiri, K., Yao, W., and Young, P. (2003b), "Business Cycle Analysis for the U.S. Transportation Sector," paper presented at the conference "Business Cycles – Theory, History, Indicators, and Forecasting" in Honor of Victor Zarnowitz, organized by CIRET and RWI, Essen, Germany, June 27-28, 2003.

Lahiri, K., and Wang, Z. (1994), "Predicting Cyclical Turning Points with Leading Index in a Markov Switching Model," *Journal of Forecasting*, Vol. 13, pp. 245-263.

Layton, A., and Moore, G.H. (1989), "Leading Indicators for the Service Sector," *Journal of Business and Economic Statistics*, Vol. 7, No. 3, July 1989, pp. 379-386.

Moore, G.H. (1987), "The Service Industries and the Business Cycle," *Business Economics*, Vol. 22, pp. 12-17.

Moore, G.H. (1961), *Business Cycle Indicators,* Volume I and Volume II. Princeton, New Jersey: Princeton University Press for NBER.

Pesaran, M.H., and Timmermann, A.G. (2003), "How Costly Is It to Ignore Breaks When Forecasting The Direction of A Time Series?" *CESINO Working Paper No. 875*, February 2003.

Pesaran, M.H., and Timmermann, A.G. (1994), "A Generalization of Nonparametric Henriksson-

Merton Test of Market Timing," *Economic Letters*, Vol. 44, pp. 1-7.

Schnader, M.H., and Stekler, H.O. (1990), "Evaluating Predictions of Change," *Journal of Business*, Vol. 63, pp. 99-107.

Stock, J.H. and Waston, M.W. (1989), "New Indexes of the Coincident and Leading Economic Indicators," *NBER Macroeconomics Annual*.

Stock, J.H., and Waston, M.W. (1991), "A Probability Model of the Coincident Economic Indicators," in Lahiri, K. and Moore, G.H. (Eds), *Leading Economic Indicators*: *New Approaches and Forecasting Records*. (Cambridge University Press, Cambridge), pp. 63-89.

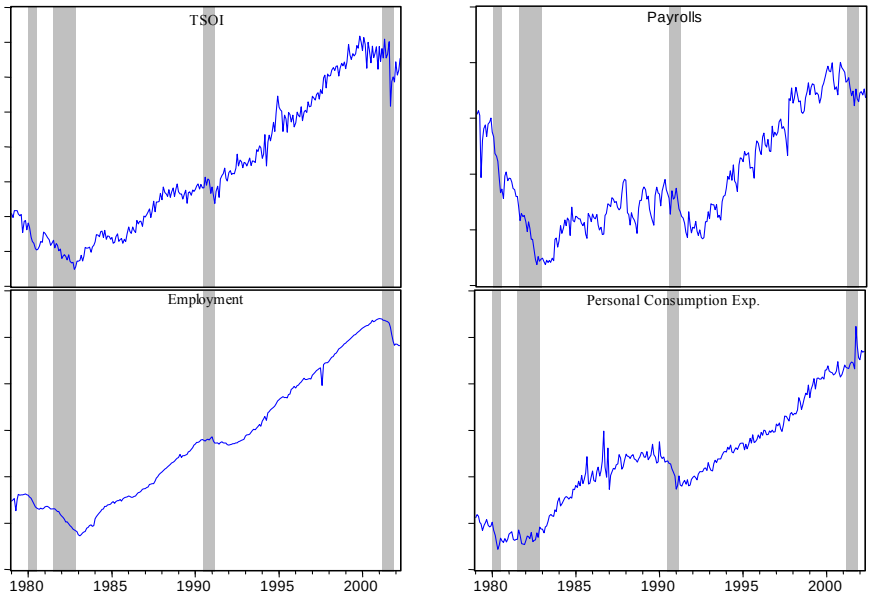


Figure 1. Coincident Indicators for the U.S Transportation Sector
*Shaded areas represent NBER-defined recessions for the U.S. economy.

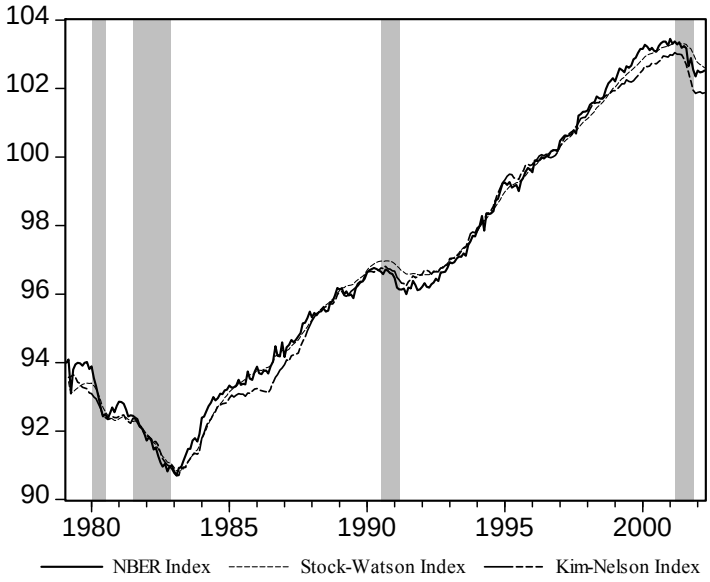


Figure 2. Coincident Indexes of the U.S. Transportation Sector
*Shaded areas represent NBER-defined recessions for the U.S. economy.

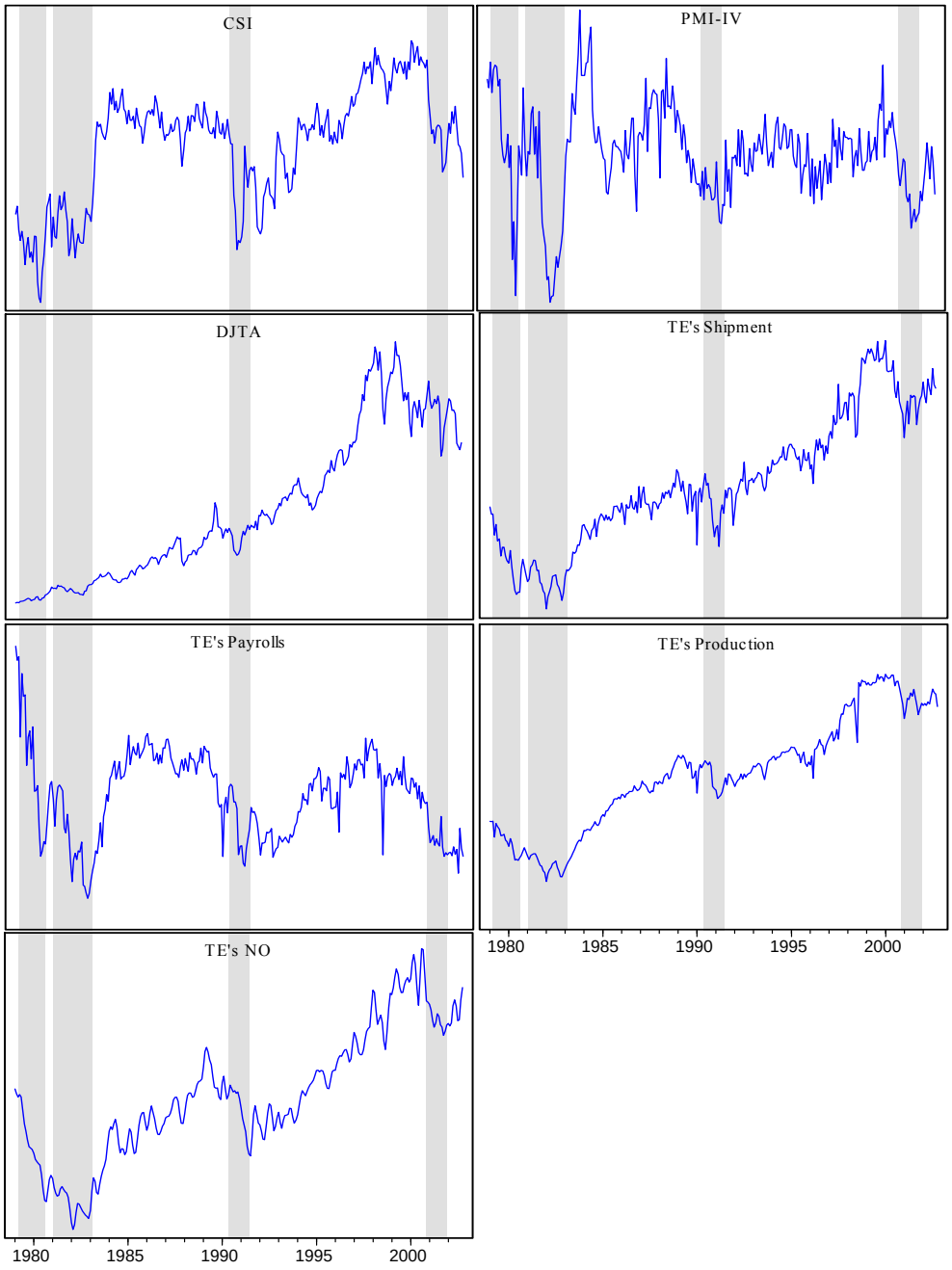


Figure 3. Leading Indicators for the U.S. Transportation Sector
* Shaded areas represent recessions defined for the U.S. transportation sector.

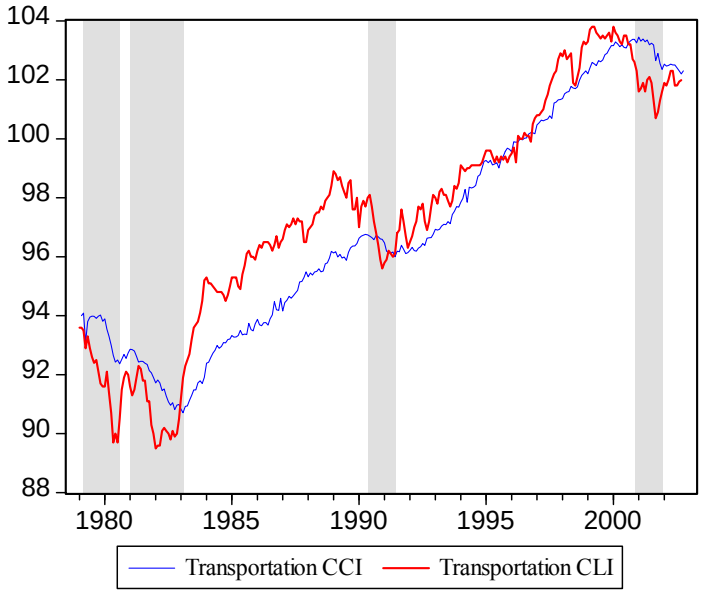


Figure 4. CLI for the U.S. Transportation Sector
* Shaded areas represent recessions defined for the U.S. transportation sector.

Table 1. Business Cycle Chronologies in U.S. Transportation Sector, 1979 – 2002

Transportation Reference Cycles		Leads (-) and Lags (+), in months, relative to Transportation reference cycle									
		NBER Index		Output		Employment		Real PCE		Real Pay	
P	T	P	T	P	T	P	T	P	T	P	T
03/79	08/80	0	-1	0	-1	3	+1	0	-3	0	0
01/81	02/83	0	0	-1	-4	+2	0	0	-9	-3	0
05/90	06/91	-3	+3	+3	-3	+8	+7	-18	+5	-1	+1
11/00	-	0		-12	-	+2	-	-12	-	-13	-
Mean		-1	+1	-3	-3	+4	+3	-8	-2	-4	0
Median		0	0	0	-3	3	1	-6	-3	-2	0
Std Dev.		1.5	2.1	2.1	1.5	2.9	3.8	9.0	7.0	6.0	0.6
Extra Turns				06/84	09/85					09/84	08/85
				12/88	07/89					11/87	08/88
				12/94	07/95					01/95	08/95

Table 2. Estimates of the Transportation Coincident Index Models

Variables	Parameters	Stock-Watson Model		Kim-Nelson Model			
				Prior	Posterior		
		Estimate	s.e.		Mean	s.e.	Median
ΔC_t	Φ_1	0.775	0.167	0.775	0.127	0.119	0.114
(State Variable)	Φ_2	0.107	0.162	0.107	0.121	0.085	0.124
ΔY_{1t}	γ_1	0.171	0.057	0.1	0.136	0.028	0.136
(Output)	φ_{11}	-0.519	0.067	-0.2	-0.637	0.057	-0.638
	φ_{12}	-0.067	0.017	0	-0.401	0.057	-0.401
	σ_1^2	5.181	0.480	2	0.652	0.057	0.648
ΔY_{2t}	γ_2	0.148	0.048	0.1	0.173	0.042	0.172
(Payrolls)	φ_{21}	-0.162	0.077	-0.1	-0.216	0.061	-0.216
	σ_2^2	2.107	0.210	2	0.782	0.071	0.778
ΔY_{3t}	γ_3	1.485	0.631	1.5	0.059	0.060	0.059
(Personal	γ_{31}	-1.364	0.626	-1.4	-0.041	0.059	-0.039
Consumption	φ_{31}	-0.149	0.122	-0.1	-0.388	0.060	-0.388
Exp.)	σ_3^2	2.443	1.831	2	0.849	0.076	0.844
ΔY_{4t}	γ_4	0.110	0.021	0.1	0.548	0.081	0.557
(Employment)	φ_{41}	-0.006	0.357	-0.1	-0.025	0.084	-0.026
	σ_4^2	0.072	0.015	2	0.125	0.081	0.120
	P_{00}			0.967	0.926	0.066	0.945
	P_{11}			0.986	0.985	0.012	0.988
	μ_0			-0.869	-1.822	0.554	-1.727
	μ_1			0.745	2.208	0.580	2.110
	δ			-	0.356	0.038	0.359
	$\mu_0 + \mu_1$			-	0.385	0.132	0.385

Table 3. Business Cycles in the U.S. Transportation Sector

Transportation Business Cycles			Leads (-) and Lags (+), in months, of Transportation Business Cycles relative to NBER Business Cycles			Leads (-) and Lags (+), in months, of Transportation Leading Index relative to Transportation Business Cycles	
P	T	Duration	P	T	Duration	P	T
03/79	08/80	17	-10	+1	6	-4	-1
01/81	2/83	25	-6	+3	16	-1	-13
05/90	06/91	13	-2	+3	8	-16	-6
11/00	12/01	13	-4	+1	8	-20	-3
Mean		17	-6	+2	10	-10	-6
Median		15	-5	+2	8	-10	-5
Std Dev.		6	3	1	4	9	5

Table 4. Standardization Factors for Constructing Transportation CLI

U.S. transportation leading indicators	Factors (Up to 10/2002)
1. DJTA (20 stocks)	0.088
2. PMI-inventory diffusion index (PMI-IV)	0.081
3. TE's new orders (NO)	0.202
4. TE's shipments (Shipment)	0.124
5. TE's industrial production index (Production)	0.219
6. TE's Payrolls (Payrolls)	0.163
7. Consumer Sentiment Index (CSI)	0.122

Data Issues and Strategies in Population Projections

Chair: Frederick W. Hollmann, U.S. Census Bureau, U.S. Department of Commerce
Discussant: Peter Johnson, U.S. Census Bureau, U.S. Department of Commerce

Overcoming Data Quality Problems Encountered in Preparing Population Projections for the Current and Former "Outlying/Insular Areas"

William H. Wannall III, U.S. Census Bureau, U.S. Department of Commerce

Developing projections for small nations is a daunting task, and island nations add other elements to the mix, including seasonal tourism, relatively small populations distributed over hundreds of islands, and large foreign-born populations. This paper will discuss the data quality issues peculiar to demographic data of America's "outlying areas"—American Samoa, the Commonwealth of the Northern Mariana Islands, Guam, the Federated States of Micronesia, the Marshall Islands, Palau, and the U.S. Virgin Islands—and the solutions used in preparing cohort-component population projections. Such issues include incomplete vital registration data, migration models, and indirect mortality and fertility measures.

Data Consistency Issues in Projecting Births and the Population Under Age 1 by Race

Myoung Ouk Kim and Ching-li Wang, U.S. Census Bureau, U.S. Department of Commerce

The Census Bureau projects population by years of age, sex, and race/Hispanic origin. Age 0 projections affect the projections of subsequent age groups. Projected births are generally used to project this population, however, the race groups used by the National Center for Health Statistics are not exactly the same as used in the census. In addition, the census coverage rates changed from census to census, and also vary from race to race. This paper examines the impact of the inconsistency of race data, and census coverage rates on the Census Bureau's state population projections. The paper also presents the procedures used to minimize the impact of these problems for upcoming State population projections.

Measurement of Internal Migration for Census Bureau's State Population Projections by Age, Sex, and Race

Caribert Irazi and Ching-li Wang, U.S. Census Bureau, U.S. Department of Commerce

IRS-extracted individual income tax returns are used to derive migration flows between counties and between States. For State population projections, the time series of migration rates are used to project migration flow rates. The migration rates by age, sex, and race derived from the census data are then used to dis-aggregate the projected IRS migration flows into age, sex, and race details. In the past, the IRS and census rates were developed and projected for all 2,550 migration flows between States and then aggregated to State totals. This paper examines the data issues in these two data sources, and presents the alternative approaches to projecting the internal migration by age, sex, and race.

DATA CONSISTENCY ISSUES IN PROJECTING BIRTHS AND THE POPULATION UNDER AGE 1 BY RACE

Myoung-Ouk Kim and Ching-li Wang
Population Division, U.S. Census Bureau

I. Introduction

The Census Bureau prepares state population projections by single years of age, sex, race and Hispanic origin. The cohort-component method is used to project state populations. Each component of population change – births, deaths, internal migration and international migration is projected separately. The vital statistics provide the basic input data for two components - births and deaths. The number of births will determine a significant part of population growth in the future as the cohort of births ages from infants to adults. The accuracy of projected births will affect the accuracy of the population at subsequent ages in the future. To project births requires the development of fertility rates, which use the live births as numerators and the female population of child-bearing age as denominators. Generally, the vital statistics are compiled by the National Center for Health Statistics (NCHS), and the population bases are derived from censuses or population estimates. In most cases, this is a straightforward procedure. However, when the fertility rates are constructed by race, the consistency of the two data sources with regard to race becomes a major concern for projections.

The data collection procedures and race classification are very different between the Census and NCHS vital statistics. The census data were largely collected through self-identification in assigning racial and ethnic groups while the vital statistics are collected by the states and submitted to the NCHS. The race of the infants generally is the race of the mother or father, depending on how they are classified (Adlakha et al, 2002). This difference results in the issue of data consistency between denominator and numerator in preparing appropriate fertility rates for projections.

In addition, the census has never been 100 percent complete, while the under-registration of births is known to be very minimal (Sink, 1997). Therefore, the birth statistics along with death statistics and other administrative records are used to evaluate population coverage in the census (Robinson, 2001). The census coverage rates vary from census to census and from race to race. Based on Demographic Analysis, the undercount rates changed from 5.4 percent in 1940 to 3.1 percent in 1960, and 1.2 percent in 1980 (Robinson et al., 1993). The net undercount rate was reduced substantially from 1.65 percent in 1990 to 0.12 percent in 2000 (Robinson and Adlakha, 2002). However, the net undercount rates for race groups other than White were substantially higher. The estimate of undercount for Whites from the Post-Enumeration Survey in 1990 was 0.9 percent for the U.S., while the estimates of undercount were 4.4 percent for Blacks, 4.5 percent for American Indian, and 2.3 percent for Asian and Pacific Islanders. (Census Bureau website).

These data issues have become more complicated after Census 2000, which allowed respondents to identify with more than one race and included a separate category of Native Hawaiian and Other Pacific Islanders to comply with the 1997 OMB directive (Federal Register, 1997). NCHS has continued to use the race classification from the 1977 OMB directive which only requires 4 groups - Whites, Blacks, American Indians, and Asian and Pacific Islanders. The changes in race classification in Census 2000 increased the inconsistency of data for denominators and numerators in calculating the fertility rates by race.

In this paper, we will examine the race and ethnicity inconsistency between NCHS birth data and the Census by comparing the difference between the Census age 0 population and age 0 estimated for the census date based on NCHS data. Then, the paper shows how this inconsistency would affect the projections of births by race, and discusses what we do to minimize the impact of these problems on projections.

II. Race and Ethnicity in Censuses

The census data were largely collected through questionnaires asking respondents to answer various demographic and socioeconomic questions. By answering the questions in the census, the respondents identify their races by themselves – self-identification in assigning racial and ethnicity groups. However, the race of a child in the Census is reported by the person who fills out the form. It is very likely that this is done by the child's parents. The race categories on the census forms changes from census to census (Bennett, 2000; Lee, 1993).

The final tabulations of race are also affected by the requirements in the Office Management and Budget (OMB) standard. This will affect comparability of race and ethnicity between censuses. To illustrate the change in race classification and requirement, let us compare the race and ethnicity groups between the 1990 census and the Census 2000.

The 1990 census questions on race included 14 separate response categories: White, Black or Negro, Indian (Amer.), Eskimo, Aleut, and nine Asian and Pacific Islander groups which are Chinese, Japanese, Filipino, Asian Indian, Hawaiian, Samoan, Korean, Guamanian and Vietnamese; and plus two residual categories (Other Asian and Pacific Islander and Other race). Three categories required write-ins: Indian (Amer.), where respondents were asked to print the name of their enrolled or principal tribe, and for those who reported as "Other Asian or Pacific Islander" or "Other race," who were asked to write in the name of their group or race. For population estimates and projections purpose, these race categories were aggregated and modified into four groups: White, Black, American Indian, and Asian and other Pacific Islanders by Hispanic origin.

Census 2000 collected race categories in 12 separate response boxes for White, Black, African (Am.) or Negro, American Indian or Alaska Native, Asian Indian, Chinese, Filipino, Japanese, Korean, Vietnamese, Native Hawaiian, Guamanian or Chamorro and Samoan, plus 3 write-in boxes - Other Asian, Other Pacific Islander, and Some Other Race. The form requests additional write-in information for several responses: American Indian or Alaska Native, where the respondent is asked to provide the name of his or her enrolled or principal tribe, in addition to the three write-in groups.

The Office of Management and Budget's 1997 revised standards for collecting and presenting data on race and ethnicity (Federal Register, 1997), identified five minimum race categories: White, Black or African American, American Indian and Alaska Native, Asian, and Native Hawaiian and other Pacific Islander. In addition, the OMB recommended to allow respondents be given the option of selecting on or more races to indicate their racial identity. With the option given to respondents to select one or more race races based on the 5 race groups, there are 31 possible combinations of the 5 race groups – 5 with one race alone, 10 with two races, 10 with three races, 5 with four races, and 1 with five races. Collection of additional detail on race is permitted as long as the additional categories can be aggregated into the minimum categories. The 1997 standards continue recommend the use of a separate question on Hispanic or Latino ethnicity.

For purpose of estimates and projections and to follow the new OMB standard for consistency with other federal agencies, the race data were modified to eliminate the "Some other race" category from the census data. Combining all race alone and the multi-race groups by Hispanic origin yields a total of 12 race and Hispanic origin groups to be used for the state projections: non-Hispanic Whites, non-Hispanic Blacks, non-Hispanic American Indian and Alaskan Natives, non-Hispanic Asian, non-Hispanic Native Hawaiian and Other Pacific Islanders, and non-Hispanic Multi-race, Hispanic White, Hispanic Black, Hispanic American Indian and Alaskan Natives, Hispanic Asian, Hispanic Native Hawaiian and other Pacific Islanders, and Hispanic Multi-race.

III. Race and Ethnicity Classification of Births in NCHS

The vital statistics data are collected by the states and submitted to the NCHS. The birth certificates ask the race and Hispanic origin of mother and father. Therefore, the determination of race of births is the matter of how to assign the race of father or mother to the babies.

In 1988 and prior years, NCHS used the minority rule to classify births by race. When the parents were of the same race, the race of child was as the same as the parents. When the parents were of different races and one parent was White, the child was assigned to the other parent's race. When the parents were of different race and neither parent was White, the child was assigned to father's race, with one exception -- If either parent was Hawaiian, the child was assigned to Hawaiian. In 1989 and later, NCHS adopted the mother rule, that is, the live births are tabulated by the race of mother. The only exception is that if the mother's race is not reported, fathers' race is used (NCHS, 1991).

The race categories used in NCHS changed over time. Prior to 1989 after 1978, the categories used were White, Black, American Indian, Chinese, Japanese, Hawaiian, Filipino, Other Asian or Pacific islander, and other. Before 1978, the category "Other Asian or Pacific Islander" was not identified separately but included with "Other" races. The separation of this category allows identification of the category "Asian or Pacific Islander" by combining the new category "Other Asian or Pacific Islander" with Chinese, Japanese, Hawaiian, and Filipino.(NCHS, 1988) From 1992 on, some areas started reporting additional Asian or Pacific Islander codes for race to include Asian

Indian, Korean, Samoan, vietnamese, and Guamania. (NCHS, 1994)

Concurrent with the 1978 revision of U.S. Standard Certificate of Live Birth, the National Center for Health Statistics recommended that States add items to identify the Hispanic or ethnic origin of the newborn's parents. Two formats were used: (1) an open-ended item to obtain the specific origin or descent of each parent, for example, Italian, Mexican, or English; and (2) an item directed toward the Hispanic population, requesting only the specific Hispanic origin (Mexican, Puerto Rican, Cuban, and so forth). In 1987, items requesting Hispanic or ethnic origin were included on the birth certificates of 23 States and the District of Columbia (NCHS, 1988). In 1989, Hispanic origin of the parents was reported on the birth certificates in 47 States and the District of Columbia, an increase of 17 states from 1988. By 1990, 48 states reported Hispanic origin in 1990 and 49 states reported in 1991 with New Hampshire the only one left until 1993 (NCHS 1993).

Since 1989, NCHS has had a new birth registration system in effect, which includes detailed racial and ethnic information about both parents. So it is possible to tabulate the race of births by the race and ethnicity of either parent. However, the NCHS continues to tabulate the race of births by race of mother with four groups: White, Black, American Indian, Asian and Pacific Islander (which including Chinese, Japanese, Hawaiian, Filipino, Other Asian or Pacific Islander).

In 2000, NCHS race data also include four groups as in the 1990 Census in which American Indian includes Aleuts and Eskimo, and Asian and Pacific Islander includes Chinese, Japanese, Hawaiian, Filipino, Asian Indian, Korean, Samoan, Vietnamese, Guamanian, and Other Asian or Pacific Islander. Hispanic Origin of mother includes Non-Hispanic and Hispanic (Mexican, Puerto Rican, Cuban, Central or South American and other and unknown Hispanic) which are very much the same as used in the 1990 Census. All the race data have not provided the option for multirace selection as in the Census 2000. In regard to the OMB's new standard, the National Vital Statistics System, which is based on data collected by the States, will not be fully implemented until later in the decade (NCHS, 2003b). Therefore, we can see that the NCHS has been trying to use the race and ethnicity classification used by the Census Bureau. However, it seems that it has been always a few years behind the Census Bureau due to different data collection procedures and needs for cooperation from state health agencies.

IV. Estimates of population age 0 by race and Hispanic origin as of April 1, 2000

To evaluate the comparability of the NCHS and Census data sources, we prepared the estimates of population age 0 based on the vital statistics. The component method was used to derive population age 0 based on live births and deaths for age 0 between 4/1/1999 and 3/31/2000, adjusted by domestic and international migration of population age 0. Those who were born between 4/1/1999 and 3/30/2000 were under age 1 on April 1, 2000. By subtracting the number of infant deaths from births in this period of time, we can get a rough estimate of age 0 on the census date. These estimates were adjusted by domestic and international migration for this age group based on census migration data and foreign born population. In order to compare the estimates based on births by race, the Census population 2000 by race are converted or bridged to old race groups as used in the NCHS vital statistics based on a "fractional assignment" procedure (see Appendix A and NCHS,2003b).

Table 1 shows the estimates of population age 0 for the United States based on NCHS data compared with two Census 2000 populations, one with original census race groups, and another one with census population bridged to old race groups as used in NCHS. As table 1 shows, the Census 2000 shows that total estimated population of age 0 based on NCHS is 3,975,913 while the Census 2000 population of age 0 is 3,805,648, a -4.3 percent. The estimate based on NCHS data is 170,265 more than the Census population age 0 in 2000.

Differences between the NCHS and Census age 0 become much more severe when disaggregated by race and Hispanic origin. Among the non-Hispanics, the relative differences between NCHS and Census 2000 are generally less for the bridged estimates than for those without bridging due to the fact that a portion of population in each race group was counted in multi-race group. The largest differences are for American Indians and Asian/Pacific Islanders. For Asian/Pacific Islanders, the un-bridged Census 2000 population is 25.3 percent below the NCHS based population age 0. For bridged race groups, Census 2000 is 12.4 percent below the NCHS-based estimates. For American Indians the Census 2000 population is 11 percent lower than NCHS without bridging, whereas with bridging the census population is 16.6 percent higher – this is the only instance among non-Hispanics where the census population is higher than NCHS figures.

Table 1. Difference of Population Age 0 Between Census and Estimates Based on NCHS Vital Statistics U.S.: April 1, 2000

Population Age 0	Total Population	NonHispanic					
		Total	White	Black	Am Indian	Asian/Paci	Multi-Race
Estimates based on NCHS (1)	3,975,913	3,193,944	2,383,185	589,416	37,827	183,516	-
% difference from Census	4.5	5.3	5.4	3.6	-14.2	14.1	-
Census 2000:							
1. Without Bridging	3,805,648	3,034,595	2,210,111	540,160	33,665	137,156	113,503
Difference from NCHS:							
Number	-170,265	-159,349	-173,074	-49,256	-4,162	-46,360	113,503
Percent	-4.3	-5.0	-7.3	-8.4	-11.0	-25.3	-
2. With Bridging (2)	3,805,648	3,034,595	2,260,488	569,183	44,099	160,826	-
Difference from NCHS:							
Number	-170,265	-159,349	-122,697	-20,233	6,272	-22,690	-
Percent	-4.3	-5.0	-5.1	-3.4	16.6	-12.4	-

	Hispanic					
	Total	White	Black	Am Indian	Asian/Paci	Multi-Race
Estimates based on NCHS (1)	781,969	761,351	13,996	2,620	4,003	-
% difference from Census	1.4	8.0	-63.3	-83.2	-67.5	-
Census 2000:						
1. Without Bridging	771,053	697,187	34,006	12,638	8,623	18,599
Difference from NCHS:						
Number	-10,916	-64,164	20,010	10,018	4,620	18,599
Percent	-1.4	-8.4	143.0	382.5	115.4	-
2. With Bridging (2)	771,053	705,041	38,123	15,591	12,299	-
Difference from NCHS:						
Number	-10,916	-56,310	24,127	12,972	8,296	-
Percent	-1.4	-7.4	172.4	495.2	207.3	

Notes: 1.The estimates are based on NCHS live births and infant deaths between 1999 and 2000,. plus international migration for population age 0.

2. A 'fractional assignment" procedure is used to convert the Census 2000 race groups into the 1990 Census race groups used by the NCHS. The fractional assignment is to distribute the multi-race group population to each of the four race groups according to the fraction of two or more races. For example, one half of the white/black group is added to White,and the other half is added to Black. One third of White/Black/Asian is added to white, one third is added to Black, and the other one third is added to Asian/Pacfic Islander.

Though the relative differences for the Hispanic population as a whole are smaller than those for non-Hispanics, the discrepancies for Hispanics are much higher when broken down by race. Except for Whites, these differences are all in excess of 100 percent! For example, the estimate of Hispanic American Indian age 0 based on NCHS data was only 2,620, while the census count age 0 was 12,638 without bridging, and 15,591 with bridging, a 495 percent difference in the latter case. This indicates that there are serious discrepancies in reporting races of Hispanics in the NCHS and Census for the age 0 population.

Since the race classification in the Census 2000 is different from the NCHS, the comparison of estimates and Censuses may not reflect the actual discrepancy. So, let us compare the NCHS-based estimates for 1990 with the 1990 census where the race classification was the same. As Table 2 shows, the relative differences for 1990 are generally less than the differences for 2000. Minority groups still show large relative differences between the NCHS and Census numbers, especially in Hispanic race groups. For example, the Census counts are much higher than NCHS estimates for American Indian (337.1 percent), Asian Pacific Islander (122.5 percent), and Black (180.7 percent) for the Hispanic age 0 population. In contrast, the Census for Hispanic White is only 6.8 percent less than NCHS estimates (The NCHS estimates are used as the standard for comparison in this paper).

Table 2. Difference of Population Age 0 Between Census and Estimates Based on NCHS Vital Statistics,U.S.: April 1,1990

Population Age 0	Total Population	NonHispanic				
		Total	White	Black	Am. Indian	Asian/Paci
Estimates based on NCHS(1)	4,042,843	3,491,872	2,667,111	653,143	37,317	134,302
1990 Census(2)	3,947,313	3,399,748	2,631,225	607,078	40,087	121,358
Diffrence from NCHS						
Number	-95,530	-92,124	-35,886	-46,065	2,770	-12,944
Percent	-2.4	-2.6	-1.3	-7.1	7.4	-9.6
		Hispanic				
		Total	White	Black	Am. Indian	Asian/Paci
Estimates based on NCHS(1)		550,971	532,697	11,288	1,783	5,203
1990 Census(2)		547,565	496,514	31,683	7,795	11,573
Diffrence from NCHS						
Number		-3,406	-36,183	20,395	6,012	6,370
Percent		-0.6	-6.8	180.7	337.1	122.5

Notes: 1.The estimates are based on NCHS live births and infant deaths between 1989 and 1990, plus international migration for population age 0.
2.The 1990 Census population is based on modified age-race-sex(MARS) file to correct mis-reporting of ages in the census, especially for the age 0.

These discrepancies suggest that if we use the births by race to estimate population age 0 directly from the NCHS, according to Table 1, we will see the estimates of age 0 increase from the census count by 5.4 percent for non-Hispanic White, 3.5 percent for Non-Hispanic Black, 14.1 percent for non-Hispanic Asian and Pacific Islanders, and 7.9 percent for Hispanic Whites in one year. We would also see 14.2 percent decrease for non-Hispanic American Indian, 63.2 percent decrease for Hispanic Black, 83.2 percent decrease for Hispanic American Indian and 67.5 percent decrease for Hispanic Asian in one year.

If we use the births by race directly from NCHS to project population we will have the same magnitude of the changes for age 0 in the first year of projection, and see the continuing decrease of Hispanic non-white groups in the future. There will be more growth of young ages in non-Hispanic than in Hispanic population because the non-Hispanic would grow 5.0 percent and Hispanic would grow only 1.4 percent in the beginning. This is completely opposite to the demographic trends we are observing. The impact of race discrepancies between NCHS and Census is a major concern in producing appropriate projections.

V. Estimates and Census Comparison by State

Tables 1 and 2 show the discrepancy between the estimates based on NCHS vital statistics and censuses at the national level. The discrepancies among states are more problematic. Table 3 shows state levels of percent difference between Census 2000 and NCHS-based estimates of the age 0 population. At the national level, total age 0 population is 4.3 percent lower than NCHS birth data. However, when broken down to the state level, it varies dramatically. The discrepancies range from -11.9 percent in the District of Columbia (DC) to 1.9 percent in North Dakota.

For non-Hispanics as whole, the discrepancies range from -14.0 percent in the District of Columbia to 0.4 percent in North Dakota. For non-Hispanic Whites, the discrepancies range from -13.0 percent in New Mexico to 30.2 percent in Hawaii. For non-Hispanic Blacks, the discrepancies range from -16.4 percent in the District of Columbia to 147.2 percent in Idaho. For non-Hispanic American Indians, the discrepancies range from -24.7 percent in Nebraska to 216.2 percent in Virginia. For non-Hispanic Asian and Pacific Islanders, the discrepancies range from -24.0 percent in Hawaii to 75.4 percent in Vermont. For the Hispanic race groups the differences are even larger. For Hispanics as a whole, the discrepancies range from -8.2 percent in Florida to 177 percent in West Virginia. Most F states have larger numbers in the Census compared to NCHS estimates. For Hispanic Whites, the discrepancies range from -13.4 percent (New York) to 158.8 percent in West Virginia. Most of the higher number in the Census were over 500 percent (California, Georgia, Michigan, New Jersey, New Mexico and New York) and reached as high as a 2,197 percent of American Indians in Texas.

VI. Projecting births and population under age one

With discrepancies of such magnitude between NCHS and Census race data for states, what can be done to develop appropriate fertility rates to project births and age 0? First of all, we need to convert the Census 2000

Table 3. Percent Difference At Age 0 Between Census and Estimates Based on NCHS Vital Statistics by State: April 1, 2000

State	Total	NonHispanic					Hispanic				
		Total	White	Black	Am Indian	Asian/Paci	Total	White	Black	Am Indian	Asian/Paci
United States	-4.3	-5.0	-5.1	-3.4	16.6	-12.4	-1.4	-7.4	172.4	495.2	207.3
Alabama	-4.7	-5.4	-5.9	-6.1	123.2	24.9	19.5	6.8	-	-	-
Alaska	-4.6	-5.9	-5.3	34.3	-19.1	23.1	14.5	10.4	-	105.6	-59.8
Arizona	-6.6	-9.4	-10.0	11.4	-12.4	-17.0	-2.3	-7.4	672.6	394.8	246.3
Arkansas	-2.3	-3.5	-4.1	-3.4	38.0	10.3	16.8	9.8	-	-	-
California	-7.4	-7.7	-6.7	-0.9	39.8	-16.3	-7.0	-12.3	448.2	634.2	451.6
Colorado	-4.3	-7.0	-8.3	7.3	17.8	-6.4	2.9	-2.9	413.4	254.5	139.3
Connecticut	-0.6	-2.3	-3.1	0.8	135.6	-4.3	8.9	-7.2	936.1	-	-
Delaware	-4.6	-6.0	-6.9	-4.4	35.8	-2.6	10.1	-6.9	-	-	-
District of Columbia	-11.9	-14.0	-11.4	-16.4	-	19.3	4.7	-11.9	-	-	-
Florida	-6.2	-5.6	-3.6	-10.3	28.8	-10.9	-8.2	-12.1	112.6	17.2	489.5
Georgia	-6.2	-6.7	-7.3	-6.3	89.7	-7.3	-1.4	-9.1	348.4	507.4	155.0
Hawaii	-7.9	-10.4	30.2	40.8	7.6	-24.0	8.5	50.0	194.8	42.5	-15.7
Idaho	-1.5	-4.5	-5.5	147.2	20.3	-7.6	20.4	14.2	-	-	-
Illinois	-4.3	-5.3	-4.2	-8.2	123.7	-10.3	-0.5	-5.3	512.1	-	-
Indiana	-1.9	-3.4	-5.0	5.8	197.1	3.1	23.0	14.5	698.2	-	-
Iowa	-0.5	-1.7	-2.6	30.4	-8.1	-9.9	22.4	16.5	-	-	-
Kansas	-2.2	-5.3	-7.2	9.4	41.5	-5.4	21.4	14.0	499.2	-	-
Kentucky	-2.7	-3.7	-5.2	8.7	79.6	1.5	51.2	41.0	-	-	-
Louisiana	-3.6	-4.5	-4.0	-5.7	54.6	-7.8	31.0	12.6	626.1	-	10.4
Maine	-0.8	-1.3	-2.2	108.0	12.9	0.6	53.3	40.0	-	-	-
Maryland	-3.8	-4.8	-5.3	-5.4	54.8	3.2	12.0	0.4	534.3	-	-57.5
Massachusetts	-3.8	-4.9	-4.1	-10.2	108.0	-12.1	5.1	4.7	-8.6	-	-
Michigan	-1.6	-3.3	-4.8	-0.7	94.4	3.6	31.0	19.9	676.6	619.9	-15.9
Minnesota	-2.1	-3.3	-4.2	9.4	6.3	-7.0	18.6	8.3	425.3	-	-
Mississippi	-3.5	-4.5	-3.8	-5.3	11.9	-15.3	77.0	45.6	-	-	-
Missouri	-3.5	-4.6	-4.5	-6.1	55.4	-11.0	29.9	20.2	440.7	-	-
Montana	0.1	-1.1	-1.5	-	-2.4	-7.4	39.2	18.5	-	-	-
Nebraska	-2.1	-4.0	-4.5	13.7	-24.7	-11.0	14.5	7.9	-	-	-
Nevada	-4.8	-7.2	-8.2	6.9	-5.1	-17.1	0.2	-4.4	290.1	277.1	135.8
New Hampshire	-1.6	-2.4	-3.8	78.0	-	14.4	29.6	20.3	-	-	-
New Jersey	-3.1	-3.2	-0.8	-9.5	115.7	-8.6	-3.0	-8.2	26.0	573.9	589.0
New Mexico	-2.3	-10.8	-13.0	39.5	-12.4	-1.9	5.8	-0.1	-	533.6	-
New York	-3.9	-4.3	-2.3	-7.1	115.4	-16.0	-2.4	-13.4	59.8	1669.0	806.7
North Carolina	-4.7	-5.3	-6.4	-2.9	14.9	-14.2	1.8	-7.1	393.5	-	-
North Dakota	1.9	0.4	0.1	62.5	-2.2	-9.1	95.0	63.4	-	-	-
Ohio	-2.4	-3.7	-5.4	4.3	111.0	-5.8	50.1	31.9	604.6	259.2	-
Oklahoma	-2.9	-5.4	-11.6	8.0	26.4	-4.4	24.4	10.3	660.7	372.3	-
Oregon	-3.1	-6.0	-7.4	30.0	27.1	-8.1	12.7	6.5	-	461.5	-
Pennsylvania	-2.2	-3.6	-4.0	-2.0	21.2	-3.3	24.7	6.2	398.5	206.2	-
Rhode Island	-1.8	-3.4	-2.1	-7.2	-20.6	-14.4	6.6	-11.7	383.3	-	-
South Carolina	-2.3	-2.9	-4.3	-1.1	79.9	-6.0	16.5	7.3	468.6	-	-43.2
South Dakota	-1.0	-2.4	-2.1	114.7	-9.4	-12.5	71.8	36.9	-	-	-
Tennessee	-4.2	-5.0	-5.6	-3.5	77.2	-5.3	18.9	8.1	656.9	-	-
Texas	-6.2	-6.3	-7.6	-1.9	74.8	-11.0	-6.2	-9.5	821.7	2197.0	1725.0
Utah	-4.5	-5.6	-6.2	87.2	0.1	-10.6	2.9	-2.6	-	-	-
Vermont	-2.4	-3.2	-5.2	145.6	-	75.4	124.4	108.2	-	-	-
Virginia	-3.8	-4.7	-5.3	-3.5	216.2	-8.1	6.7	-4.3	482.5	-	45.1
Washington	-3.1	-6.1	-7.0	20.2	0.1	-13.8	16.4	8.1	471.3	378.0	237.1
West Virginia	-1.9	-2.7	-4.3	28.2	-	15.0	177.0	158.8	-	-	-
Wisconsin	-1.2	-3.1	-3.5	0.5	4.9	-6.3	27.6	17.5	488.6	229.9	-
Wyoming	1.1	-2.0	-1.9	21.5	-9.4	-7.1	34.4	24.7	-	-	-

Notes: 1. The percentages for population less than 30 are not shown.
2. The percentages are the difference between census and estimates divided by the estimates.

population by race to the old race groups to be consistent with the NCHS race groups as much as possible. As described in estimating age 0 by race (Appendix A), the Census 2000 populations by race were converted using a "fraction assignment" procedure to convert the 2000 population by race to approximate the race groups as provided by NCHS.

Secondly, we use the concept of a "standard schedule" to apply the fertility rates from certain race/Hispanic groups with reasonable rates to other groups with data problems. The only race groups for which the fertility rates were

developed are (1) Non-Hispanic total population, (2) Non-Hispanic White, (3) Non-Hispanic Black, (4) Non-Hispanic American Indian, (5) Non-Hispanic Asian and Pacific Islander, and (6) Hispanic origin. This is because many states had small population counts in the detailed age groups for some races. The NCHS vital statistics by age, sex, race, and Hispanic origin also contain many small or empty data cells, especially in Hispanic origin race groups with severe data inconsistency as described before. The age-specific fertility rates by race and Hispanic origin will not be reliable for projections in some states.

Then, the rates for the overall Hispanic origin group were applied to each Hispanic race group (including Hispanic White, Hispanic Black, Hispanic American Indian, Hispanic Asian, and Hispanic Native Hawaiian or other Pacific Islander, and multi-race). The rates for the non-Hispanic total population were applied to the non-Hispanic multi-race group. The rates for Asian/Pacific Islander are applied to Asian and Native Hawaiian or Pacific Islanders. In addition, if the state numbers for particular race groups are too small to produce appropriate rates, the national rates are used (Graphically, when the number of births is below 350, the rates fluctuate dramatically with missing data in some groups).

Table 4 shows the projected births and population aged 0 based on NCHS births by race and based on census distribution of age 0 by race. The projections are based on the procedures described here for the purpose of evaluation, which do not reflect the official projections in progress. The projections were made for the United States as a whole for illustrative purposes.

(A) Projected births and age 0 directly from fertility rates by race

With the procedures described above along with infant

Table 4, Projected 2001 Births and Population Ages 0 and 1 for the U.S.: Based on NCHS Births by Race and Based on Census Age 0 by Race

Age	Total Population	NonHispanic						
		Total	White	Black	Am. Indian	Asian	Hawaiian	Multi-Race
Census 2000								
Age 0	3,805,648	3,034,595	2,210,111	540,160	33,665	131,494	5,662	113,503
Age 1	3,820,582	3,074,896	2,244,143	547,229	34,692	132,186	5,960	110,686
Ratio of Age 0 to 1	1.00	0.99	0.98	0.99	0.97	0.99	0.95	1.03
2001 Projections, NCHS Base								
Births	4,011,175	3,206,355	2,352,262	596,989	30,592	173,722	5,770	47,020
Age 0	3,999,467	3,192,263	2,342,644	590,409	30,415	176,253	5,679	46,863
Age 1	3,816,397	3,037,606	2,210,346	539,581	33,641	134,875	5,644	113,519
Ratio of Age 0 to 1	1.05	1.05	1.06	1.09	0.90	1.31	1.01	0.41
% difference Age 0 from Census	5.09	5.20	6.00	9.30	-9.65	34.04	0.30	-58.71
2001 Projections, Census Base								
Births	4,011,175	3,198,482	2,329,471	569,332	35,483	138,595	5,968	119,633
Age 0	3,999,398	3,184,343	2,319,950	563,071	35,275	140,937	6,032	119,078
Age 1	3,816,316	3,037,513	2,210,338	539,569	33,641	134,654	5,757	113,554
Ratio of Age 0 to 1	1.05	1.05	1.05	1.04	1.05	1.05	1.05	1.05
% difference Age 0 from Census	5.09	4.93	4.97	4.24	4.78	7.18	6.53	4.91

	Hispanic						
	Total	White	Black	Am. Indian	Asian	Hawaiian	Multi-Race
Census 2000							
Age 0	771,053	697,187	34,006	12,638	6,198	2,425	18,599
Age 1	745,686	674,180	33,539	12,240	6,028	2,334	17,365
Ratio of Age 0 to 1	1.03	1.03	1.01	1.03	1.03	1.04	1.07
2001 Projections, NCHS Base							
Births	804,820	739,065	34,228	12,747	5,634	2,267	10,879
Age 0	807,204	741,343	34,244	12,818	5,648	2,285	10,866
Age 1	778,791	704,476	34,217	12,771	6,234	2,455	18,638
Ratio of Age 0 to 1	1.04	1.05	1.00	1.00	0.91	0.93	0.58
% difference Age 0 from Census	4.69	6.33	0.70	1.42	-8.87	-5.77	-41.58
2001 Projections, Census Base							
Births	812,693	734,839	35,842	13,320	6,533	2,556	19,603
Age 0	815,055	737,119	35,849	13,394	6,558	2,571	19,564
Age 1	778,803	704,463	34,215	12,776	6,246	2,454	18,649
Ratio of Age 0 to 1	1.05	1.05	1.05	1.05	1.05	1.05	1.05
% difference Age 0 from Census	5.71	5.73	5.42	5.98	5.81	6.02	5.19

The projections are based on the Cohort-Component Method using current fertility, mortality, and international migration rates which do not reflect the official projections in progress.

mortality and migration rates, we projected the population for 2001 from Census 2000 in order to evaluate the outcome (second panel of Table 4). The projected total number of births of 4,011,175 for the United States is close to the 4,025,933 in 2001 reported by NCHS (2003). The projected population age 0 is more than age 1 by 5 percent. This is reasonable because age 0 in Census 2000 was less than the NCHS estimated births by 4.3 percent (as shown in Table 1). However, the projected populations under one year of age were very different from the Census 2000 population age 0 for most of the race groups, especially for the multi-race group. For example, the projected non-Hispanic multi-race age 0 was 46,863, a reduction of 58 percent from 113,503 in Census 2000. This is contradictory to the general perception that the multi-race population should be increasing over time.

The unexpected outcome of the projections from using the fertility rates by race directly from the NCHS data can be seen in other race groups. For example, the projections show that between 2000 and 2001, the non-Hispanic Asian age 0 would increase by 31 percent, the non-Hispanic American Indian would decrease by 10 percent, while the non-Hispanic Blacks would increase by 9 percent. In other words, despite our use of the "standard schedule" procedure based on the NCHS data to project age 0 by race, the outcomes are still not acceptable. This is because the projection of births by race based on NCHS data cannot be used to directly project the population age 0 by race.

(B) The use of census proportions at age 0 by race and Hispanic origin

Race of Mother from NCHS birth data was used to calculate fertility rate by race, which we use to project the births by race. For instance, when we apply the non-Hispanic total fertility rates to multi-race females of child-bearing ages, the number of births represents only the births to multi-race mothers. The multi-race births should include not only those born by multi-race mothers, but also those born by single race mothers with fathers of different races. How the races of these births of interracial marriage were reported in the census is another question. Thus, there always is a discrepancy of race between births based on race of mother in the NCHS and age 0 in the census.

Since we project the population from a Census base, we need to project age 0 by race as consistent with the census as possible. It was decided to use the proportion of age 0 by race in the census to distribute the projected total births and derive the projected population age 0 by race. This procedure improved the projections dramatically as shown in the third panel of Table 4.

Based on the census racial distribution, the multi-race age 0 would increase by 4.9 percent between 2000 and 2001 instead of a reduction of 59 percent based on projected NCHS births by race. The non-Hispanic American Indian would increase by 4.8 percent instead of a reduction of 9.7 percent. The dramatic improvement in projecting age 0 can be seen in all races of Hispanic origin.

VII. Conclusions and Discussions

The data consistency issues between the National Center for Health Statistics and the Census Bureau in race statistics have been a major concern in preparing appropriate population estimates and projections. The data consistency issues include differences in data collection, coverage, and race classification between the vital statistics and the census population. This paper has examined the impact of the inconsistency between these two sources of data on estimates and projections of the age 0 population. Since the coverage rates for the vital statistics is considered higher than that of the Census, the vital statistics along with other administrative records have been used to evaluate the coverage of censuses. Though the use of total births and deaths to evaluate census coverage is reasonable, the use of race statistics from NCHS would not be appropriate due to inconsistency in race reporting between NCHS and the Census.

(A) Differences between fertility by race, births by race, and age 0 by race.

This paper has shown that the use of births by race directly from the NCHS data to estimate or project the age 0 population will produce dramatic inconsistencies with the Census base population for many race and Hispanic groups. If the NCHS data are used directly to estimate or project age 0, we would see the Hispanic population grow more slowly than the non-Hispanic population and see a dramatic reduction of American Indians. This is not consistent with the demographic trends we are observing. Thus, the use of births by race directly from the NCHS will produce projections with undesirable results.

Our conclusion is that it is not appropriate to use NCHS births by race directly to estimate population age 0 by detailed race and Hispanic origin. However, it is appropriate to use fertility by race to project the total number of births. The use of births by race of mother is reasonable for projecting fertility because the level of

fertility is mostly determined by the characteristics of mother although some other social economic variables are involved.

One reason the use of births by race to estimate the population under age 1 is not appropriate is because in NCHS data, the births are tabulated by the race of mother. Based on the Census Quality Survey, the multi-race White-Black individual would report as Black more than White (3 to 1) if only one race were chosen. White-American Indians are about twice as more likely to report as White than American Indian. The ratio is about 3 to 2 for White-Asians (Bentley et al., 2003). Thus, use of the mother rule to determine the race of births is inconsistent with the race of infants reported in the census. Therefore, the use of births by race of mother is appropriate to develop fertility of all women, but cannot be used to estimate the population under age 1 by race directly.

B) The choice between NCHS consistent projections and census consistent projections

This paper has also shown that if we use the proportion of age 0 by race from the census to distribute the births to each race and Hispanic origin group, the projections for age improve dramatically. Since our projections are based on the census, the population by race and Hispanic detail in the projections must be consistent with the census data. The users of the projections will be comparing the projections with the census data to calculate growth rates or other indicators based on the census data. If our projections were based on NCHS race data, the projections would be inappropriate to monitor the demographic trends for particular race and Hispanic groups. Therefore, it is imperative that we should have census consistent projections and estimates.

(C) Projections for multi-race group

The greatest challenge in producing the current projections series is the change in race classifications found in Census 2000, especially the addition of the multi-race group. The use of the proportion of the population age 0 in the census to distribute total projected births to each race and Hispanic group improves the projection of age 0 by race substantially. However, it may only be appropriate in the first few years of the projection interval because it is assumed that the multi-race group should increase more dramatically as time goes on. For this reason, it is necessary to incorporate the interracial marriage data in to our projection model.

.

References

Adlakha, Arjun L. J. Gregory Robinson, and Amy Symens Smith, 2002. "Alternative Rules for Assigning Race of Birth: Effect on Birth Totals, Implication for Vital Rates and Census Undercount Estimates by Race." Paper presented at the Southern Demographic Association Meetings, Austin, Texas, October 10-12, 2002.

Bennett, Claudette, 2000. "Racial Categories Used in the Decennial Censuses, 1979 to the Present." Government Information Quarterly, Volume 17, Number 2, Pages 161-180

Bentley, Michael, Tracy Mattingly, Christine Hough, and Claudette Bennett, 2003. "Census Quality Survey to Evaluate Responses to the Census 2000 Question on Race: An Introduction to the Data" U.S. Census Bureau, Census 2000 Evaluation B.3 (April 3, 2003)

Jones A. Nicholas and Amy Symens Smith, 2001. "The Two or More Races Population: 2000"
U.S. Census Bureau, Census 2000 Brief, C2KBR/01-6, November, 2001.

Lee, Sharon M., 1993. "Racial Classifications in the US Census: 1890-1990." Ethnic and Racial Studies, Volume 16, Number 1, 75-94. (January 1993).

National Center for Health Statistics, 1988. "Public Use Data Tape Documentation-1988 Detail Natality," Hayattsville, Maryland.

National Center for Health Statistics, 1991. "Public Use Data Tape Documentation-1989 Detail Natality," Hayattsville, Maryland.

National Center for Health Statistics, 1993. "Public Use Data Tape Documentation-1991 Detail Natality," Hayattsville, Maryland.

National Center for Health Statistics, 1994. "Public Use Data Tape Documentation-1994 Detail Natality," Hayattsville, Maryland.

National Center for Health Statistics, 2003. "Births: Preliminary Data for 2002" Natinal Vital Statistics Reports, Volume 51, No. 11 (June 25).

National Center for Health Statistics, 2003b. "United States Census 2000 Population with Bridged Race Categories." Vital and Health Statistics, Series 2, Number 135. (September)

Office of Management and Budget, 1997. "Revisions to the Standards for Classification of Federal Data on Race and Ethnicity." Federal register, 62 (210): 58782-58790.

Robinson, J. Gregory, Bashir Ahmed, Pritwis Das Gupta, and Karen Woodrow. (1993). "Estimation of Population Coverage in the 1990 United States Census Based on Demographic Analysis." Journal of the American Statistical Association, Vol. 88, No. 423: 1061-10

Robinson, J. Gregory, 2001a, "ESCAPEII: Demographic Analysis Results" U.S. Census Bureau, Executive Steering Committee for A.C.E. Policy II, Report No. 1

Robinson, J. Gregory, 2001b, "Accuracy and Coverage Evaluation: Demographic Analysis Results," U.S. Census Bureau, DSSD Census 2000 Procedure and Operations Memorandum Series B-4 (March 12, 2001).

Robinson, J. Gregory and Arjun Adlakha, 2002. "Comparison of A.C.E. Revision II Results with Demographic Analysis," U.S. Census Bureau, DSSD A.C.E. Revision Estimates Memorandum Series # PP-41.

Sink, Larry, 1997. "Race and Ethnicity Classification Consistency Between the Census Bureau and the National Center for Health Statistics." U.S. Census Bureau, Population Division Working Paper No. 17. (February, 1997).

U.S. Census Bureau, 1990. "Age, Race, and Hispanic Origin Information from the 1990 census: a Comparison of Census Results Where Age and Race have been Modified." 1990 CPH-L-74.

U.S. Census Bureau, 1998, "Race of Wife by Race of Husband: 1960, 1970, 1980, 1991, and 1992." Census Bureau internet, http://www.census.gov/population/socdemo/race/interracetabl.txt, 06/10/98

Appendix A

The procedures to estimate population under age 1 as of April 1, 2000 are as follows.

1. Live births minus deaths for age 0 between 4/1/1999 and 3/30/2000 (old race classification):

(1999 births) * (275/365) + (2000 births) *(91/366) -
(1999 deaths) * (275/365) + (2000 deaths) *(91/366)

2. Estimates of domestic migration:

The census migration is only for the population age 5 and over based on a census question concerning residence 5 years ago. It is assumed that babies migrate along with their mothers. To estimate migration for age 0, we applied the child women ratio to female migrants of child- bearing age. This was done for domestic in-migration and out-migration separately. First, the child women ratio of age 0 to women 15-44 is used to measure the proportion of child-bearing female with children under age 1. The child/women ratios were applied to the domestic in - migration and out-migration for female age 15-44 to get the migration for age 0 in five years. The results were divided by 5 to approximate annual migration.

3. Estimates of international migration:

The international migration is measured by the foreign born population entering the U.S. in the past five years for age 0 prior to census 2000. It is not necessary to divide the foreign born population age 0 by 5 to approximate annual migration as for other age groups because the migration of age 0 only occurs in one year prior to census date.

4. Estimates by Race and Hispanic origin:

The major difference of race classification between Census 2000 and NCHS is the absence of the multi-race group in NCHS vital statistics. To evaluate the census coverage rates for age 0 by race, we need to have consistent race groups between NCHS and census for comparison. Thus, the census 2000 population by race and Hispanic origin was converted (bridged) to the old race groups using a "fraction assignment " procedure to distribute the multiple race population to each of the 4 groups - White, Black, American Indian, and Asian or Other Pacific Islanders.

5. Bridging Census 2000 race groups to 1990 Census race groups:

The race assignment procedure which was used to convert the Census 2000 groups to approximate the 1990 race classification was called the "fractional assignment" rule. A person is assigned to a factional identity depending on the number of races they reported. For example, if the population reported two races as White and Black, one half is added to White and one half is added to Black. If the population reported three races as White, Black, and Asian, one-third is added to White, one-third is added to Black, one-third is added to Asian, and so on.

MEASUREMENT OF INTERNAL MIGRATION FOR THE CENSUS BUREAU'S STATE POPULATION PROJECTIONS BY AGE, SEX, AND RACE

by

Caribert Irazi and Ching-li Wang
Population Division, U.S. Census Bureau

Abstract

The Census Bureau uses the IRS-extracted individual income tax returns to estimate migration flows between counties and between states (50 states and the District of Columbia). In the Census Bureau's state population projections, the time series of IRS migration rates for the states are used to project migration flows. The migration rates by age, sex, and race derived from census data are then used to disaggregate the projected IRS migration flows into age, sex, and race detail. In the past, the IRS and census migration rates were developed and projected for all 2,550 state-to-state migration flows, and then they were aggregated to state totals. This paper examines the data issues in these two data sources and shows that there was a very high percentage of empty data cells in the census migration flows and many small migration streams in the state-to-state migration by age, sex, and race. This problem becomes even more prevalent with expanded race/ethnic group details - White alone, Black alone, American Indian alone, Asian alone, Hawaiian alone, and multi-race by Hispanic origin for each of 2,550 flows in the census 2000. This paper also presents alternative approaches to projecting the internal migration by age, sex, and race. It was found that the use of region-to-state and state-to region migration flows, the use of 5-years age grouping for migration rates by race detail, and combining small race groups to develop migration rates for projections reduce the empty cells substantially. The projections of IRS migration rates for 204 region-to-state and 204 state-to-region migration flows provide a better fit than the use of 2,550 state-to-state migration flows.

I. Introduction

The Census Bureau prepares population projections by single years of age, sex, race, and Hispanic origin. The cohort component method is used in the projections for 50 states and the District of Columbia. Each component of population change – births, deaths, internal migration and international migration was projected separately. The migration component is the most difficult part of the projections. Unlike other components, migration is defined in terms of space as well as time. There are many sources of data to measure the movement of population. Among them, the Census Bureau has used an extract of individual income tax returns from the Internal Revenue Service (IRS) to derive migration flows between states and counties on an annual basis since 1975. In addition, the decennial censuses also provide migration flows with detailed age, sex, race, and ethnicity, and other social and economic variables. The two data sources have been considered the best information so far to measure migration not only with extended time series, but also with greater demographic details.

In the previous state projections series, the migration flows used for projections contain 2,550 state-to-state flows (51x 50) derived from IRS and Census data (PPL-47, 1996). The 2,550 state-to-state migration flows rates were projected separately, 50 flows for every state. Then the census migration by single year of age, sex, race and Hispanic origin for each of the 2,550 flows was used to dis-aggregate the projected IRS migration flows into detailed age, sex, race and Hispanic origin groups. However, the evaluation of the last state projections series indicates that the percent errors in domestic internal migration remained very high for most of the states (Wang 2002). Among 4.9 million data cells of single year of age (0 to 85 and over), sex, and race/Hispanic origin in 2,550 migration flows, there were a lot of small or empty data cells. The data in such detail may have contributed to significant large errors in addition to other data issues in the two data sources such as census and IRS coverage and data quality.

The purpose of this paper is to examine the data issues in the two data sources and present alternative approaches to projecting the internal migration by age, sex, and race detail.

II. Data Issues for Internal Migration

There are many data sources to measure migration. These include 1) the decennial census, 2) the Internal Revenue Service administrative records, 3) the Social Security Administration records, 4) Census Bureau population estimates (migration component), 5) the American Community Survey (ACS), and 6) the Current Population Survey (CPS). Each provides different measures of internal migration. However, the state-to-state migration is considered the best choice in measuring internal migration between states, capturing the movement between specific places of origin and places of destination. To date, the best data sources for state-to-state migration are the census migration data and IRS migration flows data (Isserman et al, 1982).

A. Census migration data

The decennial census includes a question on place of residence five years prior to the census date. Comparison of current and previous residence provides basic information on the volume of mobility, including movements between counties, states, and regions. The interstate migration data on residence five years ago have been collected in each census since 1960. The Census Bureau tabulates the demographic characteristics of migrants, including single year of age, sex, and race/ethnic group details. With such detailed information, these data are ideal when population projections are made using the cohort component method that requires such details.

(a). General Data Issues:

However, there are some disadvantages. The decennial census asks respondents to report the place of residence 5 years before the census date; therefore migration movements of children under age five are not available from the migration question. Estimates of migration for children under 5 were modeled on the migration patterns of children aged 5 to 9 years. Comparing places of residence in 5-year time spans to derive migration data cannot capture the multiple moves in between. Shifts in location with intermediate moves are ignored. For example, movements of individuals who resided in more than two states during the 5 years period are not available from the census data. The migration data are derived from samples in the census's long form. Both sampling and other non-sampling errors are involved.

The information obtained from the decennial census represents a snapshot of the population every 10 years, and hence does not provide insight about trends in migration patterns on a continuing basis. Moreover, detailed data from one decennial census are not always comparable to the data from another census because definitions change. As a result, it may be difficult to derive patterns of migration necessary to forecast migration. For instance, the race/ethnic categories in census 2000 are not the same as in census 1990, making difficult a comparison of migration by race or ethnicity between the two censuses.

Although census migration data are available in details that are ideal for state population projections using the cohort component method, evaluation of the data suggest that there are errors in the reporting of age. One way to illustrate this issue is to examine the sex ratio at young ages. Assuming that children of both sexes have similar migration patterns, we would expect to have sex ratios around 100. Analysis of the census 2000 data indicates that this is not always the case. For example, the sex ratio of non-Hispanic White children migrating from the Midwest to Indiana increases from 69 at age 5 to 109 at age 6 and to 319 at age 7; it decreases to 121 at age 8 but rises to 196 at age 9 then declines to 72 at age 10. The sex ratio for non-Hispanic Black children migrating from the Northeast to Alabama also fluctuates greatly by age; it increases from 87 at age 5 to 383 at age 6, then declines to 58 at age 7 before increasing at 167 at age 8. Fluctuations in the sex ratio can also be observed for non-Hispanic Asian children migrating from the West to Rhode Island. In addition to sampling errors, these unusual sex ratios may be the result of age misreporting although undercount of children of either sex cannot be ruled out.

(b). Issue of Empty Data Cells

Perhaps the most serious data issue in using the census migration by demographic detail for state projections is the empty data cells in the state-to-state migration flows. Since we will prepare the projections by single years of age (0 to 85+), sex, race, and Hispanic origin for each state-to-state migration flow, there are 2,064 (86*2*6*2) data cells. There are six race groups to be projected include White alone, Black alone, American Indian and Alaska Native alone, Asian alone, Native Hawaiian and other Pacific Islanders alone, and the group reporting more than one race. The six race groups are separated by non-Hispanic and Hispanic origin, a total of 12 race and Hispanic groups. With male and female for each race and Hispanic group, there are 24 sex and race/Hispanic groups. The 24 groups are multiplied by 81 groups of single year of age (5 to 85 and over) in each for each of 2,550 migration flows. There are a lot of empty data cells when the volume

Table 1. Percent Empty Data Cells in Migration Flows by Age, Sex, Race and Hispanic Origin: 1995-2000

Sex, Race & Hispanic		State-to-State Flows		Region-to-State Flows		State-to-Region Flows	
		Single Years of Age (1)	Five Years of Age (2)	Single Years of Age (1)	Five Years of Age (2)	Single Years of Age (1)	Five Years of Age (2)
Total Migration Flows		2550 Flows		204 Flows		204 Flows	
All Race and Hispanic Group		88.6	79.4	59.1	48.4	60.1	50.3
NonHispanic							
White	Male	34.5	15.8	2.5	0.3	3.2	0.8
	Female	33.4	13.5	1.9	0.2	2.3	0.2
Black	Male	78.0	60.2	29.4	17.0	31.3	18.5
	Female	78.2	60.0	28.6	15.6	30.8	16.5
American Indian	Male	94.7	84.1	57.6	40.4	58.2	41.3
	Female	94.6	83.7	57.4	39.5	57.5	39.1
Asian	Male	88.2	72.9	39.6	24.5	40.4	24.4
	Female	87.7	71.5	37.9	21.8	38.5	21.4
Hawaiian/PCI	Male	99.0	96.8	85.5	76.7	86.1	78.6
	Female	99.1	96.8	85.6	76.4	85.9	78.1
Multi-Race	Male	91.5	76.9	47.3	30.2	48.5	31.9
	Female	91.1	76.0	45.4	27.5	46.6	30.2
Hispanic							
White	Male	83.0	66.3	32.1	19.8	36.0	23.2
	Female	84.1	67.3	32.9	18.4	36.3	21.2
Black	Male	97.7	93.0	73.4	60.8	75.1	65.2
	Female	97.7	92.8	72.9	60.5	74.5	64.4
American Indian	Male	98.8	95.9	80.7	71.5	81.6	74.7
	Female	99.0	96.4	81.7	73.1	82.5	76.2
Asian	Male	99.5	98.1	88.8	82.6	89.5	85.8
	Female	99.5	98.2	89.1	83.9	89.5	86.1
Hawaiian/PCI	Male	99.8	99.0	92.7	90.0	93.1	91.3
	Female	99.8	99.1	93.1	90.4	93.3	92.2
Multi-Race	Male	98.8	95.7	81.1	70.4	81.8	72.8
	Female	98.7	95.5	80.5	69.3	81.2	72.1

Notes: 1 - Single years of age 5 ~ 85 and over, 2 - Five years of age 5 ~ 85 and over.
There are 1,944 cells each flows for single year of age, 408 cells each flow for 5-year age grouping.
Source: U.S. Census Bureau, Census 2000 Migration File

of interstate migrants is disaggregated in such detail. As Table 1 shows, the proportion of empty cells for all the states in the 2,550 migration flows reaches 88.6 percent. That is, almost 90 percent of cells are empty when state-to-state migrants are dis-aggregated by the age, sex, and race detail.

The age pattern of migrants could be one possible explanation for the high proportion of empty cells, individuals in some age groups having very low propensity to migrate from one state to another. After age 60, the number of migrants is significantly reduced. Another possible explanation is that some states do not send or receive significant numbers of migrants. As table 2 shows, the proportion of empty cells is greater than 90 percent for 18 states. Most of them are small states. Third, the race/ethnic factor plays an important role in explaining the high proportion of empty cells. In general, the smaller the race/ethnic group, the greater the proportion of empty cells. The lowest proportion of 33 percent is observed for non-Hispanic Whites who represent the largest percentage of the US population, and the highest proportion of empty cells of 99.8 percent is observed for

Table 2. Percent Empty Data Cells in Census Migration Flows by Age, Sex, and Race by State: 1995-2000

State	State-to-State Flows		Region-to-State Flows		State-to-Region Flows	
	Single Years of Age (1)	Five Years of Age (2)	Single Years of Age (1)	Five Years of Age (2)	Single Years of Age (1)	Five Years of Age (2)
Total	**88.6**	**79.4**	**59.1**	**48.4**	**60.1**	**50.3**
Alabama	90.0	81.5	61.6	50.2	64.5	54.0
Alaska	92.0	83.9	68.0	56.0	66.8	56.1
Arizona	85.3	73.6	47.7	35.4	50.8	39.3
Arkansas	91.1	83.4	60.5	49.7	65.6	56.6
California	68.0	50.6	33.7	23.2	26.1	17.0
Colorado	85.5	73.6	50.0	38.1	50.9	40.3
Connecticut	89.6	80.7	60.6	50.1	58.6	49.8
Delaware	95.2	89.3	72.2	64.1	74.4	66.5
District of Columbia	93.9	86.8	74.4	64.9	70.6	57.4
Florida	78.8	64.7	35.4	25.3	39.1	26.7
Georgia	85.2	74.5	47.7	35.5	54.4	43.0
Hawaii	89.0	77.4	58.5	46.4	54.9	41.6
Idaho	93.4	87.0	73.1	63.2	73.0	64.3
Illinois	81.0	69.5	47.6	39.6	44.5	33.0
Indiana	88.4	79.3	57.2	47.2	59.0	50.9
Iowa	91.4	83.6	64.7	54.9	66.7	57.5
Kansas	89.4	80.3	57.4	47.3	59.8	51.6
Kentucky	90.8	82.4	61.9	49.8	64.8	54.0
Louisiana	89.1	79.6	61.0	51.0	61.5	50.7
Maine	94.9	89.7	75.5	65.3	76.6	68.6
Maryland	86.2	75.8	54.7	43.3	55.3	44.9
Massachusetts	87.8	77.6	55.4	44.1	53.5	44.7
Michigan	85.7	75.3	48.9	41.3	51.7	43.9
Minnesota	89.4	80.3	54.6	43.1	60.0	52.3
Mississippi	91.6	83.7	66.6	55.6	69.4	59.6
Missouri	87.6	77.7	52.4	41.8	56.7	47.3
Montana	93.8	87.7	74.7	64.6	75.4	65.8
Nebraska	92.0	83.9	63.9	53.1	65.5	56.4
Nevada	89.6	79.8	54.2	40.8	59.7	46.8
New Hampshire	94.7	89.4	76.3	65.6	76.6	66.3
New Jersey	84.2	73.0	51.4	41.2	46.8	36.4
New Mexico	89.3	79.1	59.4	46.1	58.3	45.9
New York	77.0	62.9	41.5	33.5	35.0	24.3
North Carolina	85.6	74.8	47.5	34.3	53.8	43.0
North Dakota	95.1	89.6	77.4	68.8	77.6	69.9
Ohio	85.6	75.4	50.3	42.0	52.1	46.0
Oklahoma	88.2	77.6	52.5	42.2	56.6	45.7
Oregon	89.9	80.9	60.3	48.3	61.6	49.8
Pennsylvania	84.7	74.1	50.0	38.2	49.2	40.1
Rhode Island	95.0	89.0	72.2	63.4	70.5	62.4
South Carolina	89.9	81.6	59.7	47.5	64.7	54.1
South Dakota	94.6	88.6	74.4	65.3	75.5	66.3
Tennessee	88.2	78.9	57.1	44.5	61.6	50.4
Texas	76.8	62.2	35.9	26.4	36.6	27.1
Utah	91.6	83.5	65.9	53.7	65.8	55.0
Vermont	96.2	92.1	79.6	69.8	80.9	72.5
Virginia	83.9	72.8	48.3	36.8	52.4	41.1
Washington	84.6	72.5	50.7	36.2	51.3	41.0
West Virginia	94.3	88.5	73.2	62.7	74.9	65.6
Wisconsin	89.3	80.5	57.4	47.3	59.3	52.3
Wyoming	94.5	89.1	77.5	67.9	76.6	68.0

Notes: 1 - Single years of age 5 ~ 85 and over, 2 - Five years of age 5 ~ 85 and over.
Source: U.S. Census Bureau, Census 2000 Migration File.

Table 3: The Coverage of 1999-2000 Total Exemptions and Matched IRS Returns for Migration Flows As Compared with Census 2000 Population

State	Census 2000 Population	Estimated 2000 Total Exemptions	Estimated Matched Exemptions	Exemptions as Percent of Population	Matched as Percent of Population	Difference between Matched & Total
Total	**281,421,906**	**235,989,792**	**219,647,295**	**83.9**	**78.0**	**-5.8**
Alabama	4,447,100	3,718,104	3,469,162	83.6	78.0	-5.6
Alaska	626,932	527,821	484,915	84.2	77.3	-6.8
Arizona	5,130,632	4,033,038	3,694,680	78.6	72.0	-6.6
Arkansas	2,673,400	2,176,831	2,017,666	81.4	75.5	-6.0
California	33,871,648	27,633,666	25,257,872	81.6	74.6	-7.0
Colorado	4,301,261	3,625,130	3,361,479	84.3	78.2	-6.1
Connecticut	3,405,565	2,913,090	2,743,772	85.5	80.6	-5.0
Delaware	783,600	677,355	633,007	86.4	80.8	-5.7
District of Columbia	572,059	430,091	385,747	75.2	67.4	-7.8
Florida	15,982,378	13,002,641	11,866,860	81.4	74.2	-7.1
Georgia	8,186,453	6,787,201	6,222,374	82.9	76.0	-6.9
Hawaii	1,211,537	1,014,535	945,226	83.7	78.0	-5.7
Idaho	1,293,953	1,105,306	1,031,653	85.4	79.7	-5.7
Illinois	12,419,293	10,668,603	10,027,829	85.9	80.7	-5.2
Indiana	6,080,485	5,331,330	5,044,809	87.7	83.0	-4.7
Iowa	2,926,324	2,549,914	2,443,365	87.1	83.5	-3.6
Kansas	2,688,418	2,312,138	2,183,844	86.0	81.2	-4.8
Kentucky	4,041,769	3,320,818	3,117,942	82.2	77.1	-5.0
Louisiana	4,468,976	3,697,146	3,430,187	82.7	76.8	-6.0
Maine	1,274,923	1,089,094	1,026,343	85.4	80.5	-4.9
Maryland	5,296,486	4,594,593	4,288,784	86.7	81.0	-5.8
Massachusetts	6,349,097	5,322,934	5,017,214	83.8	79.0	-4.8
Michigan	9,938,444	8,496,997	7,995,699	85.5	80.5	-5.0
Minnesota	4,919,479	4,326,440	4,103,321	87.9	83.4	-4.5
Mississippi	2,844,658	2,343,420	2,175,034	82.4	76.5	-5.9
Missouri	5,595,211	4,780,142	4,493,362	85.4	80.3	-5.1
Montana	902,195	763,232	714,064	84.6	79.1	-5.4
Nebraska	1,711,263	1,512,832	1,444,584	88.4	84.4	-4.0
Nevada	1,998,257	1,664,632	1,501,556	83.3	75.1	-8.2
New Hampshire	1,235,786	1,105,897	1,044,589	89.5	84.5	-5.0
New Jersey	8,414,350	7,266,350	6,795,772	86.4	80.8	-5.6
New Mexico	1,819,046	1,513,711	1,400,987	83.2	77.0	-6.2
New York	18,976,457	15,171,991	14,079,335	80.0	74.2	-5.8
North Carolina	8,049,313	6,752,092	6,280,193	83.9	78.0	-5.9
North Dakota	642,200	563,295	543,683	87.7	84.7	-3.1
Ohio	11,353,140	9,890,512	9,320,355	87.1	82.1	-5.0
Oklahoma	3,450,654	2,786,102	2,581,179	80.7	74.8	-5.9
Oregon	3,421,399	2,815,542	2,603,217	82.3	76.1	-6.2
Pennsylvania	12,281,054	10,520,257	9,960,014	85.7	81.1	-4.6
Rhode Island	1,048,319	847,632	794,268	80.9	75.8	-5.1
South Carolina	4,012,012	3,361,418	3,134,501	83.8	78.1	-5.7
South Dakota	754,844	659,976	628,208	87.4	83.2	-4.2
Tennessee	5,689,283	4,826,918	4,512,884	84.8	79.3	-5.5
Texas	20,851,820	17,347,445	15,936,178	83.2	76.4	-6.8
Utah	2,233,169	1,917,903	1,789,898	85.9	80.2	-5.7
Vermont	608,827	529,727	501,304	87.0	82.3	-4.7
Virginia	7,078,515	6,048,658	5,643,655	85.5	79.7	-5.7
Washington	5,894,121	5,017,872	4,657,117	85.1	79.0	-6.1
West Virginia	1,808,344	1,455,249	1,374,147	80.5	76.0	-4.5
Wisconsin	5,363,675	4,737,179	4,531,978	88.3	84.5	-3.8
Wyoming	493,782	434,992	411,497	88.1	83.3	-4.8

Source: U.S. Census Bureau, Special Tabulation from Administrative Records and Methodology Research Branch, Poulation Division

Table 4. The Number and Percent of IRS State-to-State Migration Flows by Selected Size: 1999-2000

State	State-to-State In-Migration Flows						State-to-State Out-Migration Flows					
	Less Than 250		Less Than 500		Less Than 1,000		Less Than 250		Less Than 500		Less Than 1,000	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Total	**567**	**22.2**	**958**	**37.6**	**1392**	**54.6**	**657**	**25.8**	**1081**	**42.4**	**1540**	**60.4**
Alabama	11	22.0	21	42.0	27	54.0	11	22.0	22	44.0	31	62.0
Alaska	14	28.0	30	60.0	41	82.0	15	30.0	33	66.0	42	84.0
Arizona	1	2.0	5	10.0	9	18.0	4	8.0	8	16.0	19	38.0
Arkansas	14	28.0	19	38.0	31	62.0	18	36.0	26	52.0	35	70.0
California	0	0.0	0	0.0	4	8.0	0	0.0	0	0.0	4	8.0
Colorado	1	2.0	5	10.0	8	16.0	3	6.0	5	10.0	13	26.0
Connecticut	12	24.0	23	46.0	30	60.0	16	32.0	25	50.0	31	62.0
Delaware	33	66.0	40	80.0	44	88.0	34	68.0	40	80.0	45	90.0
District of Columbia	32	64.0	37	74.0	44	88.0	36	72.0	40	80.0	45	90.0
Florida	0	0.0	0	0.0	6	12.0	0	0.0	3	6.0	8	16.0
Georgia	0	0.0	6	12.0	15	30.0	4	8.0	10	20.0	19	38.0
Hawaii	14	28.0	24	48.0	37	74.0	15	30.0	27	54.0	37	74.0
Idaho	15	30.0	33	66.0	40	80.0	22	44.0	37	74.0	41	82.0
Illinois	1	2.0	11	22.0	14	28.0	1	2.0	10	20.0	15	30.0
Indiana	9	18.0	13	26.0	22	44.0	12	24.0	16	32.0	28	56.0
Iowa	11	22.0	24	48.0	35	70.0	11	22.0	24	48.0	34	68.0
Kansas	8	16.0	16	32.0	28	56.0	9	18.0	18	36.0	30	60.0
Kentucky	11	22.0	18	36.0	25	50.0	13	26.0	22	44.0	32	64.0
Louisiana	11	22.0	21	42.0	29	58.0	12	24.0	22	44.0	30	60.0
Maine	24	48.0	31	62.0	41	82.0	27	54.0	39	78.0	44	88.0
Maryland	5	10.0	10	20.0	22	44.0	7	14.0	14	28.0	28	56.0
Massachusetts	7	14.0	15	30.0	26	52.0	11	22.0	20	40.0	28	56.0
Michigan	3	6.0	12	24.0	18	36.0	5	10.0	14	28.0	22	44.0
Minnesota	7	14.0	12	24.0	25	50.0	7	14.0	17	34.0	28	56.0
Mississippi	15	30.0	25	50.0	36	72.0	18	36.0	25	50.0	37	74.0
Missouri	6	12.0	10	20.0	20	40.0	7	14.0	15	30.0	21	42.0
Montana	19	38.0	33	66.0	40	80.0	21	42.0	35	70.0	42	84.0
Nebraska	13	26.0	26	52.0	38	76.0	14	28.0	27	54.0	39	78.0
Nevada	6	12.0	13	26.0	24	48.0	8	16.0	21	42.0	36	72.0
New Hampshire	23	46.0	31	62.0	38	76.0	28	56.0	35	70.0	43	86.0
New Jersey	6	12.0	18	36.0	28	56.0	7	14.0	16	32.0	28	56.0
New Mexico	9	18.0	20	40.0	37	74.0	9	18.0	24	48.0	39	78.0
New York	3	6.0	5	10.0	12	24.0	3	6.0	7	14.0	14	28.0
North Carolina	1	2.0	5	10.0	13	26.0	3	6.0	7	14.0	18	36.0
North Dakota	35	70.0	43	86.0	47	94.0	31	62.0	39	78.0	48	96.0
Ohio	1	2.0	10	20.0	17	34.0	2	4.0	11	22.0	20	40.0
Oklahoma	8	16.0	14	28.0	30	60.0	9	18.0	17	34.0	29	58.0
Oregon	7	14.0	14	28.0	26	52.0	10	20.0	21	42.0	36	72.0
Pennsylvania	3	6.0	6	12.0	18	36.0	3	6.0	9	18.0	20	40.0
Rhode Island	29	58.0	39	78.0	44	88.0	31	62.0	38	76.0	44	88.0
South Carolina	7	14.0	17	34.0	24	48.0	8	16.0	20	40.0	31	62.0
South Dakota	26	52.0	37	74.0	42	84.0	26	52.0	38	76.0	45	90.0
Tennessee	6	12.0	12	24.0	20	40.0	7	14.0	16	32.0	25	50.0
Texas	0	0.0	1	2.0	5	10.0	0	0.0	1	2.0	7	14.0
Utah	10	20.0	19	38.0	34	68.0	12	24.0	21	42.0	36	72.0
Vermont	31	62.0	39	78.0	45	90.0	36	72.0	41	82.0	46	92.0
Virginia	0	0.0	4	8.0	7	14.0	3	6.0	6	12.0	16	32.0
Washington	3	6.0	7	14.0	12	24.0	3	6.0	7	14.0	18	36.0
West Virginia	28	56.0	34	68.0	41	82.0	30	60.0	35	70.0	39	78.0
Wisconsin	6	12.0	12	24.0	30	60.0	9	18.0	19	38.0	31	62.0
Wyoming	22	44.0	38	76.0	43	86.0	26	52.0	38	76.0	43	86.0

Source: U.S. Census Bureau, Special Tabulation from Administrative Records and Methodology Research Branch, Poulation Division

Hispanic Native Hawaiians and Pacific Islanders who are among the smallest groups in the nation. The high proportions observed for some race/ethnic groups may be related to the fact that interstate migrants from small race/ethnic groups do not represent numbers that are large enough to be disaggregated by age, sex, and race/ethnic details. For example, there were only 10 migrants of the Native Hawaiian and Pacific Islander Hispanic group who migrated to or from the state of North Dakota. It is obvious that the age detail, size of the states, and size of the racial groups affect the number and proportion of empty data cells.

B. IRS migration flows data

The Census Bureau uses extracts of IRS individual income tax returns to derive migration flows between counties and between states (50 states and the District of Columbia). To derive the migration flows, the unique IDs (based on social security numbers, which along with names are erased for confidentiality) of the individual income returns are used to match the returns filed in two consecutive years. Then the addresses are compared to identify whether the returns changed their addresses or not. Through this process, the Census Bureau can determine the migration status of the tax filers and household members in the returns. The data are aggregated to the county level and then to the state level.

(a). General Issues:

Despite the extended time series of migration data from the IRS, the matched returns from which the migration flows are derived only cover about 80 percent of the population. As Table 3 shows, the exemptions on the matched returns accounted for 78 percent of total U.S. population. The percent covered by the IRS migration data varied from state to state, ranging from 67.1 percent in the District of Columbia to 84.5 percent in Wisconsin and New Hampshire. For the states of Arizona, California, Florida, New York, Oklahoma, Nevada, and the District of Columbia, the matched filing rate has been around or below 75 percent. On the other end of the spectrum, Nebraska, New Hampshire, North Dakota, and Wisconsin are states where the rate has been around 85 percent.

Table 3 also shows the total population covered by the IRS data, which include those whose returns are matched in two consecutive years, and those whose returns are not matched. Those who are not matched are the ones who did not file returns in either of the two years due to changes in the income levels required to file returns or those who start to have income, or those whose incomes were reduced. Approximately 5 to 6 percent of the returns were not matched. (See Table 3)

In addition to the coverage issue, migration data derived from the IRS income tax returns are also limited because of inaccuracies in filing the returns. For example, college students attending school in other states reported as dependents by their parents are not captured as migrants in states where colleges are located. Although the migration data derived by processing the IRS income tax returns are highly reliable, details by age, sex, and race/ethnic groups are not available to be used for state population projections.

(b). Size of migration flows issue

As noted before, one of the key issues related to the 2,550 state-to-state migration flows is the size of flows in many small states. Table 4 shows the number of IRS state-to-state migration flows by selected size (less than 250, less than 500, and less than 1,000). Among 2,550 state-to-state in flows, 22.2 percent are less than 250 migrants, 37.6 percent are less than 500 migrants, and 54.6 percent are less than 1,000 migrants. Among out-migration flows, 25.8 percent are less than 250, 42.4 percent are less than 500, and 60.4 percent are less than 1,000. Eleven states had more than one third of their flows at less than 250 migrants. Delaware, the District of Columbia, Maine, New Hampshire, North Dakota, Rhode Island, South Dakota, Vermont, and West Virginia had more than one half of their migration flows at less than 250.

The size of migration flows may not be a serious problem, but when the migrants are disaggregated by detailed age, sex, and race for 2,064 cells, the irregularity of age, sex, and race distributions emerges.

III. Alternative Approach to the Use of Migration Data

As mentioned before, the major factors which create the empty cells are size of states, the age detail, and racial detail. Therefore, we propose to aggregate the 2,550 migration flows into region-to-state flows as in-migration and state-to-region flows as out-migration. The four regions are Northeast, Midwest, South, and West. The Northeast region includes Maine, Vermont, New Hampshire, Massachusetts, Rhode Island, Connecticut, New York, New Jersey, and Pennsylvania. The Midwest region includes Wisconsin, Michigan, Illinois, Indiana, North Dakota, and South Dakota. The South region includes Delaware, District of Columbia, Florida, Georgia, Maryland, North Carolina, South Carolina, Virginia, West Virginia, Alabama, Kentucky, Mississippi, Tennessee,

Table 5. Percent Empty Data Cells in Migration Flows with Regrouped Race Groups: 1995-2000

Total Migration Flows			Region-to State 5-year age groups 204 Flows	State-to-Region 5-year age groups 204 Flows
All Race and Hispanic Goups			23.0	23.9
I. Individual Race Groups				
NonHispanic				
	(1) White	Male	0.3	0.8
		Female	0.2	0.2
	(2) Black	Male	17.0	18.5
		Female	15.6	16.5
	(3) American Indian	Male	40.4	41.3
		Female	39.4	39.1
	(4) Asian	Male	24.5	24.4
		Female	21.8	21.4
	(5) Multi Races	Male	30.2	31.9
		Female	27.5	30.2
Hispanic	(6) White	Male	19.8	23.2
		Female	18.4	21.2
II. Combined NonHispanic Group				
	(7)Asian/Nhopi	Male	23.7	23.8
		Female	21.3	21.0
III. Hispanic and NonHispanic Combined				
	(8) Black	Male	16.6	18.0
		Female	13.5	14.7
	(9) American Indian	Male	37.9	39.6
		Female	37.7	37.5
	(10) Asian	Male	24.0	24.3
		Female	21.7	21.3
	(11) NonHispanic Asian/Hhpi	Male	23.2	23.6
		Female	21.1	20.9
	(12) Multi-Races	Male	29.5	31.3
		Female	26.6	29.4

Source: U.S. Census Bureau, Census 2000 Migration File.

Arkansas, Louisiana, Oklahoma, and Texas. The West region includes Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, Wyoming, Alaska, California, Hawaii, Oregon, and Washington. A total of 408 migration flows will be projected instead of 2,550 migration flows. Second, the census migration rates are developed for 5 year' age groups (5-9 , 10-14, ... ,85 and over) instead of single year of age to minimize the impact of empty cells on developing appropriate migration rates. Third, we will use a "residual method" to project migration by race detail in which we develop rates for only a few major race groups and combination of major and smaller race groups. In the final stages of projections of internal migration, the projections for the combined groups minus the major groups will be used to derive the projections for the small race groups. This procedure will minimize the impact of small race groups on developing appropriate migration rates by race.

(A). Region-to-State and State-to-Region Migration Flows

There are 204 region-to-state migration flows (51 states times 4 regions), and 204 state-to-region migration flows. The purpose of using the state-to-state migration flow is to capture the changes of major migration streams between the state of origin and state of destination. Using the region-to-state and state-to-region migration flows, the major migration streams are reflected in the region flows because they would contribute a heavier weight in determining the overall migration for the states.

As Table 1 shows, the empty data cells in the 2,550 census migration flows by age, sex, and race was reduced from 88.6 percent to 79.4 percent if we use the 5-year age grouping. If we use the region-to-state and state-to-region migration flows too, the empty data cells were reduced to 48.0 percent for in-migration flows, and to 50.3 percent for out-migration flows. If we regroup the race groups to calculate migration rates, the empty cells were reduced dramatically to 23.0 percent for in-migration flows, and 23.9 percent for out-migration flows as shown in Table 5. After regrouping races, all the state-to-region and region-to-state migration flows are greater than 1,000, except North Dakota, South Dakota, and Wyoming. These procedures minimize the impact of size of state, age detail, and race detail on creating unreliable migration rates by age, sex, and race.

(B). Procedures for projection of migration by race:

The "residual method" will work like this. First we prepare the migration rates for major race groups and combinations of major groups and minor groups. Then projected

migration for the combined race groups minus the major groups will be used to derive minor groups. The procedures for projecting migration by race are as follows.

a. Calculate migration rates for the following race alone and combined race groups:

(1). Non-Hispanic White*
(2). Non-Hispanic Black*
(3). Non-Hispanic AIAN
(4). Non-Hispanic Asian*
(5). Non-Hispanic multi-race*
(6). Hispanic White*

(7). Non-Hispanic Asian/NHPI combined
(8). Hispanic and non-Hispanic Black combined
(9). Hispanic and non-Hispanic American Indian combined
(10). Hispanic and non-Hispanic Asian combined
(11). Hispanic and non-Hispanic Asian/NHPI combined
(12). Hispanic and non-Hispanic multi-race combined

* Stand alone categories; AIAN represents American Indian and Alaska Native, and NHPI represents Native Hawaiian and Other Pacific Islander.

b. Subtract the projected number of migrants of major groups from combined groups to derive the projected number of migrants for other race groups; for example, the non-Hispanic Asian/Pacific Islander combined group minus non-Hispanic Asian alone to derive non-Hispanic Hawaiian and other Pacific Islander as (1) below; Hispanic and non-Hispanic Asian combined group minus non-Hispanic Asian to derive Hispanic Asian as (5) below.

(1). Non-Hispanic NHPI = a(7) - a(4)
(2). Hispanic and non-Hispanic NHPI = a(11) - a(10)
(3). Hispanic Black = a(8) - a(2)
(4). Hispanic American Indian = a(9) - a(3)
(5). Hispanic Asian = a(10) - a(4)
(6). Hispanic multi-race = a(12)- a(5)
(7). Hispanic NHPI = b(2) - b(1)

c. Disaggregation of 5-year age group migration into single years of age:

Although the residual method can be used to derive certain projected migration for certain race/Hispanic groups, some race groups in many states are very small. It is impractical to develop and project migration rates for single years of age. Therefore, only 5-year age group migration rates for the race groups listed in section (a) above were prepared. Then the Karup King procedure was used to convert the projected 5-year age group migration to single year of age in the final stage of projecting migration.

Table 6. The Mean Absolute Precentage Error (MAPE) Statistics from the Estimated Model For Selected Migration Flows: 1975-2000

Migration Flows	Stepwise Autoregressive model	Quadratic Time trend model
State-to-State		
Arkansas to California	7.27	18.94
Delaware to California	9.90	46.31
South Carolina to California	6.25	50.75
West Virginia to California	7.38	42.29
South Region to California	4.34	25.12
State-to-State		
Illinois to Georgia	15.65	18.07
Iowa to Georgia	19.80	11.50
Kansas to Georgia	46.47	8.45
Ohio to Georgia	43.51	15.87
Midwest Region to Georgia	10.00	8.49
State-to-State		
Maine to Nebraska	22.83	24.68
New York to Nebraska	15.79	19.12
Rhode Island to Nebraska	7.41	9.07
Vermont to Nebraska	29.89	33.83
Northeast Region to Nebraska	5.51	9.36
State-to-State		
Alaska to Vermont	5.65	9.82
Arizona to Vermont	9.59	17.02
California to Vermont	8.57	10.35
Hawaii to Vermont	7.19	12.30
West Region to Vermont	4.85	9.48

Source: U.S. Census Bureau, Population Division

(C). Evaluation of IRS time series when state-to-state migration rates are aggregated into region-to-state

Aggregating the census data by age groups, regions, and ethnic groups helps reduce the identified data issues. However, the census data refer to one point in time. This problem can, however, be overcome with the data from the IRS records that are available since 1975, and hence, offer the possibility of determining the trend in the level of migration.

Drawing from the observation that aggregating census data helps reduce the issues, we looked at the forecasting performance of series when state-to-state flows are grouped into region-to-state flows. For that purpose, time series data selected from all parts of the country and two models (the stepwise autoregressive and the quadratic time trend equations) were used to test the forecasting performance in the two cases. The mean absolute percentage error (MAPE) statistic was used as a diagnostic statistic.

Table 6 reports the MAPE statistics from the two models for the period 1975-2000. The overall picture is that grouping state-to-state migration flows into region-to-state movements yields more satisfactory diagnostic statistics as summarized by the MAPE for the stepwise autoregressive model. Results from the stepwise autoregressive model show that the MAPE has the lowest value when migration data are grouped by region for all the series selected. For instance, the MAPE value of 4.3 percent obtained for the migration flows from the South to California, is lower than all values of the statistic corresponding to the flows from each of the selected states to California, which vary from 7.3 percent to 9.9 percent. Results from the quadratic time trend equation do not indicate, however, that all region-to-state estimated models have the lowest MAPE values. For example, the value of 9.4 percent corresponding to the flow from the West to Vermont is higher than the MAPE value of 9.1 percent obtained from the migration series from California to Vermont. We attribute this to the fact that the quadratic time trend equation may not be a good model for the series. Moreover, in our effort to evaluate the forecasting performance of grouping state-to-state flows into region-to-state, we may not have followed all the rules required for building models designed to be used in forecasting series.

IV. Conclusions

Several deficiencies characterize the data sources used by the Census Bureau to derive internal migrations for state population projections by demographic detail. This paper has shown that many of these deficiencies can be overcome when the data sources are combined. The trend in the level of migration available from the IRS series can be combined with detailed demographic characteristics tabulated from the decennial census data. Additional steps need to be taken to ensure the quality of migration results used to develop the state population projections. These include aggregating the single year of age data into five-year age groups, combining race/ethnic groups, and measuring state-to-state migration flows as region-to-state movements. Regrouping the data reduces significantly the issue of empty cells, and corrects for errors in age reporting. Detailed state population projections can thereafter be derived using disaggregating procedures.

The use of procedures and techniques to group and disaggregate data or results has necessitated a number of assumptions concerning, for example, change in migration patterns through time and by age. Although these assumptions may be questionable, they were formulated based on data analysis and informed judgment. It remains to be seen if another set of assumptions would have led to better results in terms of reducing the identified data issues.

References

Campbell, Paul R., 1996, Population Projections for States by Age, Sex, Race, and Hispanic Origin: 1995 to 2025, U.S. Census Bureau, Population Division, PPL-47.

Eduardo E. Arriaga, 1994, "Population Analysis with Microcomputers" Volume II -- Software and Documentation. U.S. Census Bureau, International Population Center Program.

Frees, Edward W. 1992, "Forecasting State-to-State Migration Rates." Journal of Business & Economic Statistics, (April, 1992) Vol.10, No. 2, 153-167

Isserman, M. Andrew., A.D. Plane., B.D. McMillen. 1982. "Internal Migration in the United States: An Evaluation of Federal Data." Review of Public Data Use Vol.10, 285-311.

Lichter, T. Daniel, and F.G. De Jong. 1990. "The United States", in Charles Nam et al (eds), International Handbook of Internal Migration. Westport: Greenwood Press.

Wang, Ching-Li, 2002, "Evaluation of Census Bureau's 1995-2025 State Population Projections." U.S. Census Bureau, Population Division, Working Paper No. 67.

Business Cycles and Global Factors in Short-Range Forecasting

Chair: Jeff Busse, U.S. Geological Survey, U.S. Department of the Interior

An Input Output Study of the Distribution of Imports and Wages by Major Demand Category with Relevance to the 2000 to 2002 Period

Arthur Andreassen, Bureau of Labor Statistics, U.S. Department of Labor

Input-Output analysis is used to show the major components of Gross Domestic Product in a slightly different light. Input-Output analysis disaggregates the demand and supply sides of the economy and allows the measurement of the interaction of changes in one side with the other. This capability is used to allocate imports and wages by major component over the 2000 to 2002 period. Insights into some of the unique aspects of the path the economy followed to and through this downturn are gleaned.

Structural Change in the Global Soybean Market: Implications for Forecasting U.S. Commodity Prices Consistent with Forecasted U.S. Quantities

Gerald Plato and William Chambers, Economic Research Service, U.S Department of Agriculture

Major structural changes are occurring in the global soybean industry, which create problems with price forecasts. Increased South American production and growing demand are the major structural changes in the global soybean market. Our model found that the U.S. soybean stocks-to-use ratio and South American soybean production are the only quantity variables needed to forecast the implied U.S. soybean price. We estimate that each 1 percent increase in South American production reduces the U.S. soybean price by 0.52 percent and that each 1 percent increase in the U.S. soybean carryover stocks-to-use ratio reduces the U.S. soybean price by 0.41 percent.

Forecasting the Counter-Cyclical Payment Rate for U.S. Corn: An Application of the Futures Price Forecasting Model

Linwood A. Hoffman, Economic Research Service, U.S. Department of Agriculture

The 2002 Farm Act provides for counter-cyclical payment (CCPs) to owners of a qualifying base, when prices are low. Policy and budget analysts within USDA forecast counter-cyclical payments in an effort to estimate budget outlays for income safety net programs. The CCP is equal to the product of [(payment rate) x (payment acres) x (payment yield)]. Since both the payment acres and payment yield are predetermined, a model is presented that forecasts the counter-cyclical payment rate for U.S. corn. A payment rate is derived from a forecasted season-average corn price and predetermined policy parameters; target price, loan rate, and direct payment rate. The season-average corn price is provided by a model that relies on monthly futures prices, basis values (cash less futures), and marketing weights.

An Input Output Study of the Distribution of Imports and Wages by Major Demand Category with Relevance to the 2000 to 2002 Period.

Arthur Andreassen Bureau of Labor Statistics, U.S. Department of Labor

This article consists of two themes; the first is an exposition of some of the insights into Gross Domestic Product (GDP) that input output analysis provides. The second is the use of these insights to expand our understanding of the 2000 to 2002 downturn. Input output tables disaggregate the supply, the demand, the production and the income sides of the economy while numerating their interactions; this makes it a useful tool in the study of some underlying connections not obvious at an aggregate level. Specifically, this article calculates the somewhat unique response of personal consumption expenditures (PCE) to this downturn and the sources of the funds for this spending.

Introduction

By 2000 the economy was in its fifth year of healthy growth but it was starting to display fissures. In March the stock market peaked and in October industrial production began a monthly decline which would continue uninterrupted for the next fifteen month. By March, 2001 the economy was in a recession brought on mainly by a collapse in investment demand. The path followed upward from this nadir relied on extraordinary increases in consumer demand. This PCE increase depended on funding from a number of disparate sources: tax cuts, wages increasing in step with rising productivity, declines in the savings rate and rises in home values. Plunging mortgage rates encouraged increases in refinancing that in various combinations lowered monthly payments and/or allowed homeowners to "cash-out" some of their homes' appreciated value. Finally, the auto industry boosted sales with no interest loans. These positive influences more than compensated for the one very large negative - wage declines brought on by drops in investment and export demand. This study attempts to separate these sources of PCE, an exercise complicated by the circular interaction of wages and consumption, i.e., wages are a major source of consumption while consumption is an important source of wages. After this interaction is disentangled the true importance of the non-wage sources of consumption to the growth of GDP becomes apparent. Healthy economic growth depends on all the components of demand increasing, if wages are growing solely because consumers are tapping limited sources of funds these will eventually dry up. Wage growth is sustainable only with the added support from non-PCE demand categories.

Gross Domestic Product by Category

GDP and its components are shown in table 1, which displays the economic landscape of the period containing the March to November, 2001 recession. Growth during the present upturn has been lethargic, not anywhere near the 6 to 7% annual rate in nominal terms that one should expect from an economy functioning at its potential. (Nominal values are used throughout). Although total GDP grew in each of the years, not every demand component did so; investment and exports declined while PCE and Government grew. Residential construction, usually included in investment, is split out because it has more closely mirrored PCE during this period. Over these two years GDP increased by 6%, PCE grew by 9% and government by 13%. Although the government component is composed two thirds of State and local its total growth was split equally with the federal sector, this due to increases in spending on defense and homeland security. On the other hand, investment, exports and imports declined. Imports falling with the domestic economy and exports responding in its usual fashion as the rest of the world suffered a slow down in sync with our own. Residential construction, fueled by low mortgage rates and rising home values, partially replaced a ravaged stock market as a repository for investment funds. Because of its small relative size it will engender little further discussion.

In sum the economic path that the economy has since been following is obvious. GDP increased $621 billion from 2000 to 2002 while PCE rose $620 billion, on the other hand, investment fell $209 billion, Government climbed $222 billion and exports and imports offset each other. As a result of these varying growths, PCE increased its share of GDP by 1.9%, investment dropped 2.8% and Government went up 1.1%.

Demand Categories Adjusted for Imports

PCE, as is oft commented, represents over 2/3rds of GDP, a ratio that is really a misstatement of its importance. Imports, a negative, are removed in total from GDP making them appear to be completely independent of the other demand components. Since imports are actually purchased by the other components an appropriate amount should be removed from each for a correct distribution of GDP. Imports are purchased either by demand categories directly or by industries as inputs in their production process. Since almost one half of the economy's total output is sold to other industries as inputs it should not be surprising that half of imports are used as

inputs. Input output allows the allocation of imports at the industry level to final demand or to the production process, i.e., intermediate demand. [1]

GDP adjusted for imports is shown in Table. 2. Total GDP obviously remains the same but imports are now reflected in the lower amount of each component. Each was lowered by the sum of the directly allocated and the intermediate imports necessary in their production process. As can be seen, imports were relatively heavily demanded by investment and exports, even though the latter reflects only imports used in the production process. This is understandable since international trade is carried on mainly in manufactured goods, the exact industries from which investment and exports make most of their purchases. One third of merchandise trade is in capital goods, a ratio three times that of investment's share of GDP. PCE and government make a relatively greater share of their purchases from the service industries with a large portion of government being the compensation of its employees so they all consume a relatively smaller share of imports. In 2000 investment purchased 22% of all imports but took only 14% of GDP, government took 9% of all imports but accounted for 18% of GDP, and PCE took 57% while representing 68% of GDP.

Wages by Demand Category

Next GDP was further massaged to allocate by demand component the portion of PCE each created. Input output connects the demand and the supply sides of the economy so wages can be associated with demand. By making certain assumptions about relationships that are already shown in the National Accounts this interconnection is possible. Each demand component generates wages as inputs in the production process. Components differ in the relative amount of wages they generate because of variation in demand patterns along with the production processes these purchases engender, a capability specific to I/O. Industries themselves vary in the per dollar proportion of inputs that are wages since they may have either a relatively large employment component or high wage rates. Two distinct steps are required to get PCE by component, the first requires the generation of wages and the second the conversion of wages to PCE.

Data are available in the National Income Accounts for wages either in total or by industry but not by demand component, this study provides this piece. The general approach used is to convert the Total Requirements Table from one that generates industry output per dollar of demand to one that generates industry wages per dollar of demand, i.e., a Wages Requirement Table. [2]. Running the purchases of each component against this created Wages Requirement Table gives the wages generated by each, table 3.

Not surprisingly the share distribution of generated wages in table 3 closely mirrors that of domestic demand. Slight differences from table 2 are explainable by the uniqueness of the purchases of each demand category. PCE and exports have a smaller wage share because their purchases include those from the agriculture and service industries which have lower wages. Investment, on the other hand, purchases mainly from high wage manufacturing. Government generated wages contain the both the wages directly paid to its employees, in 2000 $769 billion or 44% of its total purchases, while its other purchases generated another $254 billion.

[1] Imports are removed from each row of the Use Table, i.e., the intermediate purchases, by an amount that equals the proportion intermediate is to commodity output. This creates a Use Table free of imports. On the other hand, final demand is converted to domestic demand by removal of a proportion of imports equal to the ratio of demand to commodity output. This "domestic" Use Table is then converted into a "domestic" Total Requirements table that will generate only the domestic output that satisfies domestic demand. Imports are assumed not to be re-exported so only the imports used as inputs to produce exports are removed from exports. The import portion going to intermediate demand is calculated for each demand component from the difference between the outputs generated by running the individual demand components against the "domestic" and the total requirements tables. Data sources: the basic input output tables used for this article were the 1992 benchmark tables that are SIC based. The 1997 NAICS based tables and supporting data were not yet available for use. The final demand distributions by industry for 2000 were BLS derived and that distribution was applied to the GDP controls for 2001 and 2002.

[2]Each cell in a Total Requirements Table represents the value of industry output that a dollar of demand generates. Each industry's output contains a portion that represents the compensation paid to its employees which is shown in the Use Table. Taking this proportion of industry output that is compensation and scaling the rows of the Total Requirements Table convert it to a "Compensation" Requirements Table (each row represents a specific industry's output). Since this study uses partial bills of goods the Domestic Requirements Table is used as well as domestic bills of goods. When this table is run against the individual demand components instead of generating output per dollar of demand we get compensation per dollar of demand. Depending on the differences in component demand structure and generated production processes each component will generate its specific level of compensation. Compensation is composed of both wages and benefits, (health and other insurance; retirement; paid leave; and Social Security), but it is only the wages portion that is spent on PCE. To convert from compensation to wages an assumption is made that the relationship of wages to compensation as shown in the National Accounts at the national level, wages were 84% of compensation in 2000, will hold for all components. This relationship is available on an annual basis so the shift observed over time to a larger benefits portion will be captured. Since the same ratio is applied to all the categories per year the distribution of compensation is the same as that of wages. See "Two Measures of Induced Employment" by Arthur Andreassen; 12th Federal Forecasters Conference, 2002; www.federalforecasters.org.

Induced PCE by Demand Category

The objective of this study has been to determine the sources of the funds that have been spent on PCE. After deriving wages by component, these must be converted to PCE, table 4. PCE by component is then backed out of total PCE and added into the appropriate generating component. The remaining amount of PCE was funded by independent sources. There are no data available specifically connecting wages to PCE by demand component. As was done in previous steps, relationships that hold at the national level are assumed to also hold at the micro level. Within the National Accounts is the relationship of PCE to personal income, in the table of sources and disposition of income. In 2002 72% of personal income was spent on PCE, 28% going mainly to savings and taxes. Since wages were a large part of this personal income, 56%, it is assumed the same percent went for PCE. The PCE remaining after the removal of the wages generated portion comes from other sources, e.g., transfers, dissaving, interest receipts, mortgage refinancing, etc. PCE spending of these funds also generates wages, i.e., PCE itself induces PCE. In 2002 the percent from sources independent of non-PCE wages had risen to 50% of PCE while the PCE they induced equaled another 26%.

Demand Categories Adjusted for Imports and Induced PCE

Finally we reach the denouement of this article with the combination of the separate pieces, table 5. This table shows the true impact of each component on GDP from 2000 to 2002. Comparing this table with table 1 we see PCE has declined 23% in share of GDP because of the removal of imports and non-PCE induced consumption. Although from 2000 and 2002 PCE's share has been whittled down from 70% to 47% it has still increased its relative share 3.1%. This is due to the steep drop in induced PCE from both investment and exports which lessen the decrease in PCE while increasing theirs. After adjustments investment is smaller due to its high import content while government has a relatively larger share with its combination of low imports and high induced PCE. In dollar terms, of the $621 billion increase in GDP over these three years almost all, $584 billion or 95%, comes from this pared down level of PCE balancing investment's share decline of 3.3% %. Finally, increases in revised Government offsets declines in exports. One purpose of this exercise has been to stress the fact that PCE is not a component entirely independent of the others, that it is dependent on the growth of the other components for its health. This is sometimes not stressed in discussions of cyclical upturns when PCE is often pictured as totally autonomous.

Some Derived Relationships

A benefit of this study is the derivation of specific relationships whose enumeration depends on input output analysis, table 6. Combined these relationships provide a partial multiplier for each demand component, partial because the added impact from induced investment has not been included. Column 1 shows imports per dollar of demand, a leakage that must be considered when determining the impact of changes in fiscal policy. One quarter of an increase investment goes for imports due to the concentration of its purchases in manufacturing as well as outsourcing from its own offshore plants. Government, on the other hand, imports only 7 cents per dollar, reflecting the prominence of services in its purchases plus both the political necessity and legal requirement to purchase from domestic manufacturing industries. Column 2 is wages per dollar of domestic demand a result skewed by the relatively high wages paid in manufacturing. Government tops all because of its relatively higher concentration of direct purchases of compensation. Column 3 is the induced PCE per dollar of total demand. Netting columns 1 and 3 gives a partial multiplier per dollar of total demand. All components have a positive multiplier reflecting the fact that the induced consumption of each is greater than the loss due to imports. If one accepts the calculations to this point and ignores the political ramifications it is obvious that, from a fiscal standpoint, increases in the Government component provide the most bang for the buck.

Conclusion

This paper started out to determine the impact of PCE over the recent cycle that is independent of the other demand components and this entailed the removal of the influences of imports and non-PCE wages. After quantifying this relationship some further insights were acquired. Concerning the large impact of government demand, two factors were mainly responsible, the relatively low level of government imports and the relatively high level of induced PCE. However a caveat is necessary because the spending pattern of marginal increases in government purchases, which is what is important in fiscal policy, is more similar to that of PCE or investment than the government pattern at the base of this study because it will not contain the same large proportion of government compensation as is in the government purchases used in this study.

Table 1.
Gross Domestic Product and Its Major Components.
(billions of current dollars)

	Values			Percent		
	2000	2001	2002	2000	2001	2002
Gross Domestic Product	9,825	10,082	10,446	100.0	100.0	100.0
Personal consumption expenditures	6,684	6,987	7,304	68.0	69.3	69.9
Investment less residential construction	1,329	1,141	1,121	13.5	11.3	10.7
Residential construction	426	445	472	4.3	4.4	4.5
Exports	1,101	1,034	1,015	11.2	10.3	9.7
Government	1,751	1,858	1,973	17.8	18.4	18.9
Imports	-1,466	-1,383	-1,439	-14.9	-13.7	-13.8

This table shows Gross Domestic Product and the final demand components as usually presented. Total imports are removed solely from GDP thus overstating each component by its imports.

Table 2.
Gross Domestic Product with Imports Allocated by Demand Component.
(billions of current dollars)

	Values			Percent		
	2000	2001	2002	2000	2001	2002
Gross Domestic Product	9,825	10,082	10,446	100.0	100.0	100.0
Personal consumption expenditures	5,855	6,186	6,463	59.6	61.4	61.9
Investment less residential construction	1,006	857	831	10.2	8.5	8.0
Residential construction	368	387	410	3.7	3.8	3.9
Exports	974	921	902	9.9	9.1	8.6
Government	1,622	1,731	1,839	16.5	17.2	17.6

This table shows each component's purchases less imports, i.e., domestic demand.

Imports are first allocated at an industry level to final and intermediate demand in the proportion they are of output. A separate calculation generates intermediate imports by demand category.

Table 3.
Wages and Salaries Allocated to the Generating Demand Component.
(billions of current dollars)

	Values			Percent		
	2000	2001	2002	2000	2001	2002
Total Wages and salaries	4836	4951	5004	100	100	100
Wages and salaries from:						
Personal consumption expenditures	2,639	2,788	2,839	54.6	56.3	56.7
Investment less residential construction	577	490	464	11.9	9.9	9.3
Residential construction	179	188	194	3.7	3.8	3.9
Exports	417	393	376	8.6	7.9	7.5
Government	1,023	1,092	1,131	21.2	22.1	22.6

Compensation is initially derived by multiplying a compensation requirements table by the individual bills of goods.

Derived compensation is then scaled by an annual wages/compensation ratio from the National Accounts giving wages by component.

See: "Two Measures of Induced Employment", Art Andreassen, 12th Federal Forecasters Conference, 2002.

Table 4.
Induced Personal Consumption Expenditures Allocated to Generating Demand Component.
(billions of current dollars)

	Values			Percent		
	2000	2001	2002	2000	2001	2002
Total domestic PCE	5,855	6,186	6,463	100	100	100
Induced PCE from:	1,537	1,531	1,558	26.3	24.7	24.1
Investment less residential construction	404	347	334	6.9	5.6	5.2
Residential construction	125	133	140	2.1	2.2	2.2
Exports	292	278	270	5.0	4.5	4.2
Government	716	773	814	12.2	12.5	12.6
Autonomous PCE	4,318	4,656	4,905	73.7	75.3	75.9
PCE independent of self induced PCE	2,840	3,055	3,246	48.5	49.4	50.2
Induced PCE, self generated	1,478	1,601	1,659	25.2	25.9	25.7
Total Induced PCE	3,015	3,131	3,217	51.5	50.6	49.8

Wages from table 3 are then scaled by an annual PCE/Personal Income ratio from the National Accounts deriving induced PCE.

This PCE was further adjusted to remove imports. Autonomous PCE is the residual of total domestic PCE in table 2 less induced PCE.

Table 5.
Gross Domestic Product After Re-allocation of Imports and Induced PCE.
(billions of current dollars)

	Values			Percent		
	2000	2001	2002	2000	2001	2002
Gross Domestic Product	9,825	10,082	10,446	100	100	100
Personal consumption expenditures	4,318	4,656	4,905	44.0	46.2	47.0
Investment less residential construction	1,410	1,203	1,165	14.3	11.9	11.2
Residential construction	493	520	550	5.0	5.2	5.3
Exports	1,266	1,199	1,173	12.9	11.9	11.2
Government	2,338	2,504	2,653	23.8	24.8	25.4

Table 5 combines tables 2 and 4: the domestic purchases of non-PCE demand components in table 2 are increased by their induced PCE, table 4, while domestic PCE is reduced by induced PCE.

Table 6.
Relationships Derived from the Preceding Tables: 2000.
(current dollars)

	Value			
	1	2	3	4
	Imports per Dollar of Total Demand (cents)	Wages per Dollar of Domestic Demand (cents)	Induced PCE per Dollar of Total Demand (cents)	Partial Multiplier Col 3 less Col 1 (cents)
Personal consumption expenditures	12	45	30	18
Investment less residential construction	25	57	30	5
Residential construction	14	49	29	15
Exports	12	43	27	15
Government	7	63	41	34

Imports per dollar of total demand is the difference of table 2 less table 1 divided by table 1 purchases.

Wages per dollar of domestic demand are table 3 components divided by table 2 components.

Induced PCE per dollar of total demand is table 4 components divided by table 1 components.

STRUCTURAL CHANGE IN THE GLOBAL SOYBEAN MARKET: IMPLICATIONS FOR FORECASTING U.S. COMMODITY PRICES CONSISTENT WITH FORECASTED U.S. QUANTITIES

Gerald Plato and William Chambers, U.S. Department of Agriculture

South American soybean production is a major source of structural change in the global soybean market that puts downward pressure on U.S. farm prices. This paper reviews the development of the South American soybean industry and the increases in global use and trade, which are also major structural changes in the global soybean market. The main objectives of this paper are to better understand the impact of South American soybean production on global and U.S. markets, and to estimate an equation for forecasting U.S. soybean price to assist USDA forecast efforts.

USDA commodity analysts use forecasting models and individual and consensus judgements in arriving at official USDA price and quantity forecasts for soybeans and other commodities (Vogel and Bange).

They sometimes use forecasting equations to evaluate their consensus forecasts and sometimes change the forecasts provided by price forecasting equations. Their commodity price and quantity forecasts along with historical prices and quantities and market analysis are published each month in USDA's World Agricultural Supply and Demand Estimates (WASDE). Soybean price and quantity forecasts along with historical data and analysis of the soybean market are published in Oil Crops Outlook each month except for October.

The U.S. soybean carryover stocks-to-use ratio and South American soybean production provide a strong basis for price forecasts. However, using South American production forecasts to help in forecasting the U.S. price presents a major challenge. Unlike for the U.S., there are no data on planting intentions and relatively little information on the condition of the growing crop. The rapid growth in South American soybean production also contributes to forecasting difficulties by changing the traditional relationship between U.S. stocks-to-use and price. The U.S. stocks-to-use ratio is traditionally a critical variable in forecasting commodity prices. As a test of our model, we made ex ante soybean price forecasts using only the data available to USDA commodity analysts when they made their forecasts and then compared our results with official USDA estimates at the same point in time.

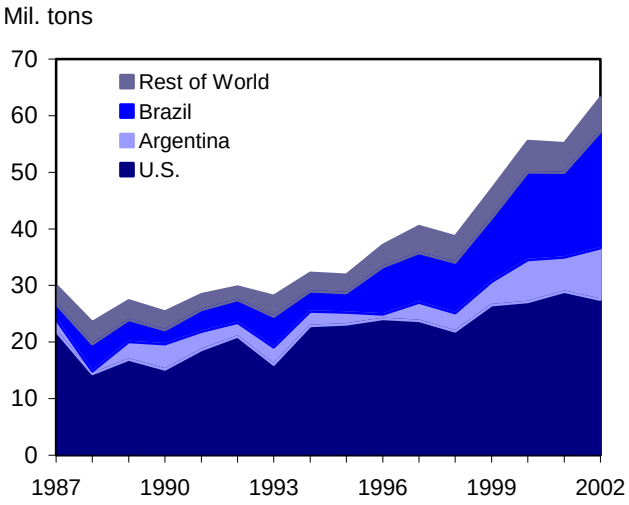
South American soybean production has a large impact on the season average soybean price received by U.S. farmers. Our soybean price forecasting equation estimates that each 1 percent increase in South American soybean production decreases the season average soybean price received by U.S. farmers by about ¼ percent. The U.S. carryover stocks-to-use ratio is smaller at each price level due to the greater potential of South American farmers to make up for any U.S. production shortfalls and due to increased South American carryover.

Background for Structural Changes in the Soybean Market

Brazil and Argentina have become major competitors to the United States in the global soybean market. This structural change has had a dramatic impact on the market dynamics of the soybean sector and complicates price forecasting efforts. Traditionally, the United States was the dominant country in the global soybean market. However, soybean production in Brazil and Argentina increased 223 percent and 204 percent, respectively, between 1990 and 2002. This led to a large increase in the South American share of world markets. U.S. soybean production also increased in the 1990's, but this increase has been much smaller than the production increase from South America.

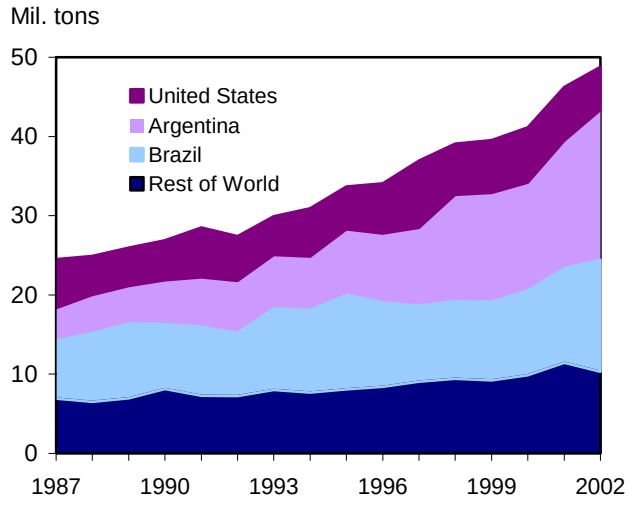
Seasonal cropping patterns in Brazil and Argentina are roughly six months different from those of the United States (e.g. they harvest their crop in the spring when the U.S. is planting). A counter-seasonal pattern has additional market implications because it makes global soybean supplies much steadier throughout the marketing year. This changes pricing, marketing, and stock holding patterns. Now there is a major harvest every six months as opposed to every 12 months.

Figure 1
World soybean exports



Source: Foreign Agricultural Service, USDA.

Figure 2
World soy-meal exports



Source: Foreign Agricultural Service, USDA.

Agricultural production in Argentina and Brazil is traditionally concentrated in the northern third of Argentina and the bordering southern portion of Brazil (this region also shares borders with Paraguay and Uruguay). This warm, humid, and semitropical area is highly productive for agriculture. A critical change has been the expansion of agricultural production into the center-west region of Brazil. Today, the center-west rivals the south as Brazil's primary agricultural production region, and there remains a large potential for further expansion (Schnepf, Dohlman, and Bolling).

The center-west lies entirely within South America's tropical zone and Brazil has developed new crop varieties that grow well in this environment. Vast tracts of virgin lands, which can be used for agricultural production remain undeveloped. A significant portion of these virgin lands are savanna-like flat lands—referred to as cerrado—which can easily and inexpensively be converted to agricultural production. Because of these untapped land resources, Brazil has a tremendous capacity to increase its agricultural production. Poorly developed transportation and marketing infrastructure has been a major problem in developing Brazil's interior regions for agricultural use. However, investments have been made to improve infrastructure and continued growth in soybean production is expected in Brazil's center-west region.

Superior infrastructure in the U.S. has been the primary competitive advantage over Brazil and Argentina in agricultural production and marketing. The United States has a widespread internal transportation network that can quickly and inexpensively move large volumes of commodities from producers to consumers. This includes a system of barges on the Mississippi River, numerous rail lines, and paved highways. The U.S. has also traditionally had greater storage capacity for agricultural commodities. Because of these advantages, transportation and marketing costs have traditionally been significantly lower for U.S.-produced commodities than commodities from either Brazil or Argentina. However, investments in Brazilian and Argentinean infrastructure are starting to narrow this gap making Brazil and Argentina more competitive in world markets.

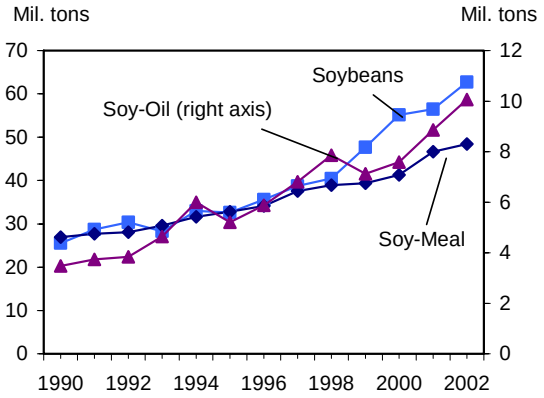
The Parana-Paraguay river system is an important waterway serving, in particular, Argentina's grain and oilseed sector. The Amazon River and its many tributaries represent significant potential for expanded/improved grain transportation in Brazil, and infrastructure development is beginning to open Brazil's interior agricultural areas to export markets. Both Brazil and Argentina have also invested in rail lines and paved highways that can be used for agricultural marketing. In addition, the transformation of both Brazil's and Argentina's economies from currencies that were pegged to the dollar during the 1990's to floating exchange rates have also improved their incentives for agricultural production. There is additional potential for both countries (but Brazil in particular) to improve their

marketing and transportation efficiencies and further enhance their global competitiveness.

Growth in consumption has kept pace with the dramatic increases in soybean production. Between 1990 and 2002, global trade in soybeans, soy-oil, and soy-meal increased 145 percent, 190 percent, and 80 percent respectively. A major factor in the oilseed sector for the past several years has been China's large soybean imports. As investment in domestic crushing capacity swelled, China's imports went from almost nothing in the early 1990's to 18 million tons in 2003. U.S. soybean trade with China increased substantially during this period. However, trade with other countries (especially Brazil and Argentina) increased even more.

Figure 3

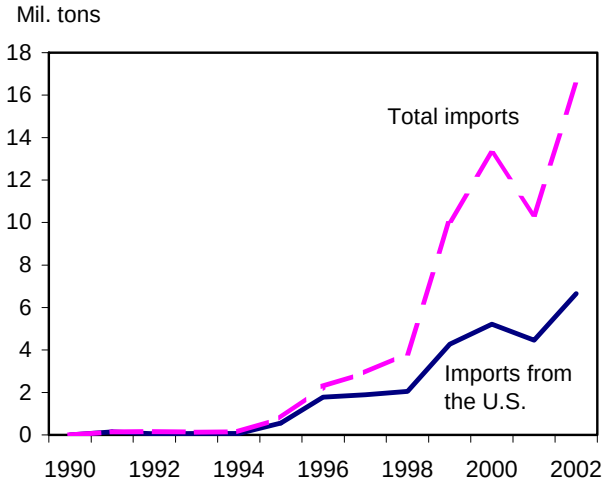
Global trade of soybeans and soybean products increased in the 1990's



Source: Foreign Agricultural Service, USDA.

Figure 4

China soybean imports



Source: Foreign Agricultural Service, USDA.

An Economic Model for Soybeans

The economic model presented in this section helped to develop our soybean price forecasting equation. It also helps in understanding how the soybean market changes previously discussed are changing the relationships among key U.S. soybean variables.

Equations 1 through 5 represent a structural model of the U.S. soybean market. It is used to explain the relationship between the stocks-to-use ratio and price. This relationship is often used to forecast a price that is consistent with forecasted quantities. Changes in the relationships between price and the dependent variables in equations 2 through 5 define structural change and can affect the relationship between the stocks-to-use ratio and price. The structural model is also used to explain how structural change from South American production and increased world use alters the relationships between price and the dependent variables in equations 2 through 5 and the relationship between stocks-to-use ratio and price.

Equation (1) is an identity describing the U.S. soybean market. It shows that carryover from the previous marketing year plus the harvest at the beginning of the current marketing year equals use in the current marketing year plus the carryover from the current marketing year into the next marketing year. Soybean imports are negligible and were left out of the equation.

$$(1) \quad C_{t-1} + H_t = U_t + C_t$$

where:

C_t = U.S. carryover in year t,

C_{t-1} = U.S. carryover in year t - 1

H_t = U.S. production (harvest) in year t,

U_t = utilization in year t (U.S. consumption and U.S. exports), and

t = represents a marketing year which begins at harvest and ends at the beginning of the following harvest.

C_t and U_t in equation 1 are determined jointly for marketing year t, given, H_t, the harvest outcome, and C_{t-1}, the carryover from the previous marketing year. H_t, is realized at the beginning of marketing year t and C_{t-1} is determined in the previous marketing year jointly with U_{t-1}.

Equations 2 through 5 show that each of the variables in equation 1 are a function of price.

(2) $H_t = f_1(E(p_t)) + e_t$

(3) $U_t = f_2(p_t)$

(4) $C_t = f_3(p_t, E(p_{t+1,}\ p_{t+2,}\ .....))$

(5) $C_{t-1} = f_4(p_{t-1}, E(p_{t,}\ p_{t+1,}\ .....))$

Equation 2 shows that the harvest outcome is a function of expected price and an error term (all yield variations are in the error term). The price expectation is formed at and prior to planting. The error term is unforeseen yield variability. Equations 3 and 4 show that year t use and carryover depend on current year price. Carryover also depends on expected price in future marketing years. Equations 3 and 4 in the structural model do not have error terms because year t supply ($H_t + C_{t-1}$) is exactly divided between current year use and carryover. Equation 5 shows that carryover for year t-1 differs from the carryover for year t in equation 4 by having all the marketing year indexes reduced by 1.

Use of mathematical algorithms, particularly dynamic programming, to solve equations 1 through 5 have greatly improved our understanding of the relationships among carryover, production, utilization and price (Makki et al.). The improved understanding helps in forming hypotheses about the relationship between the U.S. soybean carryover stocks-to-use ratio and the U.S. soybean season average price and about the relationships of structural change variables with season average price. The stocks-to-use ratio is a comprehensive variable in that it incorporates both supply and demand effects on price, and is used widely by commodity analysts for forecasting price (Westcott and Hoffman). However, the relationship between stocks-to-use ratio and price is changed by structural change.

Equations 1 through 5 imply that a large supply in year t, due to a large yield outcome, results in a large carryover and utilization and a low price. The low price makes carryover more competitive with next year's expected production. As a result, carryover is a larger portion of next year's expected supply. It is also larger relative to current year utilization resulting in a large stocks-to-use ratio. Conversely, a small supply in year t due to a low yield outcome results in low utilization and carryover and a large price. The large price makes carryover less competitive with next year's expected production. As a result, carryover is a smaller portion of next year's expected total supply and smaller relative to current year utilization resulting in a small carryover stocks-to-use ratio.

A simple way of explaining the inverse relationship between the stocks-to-use ratio and price is to assume that demand for carryover (equation 4) is more elastic than demand for current year use (equation 3).[1] The greater price elasticity for carryover implies that carryover will decrease more than current year use when supply is small and price is high, resulting in a small stocks-to-use ratio. It also implies that carryover will increase more than current year use when supply is large and price is low, resulting in a large stocks-to-use ratio.

Increased South American soybean production reduces U.S. price by increasing world supplies. It also affects the price-quantity relationships in equations 2, 3, 4, and 5.[2] Equation 2 is affected because the increased South American production decreases the U.S. expected price.[3] Equation 3 is affected because there is less export demand for U.S. soybeans at each price level; soybean exports typically account for 35-40 percent of total U.S. soybean use. Equations 4 and 5 are affected because U.S. carryover is smaller at each price level due to the larger potential for South American farmers to respond to U.S. harvest shortfalls.[4] Most likely, carryover will decrease more than current year use at each price level, resulting in a smaller stocks-to-use ratio at each price level.

Equation Estimation and Selection

A forecasting equation for U.S. season average soybean price was selected based on equation statistics and on our understanding of the soybean market as discussed in the previous section. We experimented with several structural change and policy variables. The U.S. stocks-to-use ratio was important in all our equation experiments. Our approach was to keep the forecasting model as simple as possible and avoid "mining" the data. We first tried using only the stocks-to-use ratio. We experimented with using the 1975-2000 period and several periods with later beginning dates. Most

[1] Substitution between carryover and expected production next year makes demand for carryover more price elastic than demand for current year use.

[2] A change in the relationship between quantity and price in any of the equations is a structural change.

[3] There would likely be structural change in equation 2 in absence of increased South American production due to increased U.S. productivity and policy changes.

[4] For simplicity, we are not including other variables (such as South American Production) in equations 1-5 because there influence is captured indirectly by their impact on soybean price. Changes in these other variables represent structural change in the soybean market.

likely, the stocks-to-use ratio would be the only independent variable in absence of structural change. None of the equations were satisfactory because they had low Durbin Watson statistics and low t values. We then included South American production and decided on starting the analysis in 1987, which was about the time when South American production began to increase. We also tried global use in the regression analysis but decided not to use it on statistical grounds. Our final estimated equation is shown in equation 6, and the variable definitions are provided in table 1.

Table 1—Summary of variable definitions

Variable Name	Definition
SP	US season average soybean price ($/bushel)
SUR	US soybean stocks–to-use ratio (expressed as a ratio)
PSA	Soybean production in South American (million bushels)
LN	Natural Log

(6) Ln SP = 4.62 – 0.41 • Ln SUR* – 0.52 • Ln PSA*
R-bar-sq = 0.75
F-Value = 23.41
Standard error of regression = 0.0808
Durbin-Watson statistic = 2.22
Estimation Period: 1987-2002
* Significant at the 99 percent level

Since the data were converted to logs, the variable coefficients are elasticities that estimate the percent change in price for a one percent change in the variable. The data used for estimating this and our other equations were taken from USDA's production supply and distribution database at http://www.fas.usda.gov/psd/.[5] Ordinary least squares was used to estimate this equation.

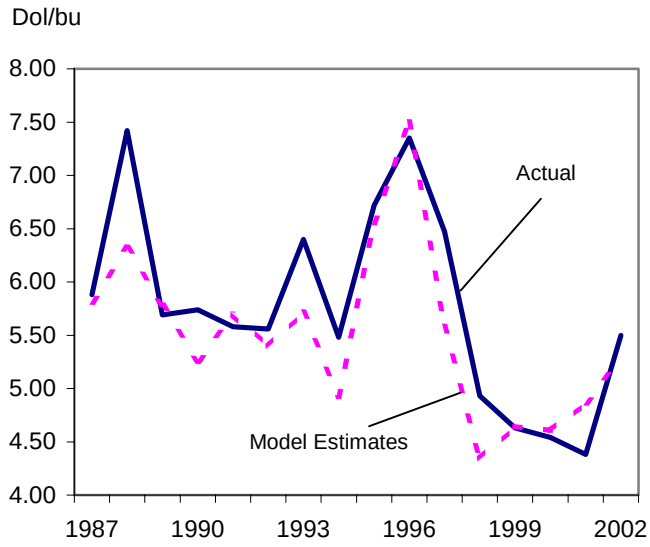
We did not include both global use and South American production as independent variables because they are highly correlated. Due to this correlation, global use had a negative sign, significant at the 1 percent level when South American production was excluded. The structural model suggests the sign should be positive.

[5] Data in this database are reported in metric tons. We converted to bushels using 1 metric ton = 36.7437 bushels (Weights, Measures, and Conversion Factors for Agricultural Commodities and Their Products, p. 10).

Ex Post Price Forecasting and Evaluation

We next examined the *ex post* forecasting capability of our regression equation. *Ex post* forecasts from the estimated equation (6) and actual outcomes are displayed in figure 5 over the model estimation period. The prices estimated from the model follow the general trend of the actual prices, and the mean absolute deviation and mean absolute percentage differences are $0.36/bu and 6 percent respectively. These relatively large errors may be because the soybean industry is in a state of flux resulting in regression parameters (relationships) changing over time. In particular, the South American industry became a more important producer during this period. This outcome alone could have had a large impact on the regression coefficients in the model. It is also important to note that the World Agricultural Supply and Demand (WASDE) projections of soybean season average farm prices provide ranges that are several times larger than the mean absolute deviation from the model (particularly early in the crop year).

Figure 5
Soybean prices: actual and model estimates



A more serious potential problem with the model has to do with turning point errors. A turning point error can be defined statistically when either of the following inequalities (7) and (8) hold.

(7) $(Predicted_t - Actual_{t-1}) * (Actual_t - Actual_{t-1}) < 0$

(8) $(Predicted_t - Predicted_{t-1}) * (Actual_t - Actual_{t-1}) < 0$

Predicted prices are derived from the models, and actual prices are those prices received by farmers as reported by the National Agricultural Statistics Service. The subscripts "t" and "t-1" represent current and lagged time periods, respectively. Defined in this way, the statistic measures whether predicted year-to-year changes from the models are directionally the same as changes in actual prices. Turning point errors can occur in two ways: first, when actual prices indicate a turning point but predicted prices do not and, second, when actual prices do not indicate a turning point but predicted prices show a turning point. The different definitions for the occurrence of a turning point in equations (7) and (8) related to whether the change in the predicted price is measured relative to the previous year's actual price (Equation 7) or the previous year's predicted price (equation 8). Both measures are useful, but the appropriate measure depends on the intended use of the model. For short term forecasting applications, where the previous year's actual price is known, the former definition is better. For longer-term applications, where the previous year's price is not known, the latter definition is better. (Westcott and Hoffman)

Turning point errors using the first definition were identified in the years 1990 and 2001. Turning point errors using the second definition were identified in 1990, 1991, 1999, and 2001. The fairly numerous turning point errors highlight the difficulty in price forecasting in the changing environment of the soybean industry.

Ex ante Price Forecasting and Evaluation

Table 2 shows the months and years in which our equation and the WASDE forecasts were made, the forecast years, and the data periods used in estimating our forecasting equation.

Table 2. Ex ante Price Forecasting and Evaluation Schematic.

Date of Forecast	Equation U.S. Soybean Price Forecast	WASDE U.S. Soybean Price Forecast 1/	WASDE U.S. Stocks-to-Use Ratio Forecast 1/	WASDE S. American Production Forecast 1/	Equation Data
Column 1	Column 2	Column 3	Column 4	Column 5	Column 6
July 2000	2000	2000	2000	2001	1987 – 1999
Aug. 2000	2000	2000	2000	2001	1987 – 1999
Sept. 2000	2000	2000	2000	2001	1987 – 1999
Oct. 2000	2000	2000	2000	2001	1987 – 1999
Nov. 2000	2000	2000	2000	2001	1987 – 1999
Dec. 2000	2000	2000	2000	2001	1987 – 1999
July 2001	2001	2001	2001	2002	1987 – 2000
Aug. 2001	2001	2001	2001	2002	1987 – 2000
Sept. 2001	2001	2001	2001	2002	1987 – 2000
Oct. 2001	2001	2001	2001	2002	1987 – 2000
Nov. 2001	2001	2001	2001	2002	1987 – 2000
Dec. 2001	2001	2001	2001	2002	1987 – 2000
July 2002	2002	2002	2002	2003	1987 – 2001
Aug. 2002	2002	2002	2002	2003	1987 – 2001
Sept. 2002	2002	2002	2002	2003	1987 – 2001
Oct. 2002	2002	2002	2002	2003	1987 – 2001
Nov. 2002	2002	2002	2002	2003	1987 – 2001
Dec. 2002	2002	2002	2002	2003	1987 – 2001

1/ The WASDE forecasts were taken from *World Agricultural Supply and Demand Estimates*, and *Oil Crop Outlook* for the months and years in column 1.

Each row in table 2 shows:

1 the month and year an equation forecast and the WASDE forecasts were made (column 1),
2 the marketing year for which the equation price forecast and the WASDE price forecast were made and the marketing year for which the WASDE stocks-to-use ratio forecast was made (columns 2, 3, and 4),
3 the year for which the South American production forecast was made (column 5), and
4 the marketing years for the equation data (column 6).

Each equation forecast uses the equation based on the marketing years in column 6 and the corresponding WASDE forecasts in columns 4 and 5. Forecast comparisons were made by comparing the forecasts in columns 2 and 3 for each row.

Table 3 contains the estimated equation coefficients used to forecast price and contains selected equation statistics. Data revisions make the 1987-2002 data for equation (6) slightly different from those for the 2002 equations in table 2. Equation 6 is based on data revisions through August 2003.

Table 3. Soybean Price Forecasting Equations.

Year 1/	Month 1/	Beta1	Beta2	Beta3	R-Sq 2/	DW 3/
2000	July	4.00	-0.40	-0.43	0.75	2.40
2000	August	4.05	-0.40	-0.43	0.73	2.34
2000	September	4.12	-0.40	-0.44	0.71	2.25
2000	October	4.01	-0.40	-0.43	0.75	2.40
2000	November	4.00	-0.40	-0.43	0.75	2.42
2000	December	4.02	-0.40	-0.43	0.75	2.40
2001	July	4.48	-0.41	-0.50	0.79	2.14
2001	August	4.47	-0.41	-0.49	0.78	2.14
2001	September	4.53	-0.41	-0.50	0.77	2.06
2001	October	4.45	-0.41	-0.49	0.79	2.18
2001	November	4.44	-0.41	-0.49	0.79	2.19
2001	December	4.43	-0.41	-0.49	0.79	2.21
2002	July	4.91	-0.40	-0.55	0.79	1.84
2002	August	4.95	-0.39	-0.55	0.77	1.78
2002	September	4.95	-0.39	-0.55	0.77	1.78
2002	October	4.89	-0.40	-0.55	0.79	1.86
2002	November	4.88	-0.40	-0.55	0.79	1.88
2002	December	4.86	-0.40	-0.55	0.79	1.89

1/ Each equation is based on the latest available data for the month and year indicated. Year is also the marketing year for which the season average soybean price is forecast. 2/ R_sq is the corrected R-square. 3/ All the Durbin-Watson test statistics are in the do-not-reject range at the 5 percent significance level.

Beta1 is the equation intercept. Beta2 is the coefficient for the U.S. stocks-to-use ratio. Beta3 is the coefficient for South American Production. All the beta coefficients are significant at the 1 percent level. Beta2 and Beta3 are elasticities—they estimate the percentage change in price from a one percent increase in the U.S. stocks-to-use ratio and from a one percent increase in South American soybean production, respectively.

Equations within each year in table 3 vary slightly because data in the last data year and sometimes in the next-to-last data year are revised from month to month. Coefficient variation across years in table 2 may be due to structural change. All the coefficients in table 2 are significant at the 1 percent level. The corrected R squares range from 0.71 to 0.79. All the Durbin-Watson statistics are in the do-not-reject range at the 5 percent level.

Table 4 contains summaries of the equation and WASDE forecast errors. Equation forecast errors were about the same as the WASDE forecast errors for 2000 and 2001, but much larger for 2002. Interestingly, the *ex ante* forecast errors for 2000 and 2001 are smaller than the *ex post* forecast errors for the 1987-2002 period as reported in the previous section.

Table 4. Mean Absolute and Mean absolute Percentage Forecast Errors for Forecasting Equation and for WASDE Forecasts.1/

Year	Equation Absolute Mean Errors	WASDE Absolute Mean Errors	Equation Absolute Percentage Errors	WASDE Absolute Percentage Errors
2000	0.20	0.22	4.3	4.9
2001	0.20	0.22	4.7	4.9
2002	0.63	0.23	11.7	4.2

1/ Mean absolute errors are in dollars per bushel.

Conclusions

This paper examines the changing structure of the global soybean industry and provides forecasts for season average soybean prices. Expanded competition from South America is having a major impact on the soybean market and on soybean price forecasting equations. We found that the U.S. stocks-to-use ratio and South American soybean production were sufficient variables for forecasting price.

Estimating a soybean price forecasting equation each year using the latest data appears to be needed due to ongoing structural change in the global soybean market. The updated equation each month can provide useful price forecasts. However, equation forecasts can only be part of the input into making price forecasts because equation forecasts can sometimes be wide of the mark.

Our results demonstrate that *ex ante* forecast evaluation is needed in addition to equation estimation and *ex post* evaluation for evaluating and choosing a soybean price forecasting equation when the soybean market is experiencing rapid structural change.

Box: Impact of South American Production on U.S. Farm Price

Our forecasting equation was used to examine the downward pressure on U.S. soybean price from South American production. Understanding this downward price pressure is important for budgeting counter cyclical payments and marketing loan assistance program for soybeans under the Farm Security and Rural Investment Act of 2002. [6]

The coefficient for South American production in equation (6) says that, other things equal, a 1 percent increase in South American production decreases U.S. soybean price by about ½ percent. However, other things are not equal. The U.S. soybean industry has responded to increased South American production by carrying fewer stocks and lowering production relative to what it would have been without the increased South American production. This latter effect also influences the U.S. soybean price though it is an indirect effect of South American production on price. This indirect effect can be combined with the effect of South American production on the U.S. soybean price.

To analyze this we used a procedure first developed by Buse (shown below) that uses the elasticity coefficients in equation (6). In addition, we had to estimate the change in the U.S. stocks-to-use ratio from a 1 percent increase in South American production, which was calculated to be -0.64. The equation below shows our calculation for the percent change in the U.S. soybean price given a 1 percent increase in South American production.

Percent U.S. soybean price change = (-0.41)(-0.64%) + (-0.52)(1%) = -0.26%

- This equation indicates that a 1 percent increase in South American production decreases U.S. soybean price by 0.26 percent. This equation combines the direct effect of South American production and an indirect effect via the effect of South American production on the U.S. stocks-to-use ratio. Decreases in the soybean season average price increase USDA counter cyclical expenditures when the season average price is between the target price minus the direct payment rate and the national loan rate.[7] A 0.26 percent decrease in price when the season average price is in this range is between 1.3 and 1.4 cents per bushel.

[6] The October and February WASDE soybean price forecasts are used in calculating advanced counter cyclical payments.

[7] The target price, direct payment rate, and national loan rate for soybeans under the 2002 Farm Act are $5.80, $0.44, and $5.00, respectively.

References

Buse, R. "Total Elasticities—A Predictive Device" Journal of Farm Economics, Vol XL, November 1958, No.4, pp. 881-891.

Makki, Shiva S., Luther G. Tweeten, and Mario J. Miranda. "Storage-Trade Interactions Under Uncertainty: Implications for Food Security", Journal of Policy Modeling, Vol. 23, 2001, pp. 127-140.

Schnepf, Randall D., Erik N. Dohlman, and Christine Bolling. Agriculture in Brazil and Argentina: Developments and Prospects for Major Field Crops. Economic research Service, U.S. Department of Agriculture, Agriculture and Trade Report, WRS-01-3, November 2001.

U.S. Department of Agriculture, Economic Research Service, *Oil Crops Outlook*, July, August, September, November and December, 2000, 2001, and 2002. www.ers.usda.gov/publications/outlook

U.S. Department of Agriculture, Office of the Chief Economist. *World Agricultural Supply and Demand Estimates*, Washington DE, July-December, 2000, 2001, 2002. www.usda/gov/oce/waob/wasde/wasde.htm

U.S. Department of Agriculture, Economic Research Service, *Weights, Measures, and Conversion Factors for Agricultural Commodities and Their Products*, Agricultural Handbook Number 697, Washington DC, June 1992.

Vogel, Frederic A. and Gerald A. Bange. *Understanding USDA Crop Forecasts*. National Agricultural Statistics Service and World Agricultural Outlook Board, U.S. Department of Agriculture, Miscellaneous Publication No. 1554, Washington D.C., March 1999.

Westcott, Paul C. and Linwood A. Hoffman. *Price Determination for corn and Wheat: The Role of Market Factors and Government Programs*. Economic Research Service, U.S. Department of Agriculture, Technical Bulletin, No, 1878, Washington DC, July 1999.

FORECASTING THE COUNTER-CYCLICAL PAYMENT RATE FOR U.S. CORN: AN APPLICATION OF THE FUTURES PRICE FORECASTING MODEL

by

Linwood A. Hoffman, U.S. Department of Agriculture, Economic Research Service

Introduction

On May 13, 2002 a Farm Act entitled the Farm Security and Rural Investment Act of 2002 was signed into law covering a period of 6 years, 2002-2007 (USDA, 2002). This Act provides income support to the U.S. corn sector through three different programs: counter-cyclical payments, direct payments, and non-recourse marketing assistance loans. The new counter-cyclical payment (CCP) program was established to provide an improved counter-cyclical income safety net. This component of the safety net was designed to stabilize producer income when prices are low. CCPs replace ad-hoc payments for market loss assistance, provided by Congress on an annual basis from 1998 to 2001. The new Act also provides for direct payments, which replace production flexibility contract payments, a type of direct payment from the 1996 Act. The non-recourse marketing assistance loan program is continued from the 1996 Act.

Both producers and program analysts need forecasts of counter-cyclical payments. Producers need to know how these potential safety net receipts will affect their cash flow. Program analysts forecast government outlays for income safety net programs, such as marketing loan benefits and now counter-cyclical payments. These forecasts are necessary for budget purposes and for the circuit breaker provision in the 2002 Farm Act, which requires the Secretary to adjust expenditures to meet URAA domestic support ceilings. This provision assures that the United States will not exceed its WTO limits (USDA (c)).

A CCP is based on producers' payment acres and payment yields and the national CCP rate. A forecast of the CCP requires a forecast of the payment rate, since both payment acres and payment yield are predetermined. The payment rate is equal to the target price less the effective price, which is equal to the higher of the national average market price or national average loan rate plus the direct payment rate. Thus, a season-average U.S. corn price received by producers is needed in order to estimate the counter-cyclical payment rate.

The U.S. Department of Agriculture analyzes agricultural commodity markets and publishes current crop year market information, including price projections (except for cotton), on a monthly basis. [1] This monthly price projection provides a price forecast that can be used to forecast the counter-cyclical payment rate. [2] However, since producers and program analysts maintain a keen interest in the magnitude of these payment rates, a weekly forecast of the season-average price may be preferable to a monthly projection. Hoffman (2001) modified a model that uses futures prices to provide weekly forecasts of the corn season-average farm price. Such forecasts are reliable, easy to provide, and can be used to forecast the counter-cyclical payment rate. This approach provides a forecast of the season-average price independent of the WASDE season-average price projection.

The objectives of this study are as follows: 1.) Forecast a season-average price received by U.S. corn producers on a weekly frequency. 2.) Forecast an annual counter-cyclical payment rate for U.S. corn on a weekly frequency.

[1] Price projections rely on economic models and analysts' judgement. Econometric price forecasting models are re-estimated periodically because of changes in policy. The most recent updates have been associated with the FAIR Act of 1996 (Westcott and Hoffman, Childs and Westcott, and Meyer).

[2] Since the passage of the 2002 Farm Act, two counter-cyclical payment tools have been developed and posted on the internet. The first tool was developed by Bradley D. Lubben, Kansas State University (http://www.agmanager.info/policy/commodity/default.asp) and the second by the Farmdoc project, University of Illinois (www.farmdoc.uiuc.edu/marketing/CounterCyclical/CCP.asp). Lubben relies on the monthly WASDE releases to compute a projected counter-cyclical payment rate and follows USDA decisions regarding advance payments of the counter-cyclical payment. The Farmdoc project provides a CCP rate for selected commodities based on the monthly WASDE projection of the season-average price, a projected weighted average price needed for the remainder of the year to meet the WASDE projected price, a projected weighted average price needed for the remainder of the marketing year to result in no counter-cyclical payment, and an estimated weighted season-average price to date based on available monthly cash prices.

Background

Agricultural commodity policy was given a greater market orientation beginning with changes made in the 1985 Farm Act and continuing through the 1996 Act. However, the enactment of the 2002 Farm Act marked a switch in this movement, especially with the introduction of the counter-cyclical payment program. Many of the farm and commodity organizations that testified before the House and Senate Agriculture Committees in 2001 requested additional counter-cyclical support be developed as a supplement to the current marketing assistance loans (marketing loan benefits) and fixed annual payments (Becker and Womack). Counter-cyclical payments were provided by the 2002 Farm Act because of low commodity prices in the 1997-2001 period, which led to supplemental emergency assistance payments. Instead of passing annual emergency economic assistance bills, crop revenue shortfalls are now to be offset with the counter-cyclical payment plan.

An example of a past counter-cyclical program is the 1990 Farm Act's deficiency payment program (USDA (e)). [3] Congress specified a target price for each major crop and if the market price was less than the target price, eligible producers received a deficiency payment to make up the difference. The payment rate was determined by the difference between the target price and the higher of the loan rate or market price. Also, the producer was required to plant the base acreage to the program crop, except for 0/92 and limited flex, and comply with the acreage reduction program (ARP).

The 2002 Farm Act's counter-cyclical payments have similarities and differences when compared to the deficiency payments that were made under the 1990 Farm Bill. Both counter-cyclical and deficiency payments are based on historical production, a base, and a target price. In contrast to the deficiency payment program, counter-cyclical payments are accompanied by nearly full planting flexibility and no acreage set-asides. Thus, under the counter-cyclical payment program the producer is not required to plant the program crop to the base acreage.

Counter-cyclical payments are made to producers with an established payment yield and base acres whenever the effective price is less than the target price. Based on the maximum corn payment rate of $0.34/bu., this program could total about $2.4 billion for crop year 2003 if the season-average price was equal to $1.98 or lower, but would be less if the season-average price were greater (table 1). Thus, it is imperative that program analysts and producers pay close attention to the season-average price. CCPs are based on historical area and yields, not a function of current production, but are related to season-average prices. Recipients of the CCP are not required to produce the crop, but they had to produce the program crop under the prior deficiency payment program, except for 0/92 and limited flex. Under the 2002 Act, landowners were able to update base acres and payment yields. Payment yields could be updated if the producer elected to update base acres to the average of planted acres in 1998-2001.

Base Acres—Under the 2002 Act, landowners were able to update their corn base acres if they desired (USDA (h)). One of five choices could be made. 1). Update corn base acres to equal the contract acreage that would have been used for 2002 production flexibility contract (PFC) payments. 2). Update the corn base acres to equal the contract acreage that would have been used for 2002 PFC payments, plus average oilseed acreage that was planted in 1998-2001, up to the base acreage maximum. 3). Update the corn base with the PFC acres plus oilseeds, with a PFC offset. This option allows the producer to add the full soybean plantings but must offset corn base or base for other crops for the soybean base added. 4). Update the corn base with the average acreage planted and prevented to corn in 1998-2001. 5). Update the corn base with the PFC acreage and add oilseed base by reducing PFC acres. This option offers greater flexibility to add oilseed base acres than either options 2 or 3. Preliminary data indicate that about 63 percent of all farmland owners chose to retain their historical PFC acreage (adding oilseeds, if applicable) for their base acreage (USDA (h)).

Landowners had a one-time opportunity to select a method for determining base acreage. Anyone not making a decision was assigned option # 2. Lastly, base acreage cannot exceed available cropland. Adjustments to base acres can be made when a contract for the conservation reserve program expires or is voluntarily terminated. This updating of base acres could lead to an expectation that yields may be allowed to be updated under future farm legislation, and thus could create an incentive for increasing yields (Westcott, Young, and Price, 2002).

Payment acres—Payment acres for counter-cyclical payments are equal to 85 percent of the base acres, which may or may not have been updated. Payment acres for the 1996 Fair Act were similarly 85 percent of the production flexibility contract acres.

3 An example of a current counter-cyclical program is the marketing loan program.

Program yield—Corn payment yields for counter-cyclical payments could be updated by producers that elected to update base acres to average planted acreage in 1998-2001 (option 4) (USDA (h)). These producers had three choices to update yields: 1). Use previously determined program yields. 2). Add to program yields 70 percent of the difference between program yields for the 2002 crop and the farm's average yields per planted acre for 1998-2001. 3). Use 93.5 percent of the 1998-2001 average yields per planted acre. This updating of payment yields could lead to an expectation that yields may be allowed to be updated under future farm legislation, and thus could create an incentive for increasing yields (Westcott, Young, and Price, 2002).

CCP Rate—The CCP rate equals the difference between the target price minus the effective price (Fig. 1). Figure 1 applies to program provisions for crop year 2002 and 2003. The effective price is equal to the sum of 1) the higher of the national average price received (SAP) for corn for the marketing year, or the national average loan rate (NALR) for corn and 2) the direct payment rate (DP) for the commodity.

Equation (1) and (2) consist of six variables. The season-average price, counter-cyclical payment rate and effective price are initially unknown but the value for the target price, loan rate, and direct payment are predetermined (table 3). After a value for the season-average price is derived, the effective price can be determined followed by the counter-cyclical payment rate.

(1). CCP rate ($? /Bu.) = Target Price ($2.60/Bu.) — Effective Price ($? /Bu.).

(2). Effective Price = [(Higher of SAP ($? /Bu.) or NALR($1.98/Bu.)) + (DP)($.28/Bu.)].

The season-average price and counter-cyclical payment rate relationship is illustrated in figure 1 for crop year 2002 and 2003. When the market price is $1.98/bu. (loan rate) the counter-cyclical payment rate is at its maximum of $.34/bu., but declines to zero as the market price rises to $2.32/bu. The market price of $2.32/bu. is called the CCP trigger price because if the season-average price is less than $2.32 per bushel a counter-cyclical payment can be expected. Note the difference between line segment ADE and CFG is $.28/bu. or the direct payment rate.

The relationship between the effective price and the counter-cyclical payment rate is also illustrated in figure 1. The difference between the target price (line segment AB) and the effective price (line segment ADE) equals the counter-cyclical payment rate. This rate remains zero as long as the effective price is equal to or greater than the target price of $2.60/bu., but the payment rate increases to $0.34/bu. as the effective price declines to $2.26/bu. The maximum counter-cyclical payment rate is $0.34/bu.

The 2002 Farm Act states that, if it is determined that a counter-cyclical payment is required, USDA shall pay up to 35 percent of the expected amount in October of the year the crop is harvested, 35 percent after February 1st of the following year, and the remainder as soon as possible after the end of the 12-month marketing year (USDA, 202a).

Forecast Model Justification

Price forecasts have always been useful to market participants when making production and marketing decisions. Many market participants usually forecast a price for a given location and time period when they plan to buy or sell a commodity. One indicator of prices is the futures market, which then requires a prediction of the basis, the difference between the local cash price and the observed futures price. The futures price is an unbiased predictor of the cash price at a delivery location based on the efficient market hypothesis (Fama 1970, 1991). Consequently, the futures price can be combined with a basis forecast to generate a forecast of the cash price at a non-delivery location (an average of locations as in the season-average price received). Futures prices reflect both expected supply and use and thus can be used to forecast short-run farm prices (Danthine, Garnder, Peck, Rausser and Just, and Tomek). Tomek (1997) states that, "futures prices can be viewed as forecasts of maturity-month prices and the evidence suggests that it is difficult for structural or time-series econometric models to improve on the forecasts that futures markets provide."

Season-average price forecasts are of interest to producers and program analysts, especially since counter-cyclical payments are linked to the crop year's season-average price received. Hoffman (1992) developed a model that uses futures prices to forecast the season-average cash price of corn at the U.S. farm level. His model provided forecasts with a mean absolute percentage error of 15 percent beginning in May prior to the crop year but declining to 1 percent for August, the last month of the crop year (Hoffman 2001). This forecasting framework will be used to forecast the season-average U.S. corn price received on a weekly basis. This price forecast will then be used in the computation of the annual counter-cyclical payment rate on a weekly frequency.

Methodology

Procedures for season-average price and counter-cyclical forecasts are discussed in this section.

Forecast Model for Season-Average Prices Received

The futures forecasting model consists of several components: futures prices, cash prices received, basis values (cash less futures), and marketing weights. A forecast of the season-average corn price received is derived from weekly price forecasts, which in turn are based on five futures contracts traded throughout the crop year. The forecast period for each crop year covers 16 months, beginning in May, four months before the start of the crop year, and concluding with August, and the last month of the crop year. [4] The season-average forecast is initially based on futures prices but these prices are replaced with actual monthly cash prices, as they become available from the National Agricultural Statistics Service. Consequently, the season-average price forecast becomes a composite of monthly forecasts and actual cash prices. As the months in which forecasts are made (May…September…January…May…August) move closer to the end of the marketing year, there are more months with actual cash prices and fewer months with forecast prices. The forecast error is expected to decline as the forecast period moves closer to the end of the crop year, as a greater portion of the season-average price becomes known and as information regarding the remainder of the crop year becomes more certain.

The crop year forecast of the season-average farm price (SAP) is computed as follows:

$$SAP_m = \begin{cases} \sum_{i=1}^{12} W_i (F_{mi} + B_i) & \text{for } m = 1 \text{ to } 5. \\ \sum_{i=1}^{m-5} W_i P_i + \sum_{i=m-4}^{12} W_i (F_{mi} + B_i) & \text{for } m = 6 \text{ to } 16. \end{cases}$$

where:

SAP_m = forecast of the season average price made in month m.

W_i = marketing weight for month i.

P_i = cash price in month i.

F_{mi} = observed weekly price in month m for the nearby futures of month i. [5]

B_i = expected basis, which is equal to average cash price in month i minus average futures price in month i for the nearby futures contract. This basis is usually a negative number.

m = 1, 2, 3, ... , 16 months during which forecasts are made (May – August). [6]

i = 1, 2, 3, …, 12 crop year months, September through August.

[4] The forecast period for each crop year is similar for both the futures model forecast and USDA's WASDE forecast.

[5] The nearby futures price is always used except when the forecast month coincides with the nearby futures. For this situation, the next nearby futures is used.

[6] Forecast begins in May, four months before the start of the crop year.

Basis—The difference between the cash price at a specific location and the price of the nearby futures contract is known as the basis. The basis tends to be more stable or predictable than either the cash price or futures price. Several factors affect the basis and help explain why the basis varies from one location to another. Some of these factors include: local supply and demand conditions for the commodity and its substitutes, transportation and handling charges, transportation bottlenecks, availability of storage space, storage costs, conditioning capacities, and market expectations.

The basis computed for this analysis is a 5-year moving average of the monthly U.S. average corn price received by producers less a monthly average of the nearby futures settlement price observed for the particular month. [5] For example, the September basis is the difference between the September average cash price received by producers and September's average settlement price of the nearby December futures contract. The basis for each month is updated at the end of each crop year for use in subsequent years. The basis used in this study therefore reflects a composite of the basis-influencing factors because it represents an average of U.S. conditions, rather than a specific geographic location.

Marketing Weights--{tc \l3 "Monthly Weights}Monthly marketings are used to construct a weighted season-average price. Each month's weight represents the proportion of the year's crop marketed in that month. A 5-year moving average of these monthly weights is constructed and updated annually.

Forecast Procedure

The steps taken to provide the futures price forecast are explained in more detail in this section. Table 2 illustrates the method used in forecasting the season-average corn price for the crop year 2003/04. This method computes a forecast of the season-average price based on futures settlement prices. The forecast is computed weekly, but could be computed monthly or daily. The Thursday futures settlement price for each of the nearby contracts is used for the weekly futures price. [7]

Ten steps are involved in the forecast process:

1. The latest available futures settlement prices are gathered for the contracts that are trading. Settlement prices for Thursday, October 16, 2003 are used for illustration. Futures quotes are for the following contracts: December 2003, and March, May, July, and September 2004 and are stored in line 1 of the model's spreadsheet (table 2).

2. The futures price for September, October, and November 2003 (line 2, table 2) represents the October 16th settlement price of the nearby contract, December 2003. The settlement price for the nearby (March) contract is used for the months of December, January, and February. For those months when a futures contract matures, the next nearby contract is used because of greater price stability. Futures prices for the maturing contract are affected by a decline in liquidity during the month of maturity. Also, a contract usually closes about the third week of the month, and using the current futures contract during its closing month would lower the number of observations that could be used to calculate the average monthly closing price and corresponding basis.

3. A 5-year moving average basis (monthly cash price minus the nearby futures price) is on line 3 of table 2. This average basis is updated during the first week of October, when the full-month August cash price is available thus completing all the monthly cash prices for the prior marketing year.

4. A forecast of the monthly average farm price (line 4 of table 2) is computed by adding the basis (line 3) to the monthly futures price (line 2).

5. The actual monthly average farm price is on line 5 of table 2, as it becomes available. The $2.13 per bushel on line 5 represents the mid-month September price as obtained from the Agricultural Prices report issued in late September. On November 6, 2003 the actual full-month September cash price will be entered as obtained from the Agricultural Prices report issued in late October and the mid-month October cash price is also entered.

6. The actual and forecast farm prices are spliced together on line 6. The price forecast for crop year 2003/2004, as computed on October 16, 2003, uses futures forecasts for 11 of the 12 months of the marketing year, October through

[7] Thursday is picked because there are fewer holidays and no beginning or end of week surprises.

August (from line 4), because cash prices are available for only September.

7. The monthly weights, expressed as a percent of total crop year marketings, are on line 7 of table 2. A 5-year moving average is used and updated in early October, after the release of the September Agricultural Prices report.

8. A weighted season-average U.S. farm price received forecast is computed (line 8) by multiplying the monthly weights on line 7 by the monthly farm prices on line 6 and summing their products.

9. A simple average price forecast is also computed (line 9).

10. A forecast of the Counter-Cyclical Payment Rate is computed (line 10).

Data

The futures forecasting model requires monthly data by crop year for the following items: 1) monthly settlement prices from the nearby futures contracts; 2) monthly (mid- and full-month) producer cash prices; and 3) monthly marketing weights. These data are collected for crop years 1981 through 2002 and are used to construct the 5-year moving average basis and marketing weights. The 5-year averages for bases and monthly marketing weights begin with 1981-85 data and are updated to the present. These data are used to evaluate the futures model's historical performance.

Weekly settlement prices from the nearby futures contracts are collected for crop years 2002 and 2003. These futures prices are used to produce a cash price forecast for crop year 2002 and 2003. A weekly season-average price forecast requires an update of weekly futures prices, available cash prices, and marketing weights on a periodic basis.

Historical daily settlement prices by contract (December, March, May, July, and September) are obtained from the Chicago Board of Trade for crop years 1981 through 2002. Cash prices received are obtained from Agricultural Prices, published by USDA's National Agricultural Statistics Service. Price projections from the U.S. Department of Agriculture are obtained from World Agricultural Supply and Demand Estimates (WASDE) published by USDA's World Agricultural Outlook Board. Weights for monthly marketings are derived from data published in various issues of USDA's December Crop Production. Beginning in 1997, monthly marketing weights are published in the November issue of Agricultural Prices. Beginning in 2003, monthly marketing weights are published in the September issue of Agricultural Prices. Policy parameters for the new farm bill are taken from the legislation (USDA, 2002 c).

Forecast Accuracy

Hoffman (2001) found the mean absolute percentage error generally largest in the beginning of the forecast period but it gradually declined as the forecasts were made later into the crop year, reflecting the availability of more actual information. For example, we first start with planting intentions and yield trends, next actual acreage planted becomes available in NASS's June Acreage Report, next yield estimates are published by NASS in August's Crop Production, followed by monthly production estimates and reports of quarterly stocks. Monthly exports are available from the Census Bureau approximately two months after the month observed.

Hoffman compared both WASDE and the futures model forecasts. For May, the beginning of the forecast period, the mean absolute error was 15 percent for the futures model compared to 14 percent for the WASDE projections (fig. 2). But this percentage error declined for both WASDE and the futures model forecasts to less than one percent for August, the last month of the crop year. Forecast accuracy for season-average prices can be expected to affect the CCP forecasts.

Recent Results for 2002/03 and 2003/04

The futures model provides a weekly season-average forecast of the U.S. corn price received by producers for crop years 2002/03 and 2003/04 (fig. 3 and fig. 4). These price forecasts are used to forecast the CCP rate. Forecasts of the CCP rate for 2002/03 ranged from $0.24 to $0.0/bu. As of early October 2003 USDA announced that corn's counter-cyclical payment rate for crop year 2002/03 would be zero and consequently its payment would also be zero.

As of October 16, 2003, the season-average price for U.S. corn was forecast to be $2.04 for 2003/04 implying a CCP rate of $0.28/bu. (table 3) or a counter-cyclical payment of $2.0 billion (table 4).

Crop Year 2002/03

During the crop year, forecasts of the CCP rate ranged from $0.24/bu. to $0.0/bu. (fig. 3). It is interesting to note that with production uncertainty prices were above the CCP trigger price between late July 2002 and early November. The forecast for the

CCP rate exhibited a fair amount of variability, reflecting variability in the forecast of the season-average price received. However, price forecast variability declined significantly in early November, as more information about the crop size became available.

Season-average price forecasts from the futures model are based on expectations reflected in the futures market and, if available, actual monthly farm prices. The futures model season-average price forecast for 2002/03 started at $2.07/bu. in May of 2002, compared to the WASDE mid-point projection of $1.95/bu. The U.S. 2002/03 corn crop was projected at 9.9 billion bushels, up almost 5 percent from the prior year. Expected supplies were up only slightly because of the smaller expected carryin stocks. Total use in 2002/03 was expected to expand due to gains in industrial use and exports. With use exceeding production, 2002/03 ending stocks of corn were expected down slightly from the forecasted carryin.

However, corn production for 2002/03 was reduced to 9 billion bushels by drought. A 6-percent drop in yield accounted for all of the decline because harvested area was up slightly. Futures model forecasts reflect the uncertainty of the crop size between June and early September as forecasts rose from about $2.15/bu. to about $2.63/bu. in early September. Although price forecasts declined from September to $2.30/bu. in August of 2003, total domestic use is projected at a record and tighter stocks have lead to higher prices than the initial forecast made in May of 2002.

The 2002 Farm Act indicates that advance CCPs shall be made if it is determined that a counter-cyclical payment is required for the crop year. An advance of up to 35 percent could have been made in October of the production year and another 35 percent could have been made the following February with the remainder to be made shortly after the conclusion of the crop year. But there were no advance payments made during 2002/03 most likely because the Department of Agriculture's season-average prices received projection (WASDE) was above the CCP trigger price of $2.32/bu. during these decision periods. Thus, the effective price was greater than the target price during both decision periods for advance CCPs.

At the beginning of the forecast period, May 2002, the producer observed a forecasted CCP rate of $0.24/bu., but this declined to $0.0/bu. Some grain-marketing professionals claim there is a way to protect the CCP in the earlier part of the crop year. Wisner (2003) states that, "While a precise hedge of CCPs is not possible, the risk of losing these payments may be somewhat reduced in time of low futures prices by using a vertical call option spread." Research is being conducted by the Economic Research Service to determine how CCPs affect producers' risk management and crop production.

The futures model also provides a forecast tool for the program analyst. Forecasts of prices received and CCP rates can be used to forecast CCP budget outlays. The CCP's impact on budget outlays for the crop year based on these forecasts would have ranged from $1.7 billion to zero.

Crop Year 2003/04

Forecasts of the CCP rate for crop year 2003/04 have ranged from $0.33/bu. to $0.0/bu. with an October 16, 2003 forecast of $0.28/bu. (fig. 4).

The futures forecast of the season-average price as of May 1, 2003 was $2.17 but rose to $2.32/bu. on fears of planting difficulties. However, prices declined to $1.99/bu. in July as initial indications were of a record large crop. However, these production estimates were reduced in August and the futures forecast of the season-average price was $2.19 as of August 28, 2003. However, production estimates were increased in October and the futures forecast of the season-average price was $2.04/bu. as of October 16, 2003 (fig. 4).

USDA's May 2003 price projection for 2003/04 corn was $2.10/bu., compared to the futures model forecast of $2.32/bu. The futures forecast was significantly higher than the WASDE projection most likely because the market did not believe that this year's crop would achieve the assumed trend yield, thereby including a weather-uncertainty premium.

The USDA outlook for U.S. corn in 2003/2004, as of May 2003, was based on March planting intentions, a recent 3-year average of harvested-to-planted relationships and trend yields. These assumptions provided a supply that exceeded last year's by 5 percent. Total corn use in 2003/2004 was expected to expand due to gains in domestic use and exports. Domestic use was expected to rise slightly as expanding industrial use more than offset reduced feed and residual use because of a decline in cattle on feed. U.S. corn exports were projected up 225 million bushels due to less competition from foreign corn exporters and reduced global feed wheat supplies. Ending stocks were expected to increase by 250 million bushels, as production exceeds use.

However, in August 2003, USDA's expected production reflected acres planted and a yield survey resulting in lower supply and stocks for 2003/04, and use was not expected to decline as much as supply. Thus, USDA's August mid-point price projection rose to $2.20/bu., while the futures forecast rose to $2.19/bu. In contrast, expected production was revised and reached record levels in October and USDA's October mid-point price projection declined to $2.10/bu, while the futures forecast for October 9, 2003 dropped to $2.11/bu.

Conclusions

The futures forecast method is used to forecast a season-average price for U.S. corn on a weekly frequency. The season-average price forecast provides producers and program analysts with information to estimate counter-cyclical payment rates and total counter-cyclical payments for corn. The futures forecasting procedure provides a useful tool for both producers and policy analysts and provides a useful crosscheck with other season-average price forecasts.

References

Becker, Geoffrey S. and Jasper Womach. *Farm "Counter-Cyclical Assistance"* CRS Report for Congress. Congressional Research Service, The Library of Congress. May 31, 2002. 6 pages.

Childs, Nathan W. and Paul C. Westcott. "Projecting the Season Average Price For U.S. Rough Rice." *Rice Situation and Outlook Yearbook.* U.S. Department of Agriculture, Economic Research Service, RCS-1997, December 1997, pp. 18-24.

Danthine, J. "Information, Futures Prices, and Stabilizing Speculation." *Journal of Economic Theory.* 17 (1978): pp. 79-98.

Fama, E.F. "Efficient Capital Markets: A Review of Theory and Empirical Work." Journal of Finance. 25(1970):383-423.

Fama, E.F. "Efficient Capital Markets: II." Journal of Finance. 46(1991):1575-1617.

Gardner, Bruce L. "Futures Prices in Supply Analysis." *American Journal of Agricultural Economics.* 58 (1976): pp. 81-84.

Hoffman, Linwood A. "Forecasting Season-Average Corn Prices Using Current Futures Prices." *Feed Situation and Outlook Report.* U.S. Department of Agriculture, Economic Research Service, FDS-318, May 1991, pp. 24-30.

Hoffman, Linwood A. "Evaluating the Use of Futures Prices to Forecast the Farm Level U.S. Corn Price." Ag Econ Search, July 2001. 24 pages.

Jiang, Bingrong and Marvin Hayenga. "Corn and Soybean Basis Behavior and Forecasting: Fundamental and Alternative Approaches." Unpublished manuscript. Department of Economics, Iowa State University, 1997.

Meyer, Leslie A. "Factors Affecting the U.S. Farm Price of Upland Cotton." *Cotton and Wool Situation and Outlook Yearbook.* U.S. Department of Agriculture, Economic Research Service, CWS-1998, November 1998, pp. 16-22.

Peck, Anne E. "Futures Markets, Supply Response, and Price Stability." *Quarterly Journal of Economics*, 90 (1976): pp. 407-23.

Rausser, G.C., and R.E. Just. "Agricultural Commodity Price Forecasting Accuracy: Futures Markets versus Commercial Econometric Models." *Futures Trading Seminar*, Vol. 6. Chicago: Board of Trade of the City of Chicago, 1979, pp. 117-153.

Tomek, William G. "Commodity Futures Prices as Forecasts." *Review of Agricultural Economics.* Volume 19, Number 1, Spring/Summer 1997, pp. 23-44.

U.S. Department of Agriculture (a), National Agricultural Statistics Service. *Agricultural Prices.* Annual summaries and monthly issues, 1981-2003.

_____ (b). *Crop Production.* December issues, 1981-97.

U.S. Department of Agriculture (c), Economic Research Service. Farm and Commodity Policy Briefing Room. "The 2002 Farm Bill: Provisions and Economic Implications." May2002 a.. http://www.ers.usda.gov/features/farmbill/

U.S. Department of Agriculture (d), World Agricultural Outlook Board. *World Agricultural Supply and Demand Estimates.* Monthly issues, 1981-2003.

U.S. Department of Agriculture (e), Economic Research Service. Farm and Commodity Policy Briefing Room. "Glossary of Policy Terms.". http://www.ers.usda.gov/briefing/WTO/Glossaries.htm

U.S. Department of Agriculture (f). *Fact Sheet: Direct and Counter-cyclical Payment Program* Farm Service Agency, May 2003. http://www.fsa.usda.gov/pas/publications/facts/html/dcp03.htm

U.S. Department of Agriculture (g). Farm Service Agency. Production flexibility contract data for crop year 2001.

U.S. Department of Agriculture (h). Economic Research Service. Farm and Commodity Policy Briefing Room. *"Updating base acres and payment yields."* 3 pages.

Westcott, Paul C. and Linwood A. Hoffman. *Price Determination for Corn and Wheat: The Role of Market Factors and Government Programs*. U.S. Department of Agriculture, Economic Research Service, Technical Bulletin No. 1878, July 1999.

Westcott, Paul, C. Edwin Young, and J. Michael Price. *The 2002 Farm Act: Provisions and Implications for Commodity Markets.* Ag. Inf. Bull. No. 778. USDA/ERS, November 2002.

Wisner, Robert N. "Market Outlook for Feed Grains in 2003-04." Presented at the Midwest, Great Plains and Western Outlook Conference. Indianapolis, Indiana, August 14-15, 2003. 22 pages.

Table 1. Projected Annual Maximum Corn Counter-Cyclical Payments with the 2002 Farm Act, 2003 Crop Year

Maximum Payment Rate		Estimated a/ Payment Acres		Estimated CCP a/ Payment Yield	=	Total Payment
$0.34/bu.	X	69.4 million acres	X	102.6 bushels/acre	=	$ 2.4 Bil.

a/ Base on crop year 2001.
Source: (USDA (g)).

Figure 1. Relationship of U.S. Season-Average Corn Price and Policy Parameters to the Counter-Cyclical Payment Rate

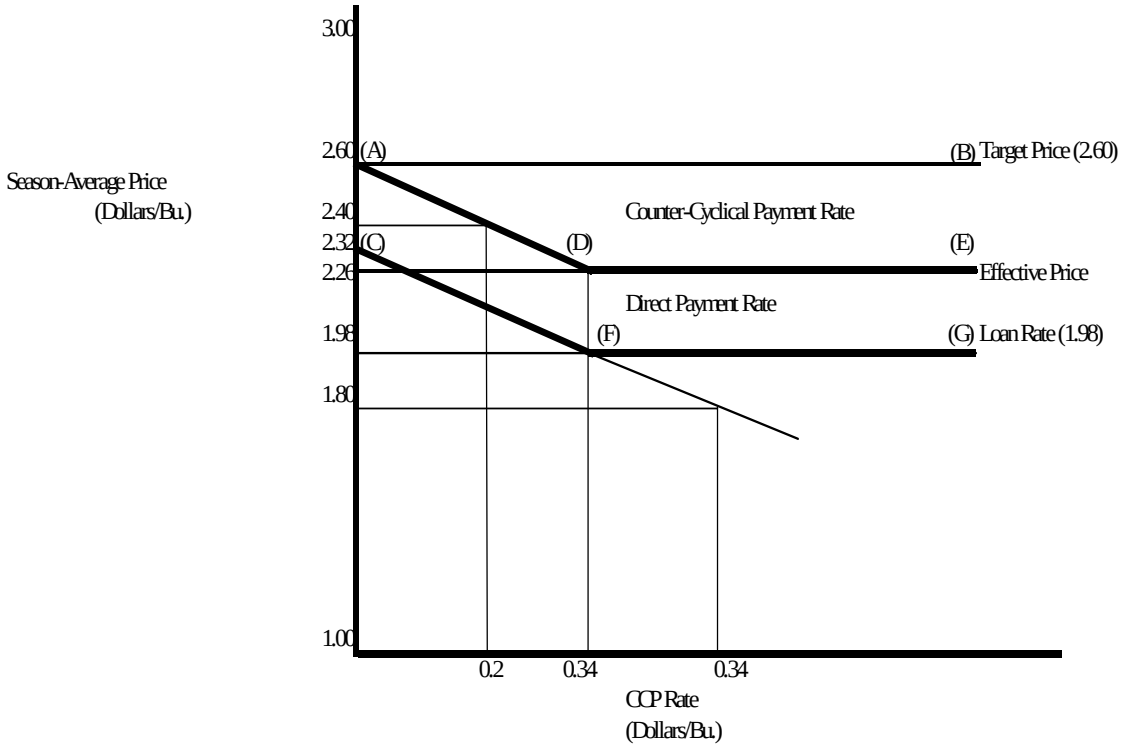


Table 2--Futures Model Forecast of U.S. Corn Producers' Season-Average Price and CCP Rate, Crop Year 2003-2004

	Item	Sept.	Oct.	Nov.	Dec.	Jan.	Feb.	March	April	May	June	July	August	Sept.
								Dollars per bushel						
(1)	Current futures price 1/ by contract (settlement)				2.15			2.22		2.27		2.30		2.32
(2)	Monthly futures price based on nearby contract	2.15	2.15	2.15	2.22	2.22	2.22	2.27	2.27	2.30	2.30	2.32	2.32	
(3)	Plus the historical basis (cash less futures)	-0.26	-0.25	-0.22	-0.20	-0.15	-0.12	-0.16	-0.13	-0.19	-0.17	-0.25	-0.27	
(4)	Forecast of monthly average farm price	1.89	1.90	1.93	2.02	2.07	2.10	2.11	2.14	2.11	2.13	2.07	2.05	
(5)	Actual monthly farm price	2.13												
(6)	Spliced actual/forecast monthly farm price	2.13	1.90	1.93	2.02	2.07	2.10	2.11	2.14	2.11	2.13	2.07	2.05	
(7)	Marketing weights (in percent)	8.46	13.78	10.88	7.14	14.00	6.34	7.26	5.54	5.18	5.66	7.30	8.28	

Forecast of Season-average prices received:

(8)	Weighted average	2.04
(9)	Simple average	2.06

Forecast of the Counter Cyclical Payment Rate (CCP):

(10) Effective price ($ 2.32/bu.) = [Higher of (national average farm price for the marketing year ($2.04/bu) or (national loan rate ($1.98/bu.) + direct payment rate ($0.28/bu.).
CCP Rate ($0.28/bu.) = Target price (2.60/bu.) - Effective price (2.32/bu.).

1/ Contract months include December, March, May, July, and September. Futures price quotation from the Chicago Board of Trade, October 16, 2003 settlement prices.

Table 3. Computation of the counter-cyclical payment rate for U.S. corn, 2002-07

Year	Target Price	-	Effective Price	((Higher of SAP or NALR) + (DP)) = CCP rate.
			Dollars per bushel	
2002	2.60	-	(2.32 [8]	or 1.98) + (.28) = 0.00
2003	2.60	-	(2.04 [8]	or 1.98) + (.28) = 0.28
2004	2.63	-	(?	or 1.95) + (.28) = ?
2005	2.63	-	(?	or 1.95) + (.28) = ?
2006	2.63	-	(?	or 1.95) + (.28) = ?
2007	2.63	-	(?	or 1.95) + (.28) = ?

Table 4. Actual and Forecasted Annual Counter-Cyclical Payments with the 2002 Farm Act

Payment Rate		Payment Acres		Payment Yield	=	Total Payment
Actual for Crop Year 2002/03						
$0.00/bushel	X	69.4 million acres	X	102.6 bushels per acre	=	$ 0.0
Forecast for Crop Year 2003/04						
$0.28/bushel	X	69.4million acres	X	102.6 bushels per acre	=	$ 2.0 Bil.

8 Based on October 10, 2003 WASDE report.

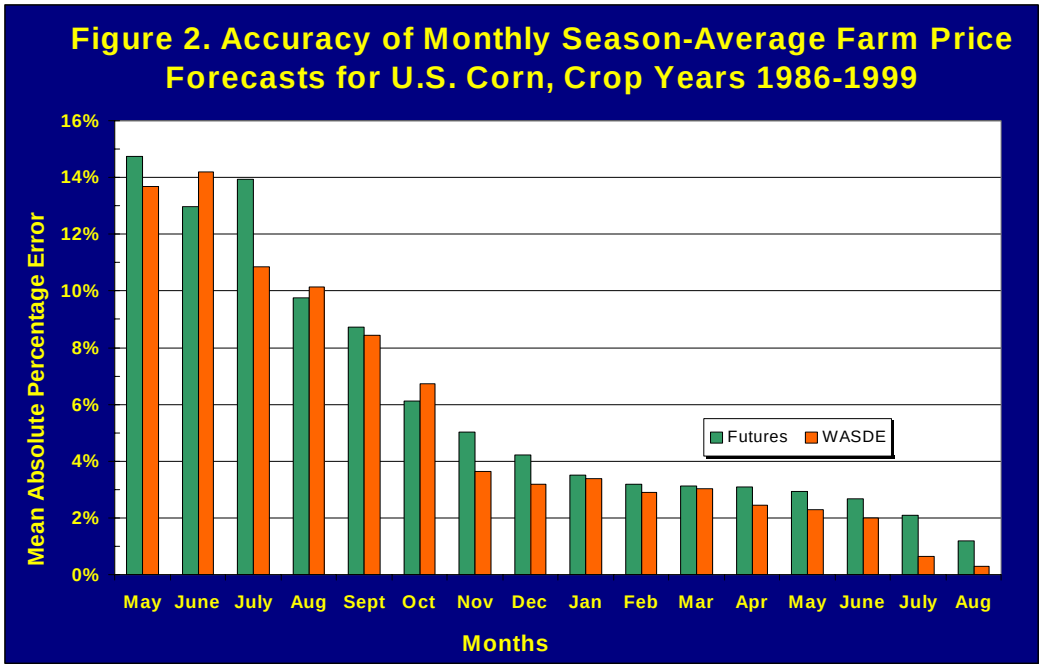


Figure 2. Accuracy of Monthly Season-Average Farm Price Forecasts for U.S. Corn, Crop Years 1986-1999
Mean Absolute Percentage Error
16%
14%
12%
10%
8%
6%
4%
2%
0%
Futures
WASDE
May June July Aug Sept Oct Nov Dec Jan Feb Mar Apr May June July Aug
Months

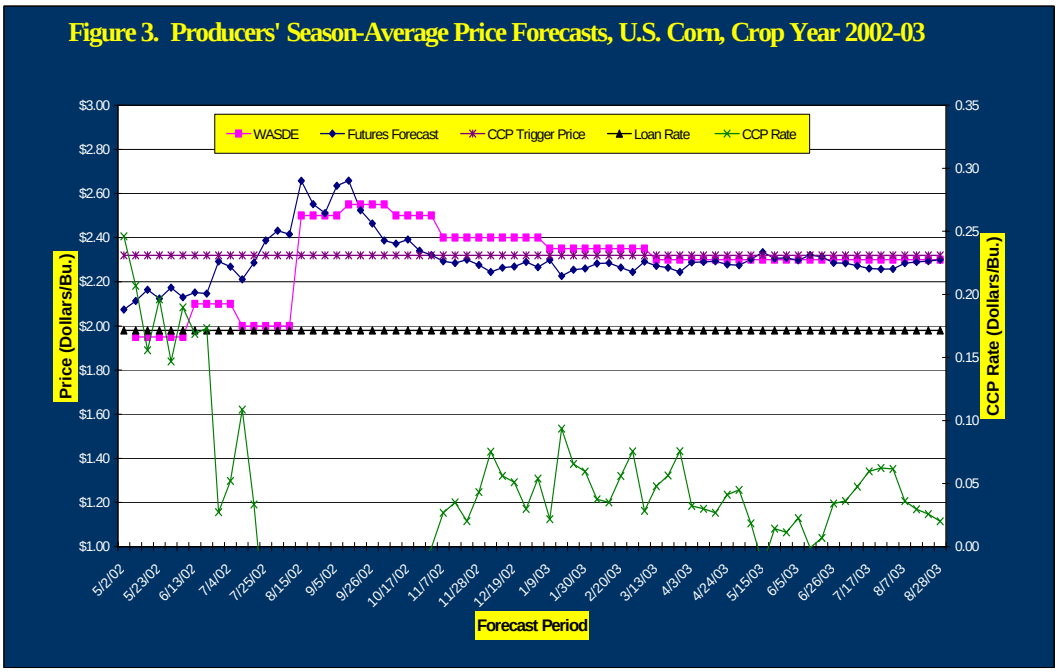


Figure 3. Producers' Season-Average Price Forecasts, U.S. Corn, Crop Year 2002-03
WASDE
Futures Forecast
CCP Trigger Price
Loan Rate
CCP Rate
Price (Dollars/Bu.)
$3.00
$2.80
$2.60
$2.40
$2.20
$2.00
$1.80
$1.60
$1.40
$1.20
$1.00
CCP Rate (Dollars/Bu.)
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
5/2/02 5/23/02 6/13/02 7/4/02 7/25/02 8/15/02 9/5/02 9/26/02 10/17/02 11/7/02 11/28/02 12/19/02 1/9/03 1/30/03 2/20/03 3/13/03 4/3/03 4/24/03 5/15/03 6/5/03 6/26/03 7/17/03 8/7/03 8/28/03
Forecast Period

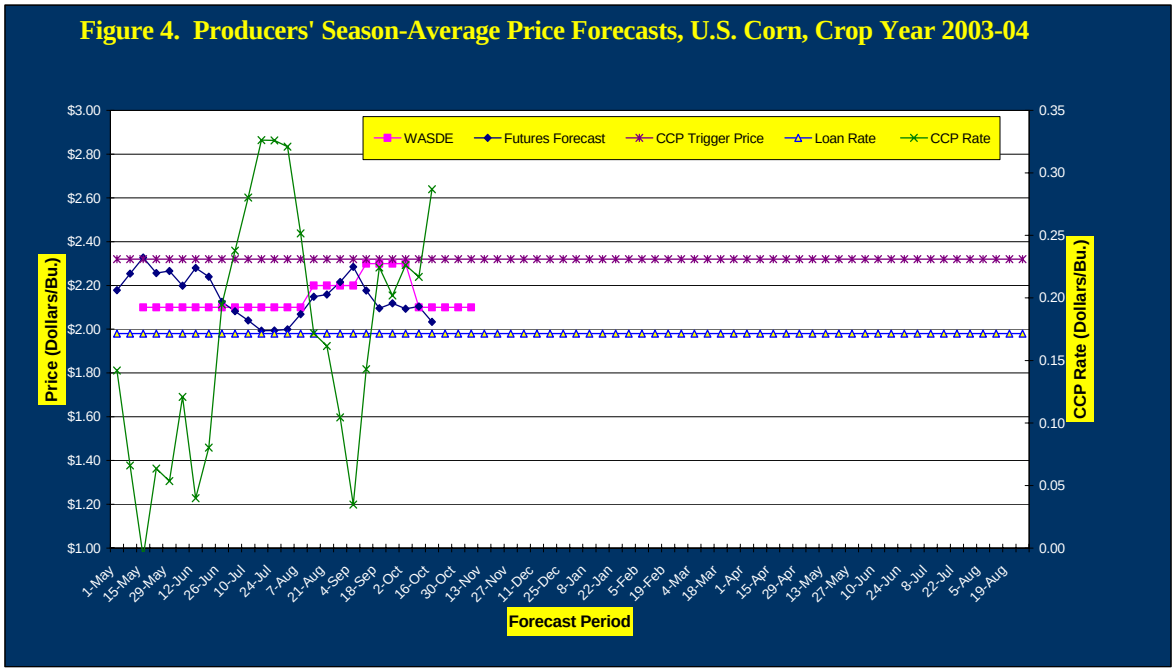


Figure 4. Producers' Season-Average Price Forecasts, U.S. Corn, Crop Year 2003-04
WASDE
Futures Forecast
CCP Trigger Price
Loan Rate
CCP Rate
Price (Dollars/Bu.)
CCP Rate (Dollars/Bu.)
$3.00
$2.80
$2.60
$2.40
$2.20
$2.00
$1.80
$1.60
$1.40
$1.20
$1.00
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
1-May
15-May
29-May
12-Jun
26-Jun
10-Jul
24-Jul
7-Aug
21-Aug
4-Sep
18-Sep
2-Oct
16-Oct
30-Oct
13-Nov
27-Nov
11-Dec
25-Dec
8-Jan
22-Jan
5-Feb
19-Feb
4-Mar
18-Mar
1-Apr
15-Apr
29-Apr
13-May
27-May
10-Jun
24-Jun
8-Jul
22-Jul
5-Aug
19-Aug
Forecast Period

Forecasting Techniques

Chair: Karen S. Hamrick, Economic Research Service, U.S. Department of Agriculture

Loss Functions for Detecting Outliers in Panel Data: An Introduction

Charles D. Coleman, U.S. Census Bureau, U.S. Department of Commerce

The detection of outliers is of critical importance for data quality assurance. An outlier either indicates a problem with its data generation process or is a true statement about the world. This paper illustrates the development and use of loss functions to detect nonparametrically outliers in positive panel data. Positive and negative outliers can be defined separately. In the case of nominal time an exact parametrization of these loss functions is obtained. A time-invariant loss function permits the comparison of data at multiple times on the same basis. A generalization is developed for any real-valued data. Several examples will be discussed.

MARS: An Alternative to Neural Net

Dan Steinberg, N. Scott Cardell, and Mikhail Golovnya, Salford Systems

One of the most effective forecasting tools is linear regression. Linear Regression, however, has a number of shortcomings, including the inability to accommodate highly non-linear relationships, intolerance for missing values, and sensitivity to outliers. This presentation will discuss a non-linear, fully automated regression methodology called MARS (Multivariate Adaptive Regression Splines). MARS was initially designed to address the most challenging of forecasting problems. MARS was developed using regression, splines, and binary recursive partitioning techniques.

Forecasting Waiting-Time for Health Services: Capturing the Nonlinear Dynamics Implied by a Constrained Decision Maker

Trond Jorgensen, Altarum Institute

Altarum Institute is currently studying the problem of long waiting times in health service organizations. One area of importance is forecasting future waiting time given a particular budget constraint. Altarum is attempting to determine the appropriate model among alternative methodologies. Here we look at an approach for forecasting waiting time for health services where we, by the choice of mathematical structure, take into account the existence of a rational decision maker trying to minimize the waiting time by allocating the optimal mix of resources, subject to a budget constraint. The decision maker is modeled as having a profit-maximizing goal. We discuss the mathematical structure of the overall forecasting model when incorporating this prior knowledge and how it differs from other empirical approaches.

LOSS FUNCTIONS FOR DETECTING OUTLIERS IN PANEL DATA: AN INTRODUCTION

Charles D. Coleman, U.S. Census Bureau, 4700 Silver Hill Rd., Stop 8800, Washington, DC 20233-8800
Email: ccoleman@census.gov

1. Introduction

In assuring data quality in forecasting, one would like to know that the data generation processes are free from anomalies. One interpretation of this is that the data do not have unexplainable outliers. In general, an outlier is an observation which departs from the norm (however defined) in a set of observations. Outliers can indicate problems with their data generation processes (i.e., anomalies) or may be true, but unusual, statements about reality.[1] In terms of Barnett and Lewis (1994, p. 37), we are testing for discordancy. This paper specializes the problem of detecting outliers to panel data, such as estimates and forecasts. Panel data are cross-sectional time series, such as a time series of population estimates for a set of areas.[2] Time may be either chronological or nominal. Nominal time indexes different sets of predictions (i.e., estimates or forecasts) for the same cross-sectional units and chronological time. Time is nominal in this context because the different predictions sets have no natural ordering. Comparing cross-sectional estimates to their true values is an instance of nominal time. The method this paper uses is to develop loss functions to identify discordant observations for further analysis. The loss functions are developed for panels of two dates and then extended to panels with arbitrary numbers of observations with arbitrary differences between dates.

Initially, the data are assumed to be positive.[3] In this context, the subject matter analyst's judgment is needed to determine the exact parametrization of the loss function, except for the special case described in Subsection 2.4.[4] The exact parametrization thus depends on the subject matter analyst and context. It is, thus, subjective. When the data can take on any real value, mathematical considerations dictate the exact parametrization.

The Population Division of the U.S. Census Bureau has been successfully using loss functions to detect outliers in the preparation of population estimates and geographic base files. Loss functions have been applied to input, intermediate and final data. Rather than use actual data, a numerical example illustrates how loss functions are used and how they avoid the pitfalls associated with taking numerical and percent differences. A map illustrates the use of loss functions with GIS and provides an illustration of the need for subject matter analyst expertise.

Section 2 develops loss functions for positive data. No distributional assumptions are made, as the natures of the data generation processes are assumed unknown and nonidentical.[5] Thus, this is an example of the nonparametric approach to outlier detection.[6] An important upshot of this approach is that data from a wide range of values are put on the same basis. This Section specifies the assumptions and develops the simplest loss function that satisfies these assumptions. Loss functions are developed for more general settings. Section 3 discusses some applications, including general usage of loss functions, parametrizing loss functions from preexisting outlier criteria and using loss functions with GIS. These examples are based on actual Census Bureau applications. Section 4 generalizes the framework to data that can take any real value. Section 5 concludes this paper.

2. The Loss Function[7]

This section describes the assumptions used to generate the loss function $L(F;B)$ and its variants, where F is the future value and B is the base period value. The loss function is the penalty, cost, or "badness" associated with the difference between F and B. Roughly speaking, the greater the difference between F and B, the greater the loss. Initially, F is assumed to be one period after B. After the necessary assumptions are made, the simplest form of L is specified. Restrictions on the values of the parameters of L which make it increase in B for a given relative difference are then specified. Subsection 2.1 axiomatically develops the simplest unsigned loss function L which satisfies these properties for data exactly

[1] This is similar to Hoaglin's (1983, p. 39-40) use of "outside cutoffs" to identify "outside values."
[2] The bidimensionality of data searched for outliers is not unique: DuMouchel (1999), Albert (1997) and Rudas, Clogg and Lindsay (1994) search for outliers in contingency tables. The contingency table approach differs in that time need not be a dimension and that parametric assumptions are made.
[3] Zeroes are permissible by adding a small constant, as discussed in Section 2 below.
[4] The subject matter analyst's judgment may already be incorporated in discrete outlier criteria. See Subsection 3.2.
[5] This obviates the use of parametric techniques, in which observations are tested for departure from a predetermined, hypothesized distribution.
[6] Barnett and Lewis (1994, pp. 107, 364-365) provide some references to nonparametric approaches in other contexts. Tukey (1977) proposed perhaps the most familiar nonparametric technique for detecting univariate outliers: the boxplot or box-and-whiskers plot. Rouseeuw, Ruts and Tukey (1999) propose the bagplot, a bivariate generalization of the boxplot.
[7] This exposition is based on Coleman, Bryan and Devine (2003, Section 2).

one period length apart. Subsection 2.2 generalizes L to situations in which F and B may not be exactly one period apart. Subsection 2.3 introduces the signed loss function for cases in which the sign of the difference is an additional important criterion. Subsection 2.4 parametrizes L for comparing two sets of estimates of the same parameters. Throughout this paper, B and F are assumed positive. Zeroes, which frequently arise in practice, are either recoded to small values or omitted from the analysis.

2.1 The Unsigned Loss Function

The unsigned loss function L is constructed by specifying three assumptions. The first assumption is that L is symmetric in the differences:

Assumption 1 (symmetry): $L(B+\varepsilon;B) = L(B-\varepsilon;B)$ for all B, $\varepsilon > 0$.

This assumption is not as innocuous as it looks. It is quite possible that, at least for some range of B, that positive and negative differences have differential impacts. However, the resulting asymmetry complicates the definition of L. Subsection 2.3 relaxes this assumption by developing the signed loss function, which allows the possibility of asymmetrically incorporating the direction of the difference ε. The symmetry of L allows us to use the equivalent notation $\lambda(\varepsilon, B) \equiv L(F, B)$ where $\varepsilon = |F - B|$.

The next assumption makes L, or, equivalently, ℓ, increasing in the difference ε:

Assumption 2 (monotonically increasing in difference): $\partial\lambda/\partial\varepsilon > 0$ for all $\varepsilon > 0$.

Note that this assumption is stated in terms of ℓ, rather than L. This assumption is quite intuitive, as it states that smaller differences are preferred to larger ones.

Finally, we want L, or, equivalently, ℓ, to decrease in B. This means that for a given value of ε, the loss associated with it decreases with its associated initial value. This has two justifications. First, for example, a difference of 500 when the initial value is 1,000 is a whopping 50%, a highly significant difference. However, the same difference, when the initial value is 1,000,000 is akin to a roundoff error. Second, when performing estimates or taking samples, the coefficient of variation, σ^2/μ^2, where σ^2 is the variance and μ is the expected value, decreases in B. This author's experience is that all areas tend to have about the same roundoff errors. Again, these are proportionately greater in small areas. We state this formally as:

Assumption 3 (monotonically decreasing in base value): $\partial\lambda/\partial B < 0$, or, equivalently $\partial L/\partial B < 0$, for all $B > 0$.

This simplest function which satisfies Assumptions 1–3 and admits Property 1 below is the Cobb-Douglas function[8]

$$L(F;B) = |F - B|B^q \qquad (1a)$$

or, equivalently,

$$\lambda(\varepsilon, B) = \varepsilon B^q \qquad (1b)$$

where $\varepsilon > 0$ and $q < 0$.[9]

An observed pair $(F;B)$ is an outlier whenever $L(F;B) > C$, where C is a predetermined critical value.[10] We will also refer to outliers as being critical. Additionally, we will refer to the equation $L(F;B) = C$ as the equation of criticality. The choice of q and C is an empirical matter.[11] Only a practitioner's experience with data can determine when data are suspect and incorporate these suspicions into parameters. One thing to note is that the loss function is ordinal: raising L and C to any positive power m leaves the rankings of losses unchanged.[12] It is only the rankings of losses that are important.[13] Another important quality is that loss is not necessarily interpretable. This is generally true of loss functions (Lindley, 1953, p. 46).

A desirable property of the loss function is that it increases in B for a given absolute relative difference. The absolute relative difference is:

$$|F - B|B^{-1} \qquad (2)$$

Note that, in this case, $q = -1$. Choosing $q > -1$ makes the loss function increase in B, for a given absolute relative difference. We state this as Property 1:

Property 1: The loss function defined by equations (1a) and (1b) increases in B for any given absolute relative difference. This is assured whenever $q > -1$.

The reader may note that $q = 0$ turns equations (1a) and (1b) into the absolute values of the differences. Thus, values of q between 0 and –1 represent various

[8] It should be noted that an infinite number of loss functions satisfy Assumptions 1-3 and admit Property 1. This one is merely the simplest.

[9] Unlike Coleman (2000, 2002, 2003), no exponent on the difference is needed due to a Lie symmetry. See Coleman, Bryan and Devine (2003) for the explanation.

[10] Alternatively, C can also be determined from the data by taking a predetermined quantile or a multiple of the interquartile range of L (Tukey 1977).

[11] Subsection 2.4 below investigates a case in which q can be determined exactly.

[12] This is at the heart of the Lie symmetry noted in footnote 9.

[13] This is similar to the economic concept of ordinal utility. Coleman (2000, 2002, 2003) differs in using a cardinal framework: the values of the loss function can be compared to each other and operated upon arithmetically.

tradeoffs of absolute differences and absolute relative differences. Consider the product of the *r*th power of the absolute difference and the *s*th power of the absolute relative difference, where $r, s > 0$: $|F - B|^r \left(|F - B|B^{-1}\right)^s$. By the Lie symmetry invoked in footnote 9, this function is isomorphic to the loss function $|F - B|B^{-\frac{s}{r+s}}$. Thus, any value of q corresponds to an infinite number of pairs (r, s) where $q = -s / (r + s)$. Geometrically, the same loss function is generated for all (r, s) lying on the line $r = -(1 + q)\, s$.

2.2 The Time-Invariant Loss Function

Instead of considering the single set of future data, $\mathbf{F} = \{F_i\}_{i=1}^n$, where i indexes the n observations, consider the sets $\mathbf{F}_t = \{F_{it}\}_{i=1}^n$, where t is the amount of time elapsed since the base date and i indexes the cross-sectional units. We wish to develop a loss function which allows us to make comparisons across time on the same basis, by explicitly incorporating t into the loss function. One way of incorporating time-invariance is to substitute the geometric average absolute relative change

$$\left(\frac{|F_{it} - B_i|}{B}\right)^{1/t} \tag{3}$$

for the absolute relative change implicit in equation (1a) to create the time-invariant loss function[14]

$$L(F_{it}; B_i, t) = |F_{it} - B_i| B_i^{tq+t-1}. \tag{4}$$

Given this paper's framework, equation (4) should be used to make comparisons across time, as it puts the geometric average absolute relative difference on the same basis for all t. The reader can verify that $-1 < tq + t - 1 < 0$ for $t > 0$ and $0 > q > -1$.

2.3 The Signed Loss Function

At times, not only is the value of the loss function important, but also the sign of the difference. Different outlier generation processes may manifest themselves by producing predominantly positive or negative differences. We can account for these by creating the signed loss function S, which is simply the loss function L, multiplied by the signum function of the difference:

$$S(F; B) = |F - B|B^q \operatorname{sgn}(F - B) = (F - B)B^q \tag{5}$$

where $\operatorname{sgn} x = +1$ for $x > 0$, 0 for $x = 0$, and -1 for $x < 0$.

[14] For details, see Coleman, Bryan and Devine (2003), Subsection 2.3.

Using S, one can create different critical values for loss, depending on whether the difference is positive or negative. To wit, one can pick C_+, C_-, $C_+ \neq -C_-$, such that a pair $(F; B)$ is declared an outlier if either $S(F;B) < C_-$ or $S(F;B) > C_+$. Again, the choice of whether to use S and then use asymmetric critical bounds is an empirical matter.[15] For example, since, by assumption, negative values of F are impossible, then asymmetric critical bounds and/or parameters may be necessary to detect cases in which F becomes very small relative to B.

The time-invariant signed loss function is

$$S(F_{it}; B_i, t) = (F_{it} - B_i)B_i^{tq+t-1}. \tag{6}$$

2.4 Comparing Two Sets of Data: A Specialization of the Loss Function

Often, one is interested in comparing two sets of estimates of the same cross-sectional units. Suppose that the sets $\mathbf{B} = \{B_i\}$ and $\mathbf{F} = \{F_i\}$ represent two versions of estimates of the true values $\mathbf{A} = \{A_i\}$. This is an instance of nominal time. Suppose that both the B_i and F_i are unbiased estimators of the A_i and that their variances are proportionate to the A_i (i.e., $\operatorname{Var}(B_i) = \operatorname{Var}(F_i) = \sigma^2 A_i$.) One way one can think of this situation as that both B_i and F_i are constructed summing A_i jointly uncorrelated random variables with mean 1 and variance σ^2.[16] In this situation, we can use the loss functions (1a) and (1b) with $q = -½$. Since the null distributions of **B** and **F** are assumed unknown, it is impossible to do any significance testing. Moreover, since we are usually dealing with the entire population, sampling theory is not appropriate.

Of course, if the processes generating **B** and **F** are not as assumed, no theoretical guidance is available for the choice of q.

Again, the signed loss function (5) can be used with $q = -½$.

3. Applications

This section illustrates the use of loss functions by first outlining a general procedure for using loss functions in Subsection 3.1. Next, three different examples of loss functions are shown. In the first example, in Subsection 3.2, preexisting outlier criteria in terms of critical ratios by size class are transformed into a loss function. The second example, in Subsection 3.3, uses real-world data and GIS to compare two sets of real-

[15] The asymmetry need not be limited to the critical values. The signed loss function can incorporate different values of q, depending on the sign of the difference.

[16] Note that independent, identically distributed variables are a special case of this assumption.

world estimates using the $q = -½$ loss function of Subsection 2.4. The results of using absolute and absolute relative differences to evaluate differences between these two sets of estimates are discussed for comparison. Coleman et al.'s (2003, Subsection 3.4) method of using a reference variable to detect outliers is not discussed.

3.1 General Procedure for Using Loss Functions

Loss function evaluations usually begin by recoding zero base values to a small positive value,[17] (the exact value determined by the range and smallest value of the data and smaller than the smallest value) and setting $q = -0.5$. If time is chronological, the subject matter analyst then has to examine the data and the rankings of their associated losses.[18] If, in the subject matter analyst's opinion, too many observations with small changes occurring to small base values are ranked highly, then q should be increased.[19] If, on the other hand, too many observations with small changes to large base values are ranked highly, then q should be decreased. This process continues until the analyst is satisfied with the loss rankings. This author has found that changing q by increments of .1 is satisfactory. Finer increments appear to have little effect.

3.2 Creating Loss Functions From Discrete Outlier Criteria

Sometimes, discrete outlier criteria have already been developed. These discrete outlier criteria can be converted into a loss function using regression. Given a set of critical pairs (ε, B), the regression

$$\log \varepsilon = -q \log B + K + \text{error} \qquad (7)$$

is estimated. q is immediately obtained from equation (7). C is then obtained as $C = e^K$.

Often, outlier criteria do not come in discete pairs. Instead, they come in ranges $\left[\underline{B}, \overline{B}\right]$ for which an outlier is declared whenever ε / B exceeds a prescribed value. Coleman et al. (2003, Subsection 3.3) recommend using the midpoints of these ranges to form the pairs (ε, B). If an unbounded uppermost range is present, its lower bound is used.

A further complication is that the outlier criteria may be inconsistent with the assumptions used to develop a loss function. For example, two different ranges may have the same minimum ε, thereby violating Assumption 3. In these cases, the offending ranges have to be either modified or removed. They may be modified if a developer of outlier criteria can be queried to produce satisfactory criteria. If this is not possible, these ranges must be omitted from regression (7).

3.3 A Numerical Example

Table 1 presents an example of two cross-sectional series, their absolute differences and their absolute percent differences and loss functions with $q = -0.5$ using Column 'B_i' as the base. These data are presented in increasing order of B_i (or, equivalently, F_i). Normally, the data are presented to the subject matter analyst in decreasing order of loss (or absolute difference or absolute percent difference).

Table 1
Numerical Example of Loss Functions

i	B_i	F_i	Absolute Difference	Absolute Percent Difference	Loss
1	1	2	1	100	1.00
2	100	105	5	5	0.50
3	500	525	25	5	1.12
4	600	624	24	4	0.98
5	700	735	35	5	1.32
6	1000	1040	40	4	1.26
7	10000	10100	100	1	1.00

Note that the absolute difference is increasing in B (and, equivalently, in F.) If one were to use absolute difference as the measure of "outlierhood," one would generally find that the observations with the largest base values are the most likely to be outliers. Conversely, focusing on the percent absolute differences would cause the observations with the smallest base values to generally be classified as outliers. The extreme case of this is shown in the first row of Table 1. The pair (1, 2) has an absolute percent difference of 100%. Yet, in many contexts, this difference is meaningless. For example, one data source may show one birth in a county, while another shows two. If a component method is used to estimate population in that county, the two data sources will produce a difference of exactly one person. This difference is generally meaningless. For example, the difference between population estimates of 10,000 and 10,001 is meaningless, falling well within the overall error of the estimates.

The loss function effectively trades off the

[17] In some instances, this step should be omitted, as it can cause spurious identification of true zeroes as outliers. Only examination of the results can determine whether this is the case.
[18] The same can be done in nominal time. If the assumptions of Subsection 2.4 are violated, then no particular value of q is prescribed.
[19] That is, q is made closer to zero, say, –0.4.

absolute and absolute percent differences.[20] The large absolute percent difference in row 1 is severely downweighted by its small absolute difference. Likewise, the last row has a large absolute difference, but small absolute percent difference. These two cases have the same loss.

Rows 5 and 6 have similar loss. Because loss is ordinal, no meaning can be placed on this difference, other than row 5 is "worse" than row 6. Instead, the subject matter analyst examines the data process generating row 5 before examining row 6. If, in his opinion, the losses are not properly reflecting the severity of the outliers, the loss functions should be recomputed with a different value of q.

3.4 An Example Using GIS

Geographic information systems can be used with loss functions to find outliers. GIS is particularly helpful for finding geographic patterns in outliers. Map 1 at the end of this paper shows the $q = -1/2$ loss function applied to two different sets of county population estimates.[21] This is an example of nominal time. The base population is the Vintage 1998 published number obtained by the "tax method" component change model.[22] The comparison population is the county household population implied by the subcounty population estimates system, including overrides,[23] before constraining to any higher level totals.[24,25] Southern California, the Dallas-Fort Worth Metropolitan Area, northern Nevada and northern Maine stand out, among others. Most of the counties in the Great Plains that stood out on a map of absolute percentage differences[26] no longer stand out. This is because their populations are very small. Other areas stand out which do not appear on maps of absolute and absolute percent differences include the outer suburbs of Detroit and the Denver area. Northern Maine and Nevada have large enough populations to make their percentage changes stand out. In the cases of Southern California and Dallas-Fort Worth, the populations are so large that small percentage changes create large losses. This may lead the subject matter analyst to conclude that a different value of q should be used. In the other cases, it is the combination of moderate population bases and moderate percentage changes that causes high loss. In any case, the interpretation of the losses is clear: high losses indicate large divergences between the two methods. It is these areas upon which an analyst should focus his attention. By varying q and examining maps and ranked lists of outliers, the analyst can obtain an appropriate value of q, which yields the greatest information about the outliers.

4. Extending the Loss Function to All Real Pairs[27]

Sections 1 through 3 developed a loss function to find outliers in positive data. In many cases, however, data can take on any real value, such as the Census Bureau's net migration data. Thus, the arguments to the loss function are a real pair. For this problem, a new set of assumptions is required. An important difference is that the parameter q is no longer adjusted as a result of subject matter analyst's review. Instead, geometric considerations dictate the choice of q. Another difference is that the assumptions involved become more elaborate. The Census Bureau has used this loss function to find outliers in raw net migration data.

Subsection 4.1 axiomatically develops the simplest unsigned loss function L. Subsection 4.2 develops the signed loss function, similar to that developed earlier. Subsection 4.3 uses geometry to determine q.

4.1 The Unsigned Loss Function

The unsigned loss function L is constructed by making five assumptions. The first assumption is that L is defined everywhere in the real plane $\Re^2$:

Assumption 4 (unrestricted domain): For all $(F, B) \in \Re^2$, $L(F, B)$ is defined and single valued.

The next assumption is that L is symmetric in the difference between B and F:

Assumption 5 (symmetry in difference): $L(B+\varepsilon; B) = L(B-\varepsilon; B)$ and $L(F, F+\varepsilon) = L(F, F-\varepsilon)$ for all B, F and $\varepsilon \in \Re$.

Like Assumption 1, this assumption is not as innocuous as it looks. It is quite possible that, at least for some ranges of B and F, that positive and negative differences have differential impacts. However, the resulting asymmetry

[20] The discussion in the last paragraph of Subsection 2.1 formally demonstrated this.

[21] Counties with "no data" on this map are those which have no subcounty geography per the Census Bureau's Population Estimates Branch's definitions.

[22] These are contained in the Census Bureau's file *98C8_00.txt*, which was released to the public in 1999.

[23] The overrides, or administrative changes, consist of numbers obtained by special censuses, challenges and other corrections to the initial estimates.

[24] In terms of Section 2, the published populations are the B_i and the subcounty estimate-derived data are the F_i.

[25] The subcounty estimates methodology may be found at *http://www.census.gov/population/methods/e98scdoc.txt*.

[26] Coleman et. al (2003) Map 2. Map 1 of that paper displays absolute differences.

[27] This Section is based on Coleman and Bryan (2003).

complicates the definition of L. Subsection 4.2 relaxes this assumption somewhat by developing the signed loss function, which allows the possibility of incorporating the direction of the difference ε. However, as Subsection 4.2 states, this relaxation only affects the critical values used.

A desirable property is that L be symmetric with respect to its arguments. To give a concrete example, we want $L(-1,1000) = L(1000, -1)$. This stated formally as Assumption 6:

Assumption 6 (symmetry in arguments): $L(B, F) = L(F, B)$.

At this point, it useful to introduce some new notation. Let $X=|F|$ and $Y=|B|$. Let the new loss function $\lambda(\varepsilon,\Sigma) \equiv L(F,B)$, where $\varepsilon = |F-B|$ and $\Sigma = \Sigma(X,Y)$ is a function such that $\partial\Sigma/\partial X > 0$ and $\partial\Sigma/\partial Y > 0$. Assumption 6 implies that $\Sigma(X,Y) = \Sigma(Y,X)$, so that Σ is symmetric in its arguments. The remaining Assumptions are stated in terms of λ.

Assumption 2 of Section 2 is repeated to make λ (and L) increase in the difference ε:

Assumption 2 (monotonically increasing in difference): $\partial\lambda/\partial\varepsilon > 0$ for all $\varepsilon \geq 0$.

Finally, we want to create an assumption analogous to Assumption 3 of Section 2 to make λ to decrease in Σ, for similar reasons. We state this formally as:

Assumption 7 (monotonically decreasing in arguments): $\partial\lambda/\partial\Sigma < 0$ for all $\Sigma > 0$.

This simplest function which satisfies Assumptions 2 and 4–7 is (after invoking a Lie symmetry)[28]

$$\lambda(\varepsilon,\Sigma) = \begin{cases} \varepsilon\Sigma^q & \Sigma \neq 0 \\ 0 & \Sigma = 0 \end{cases} \tag{8}$$

where $q < 0$. Note that equation (8) is stated in terms of ε and Σ. The simplest form of Σ will be determined in equation (9) below. Theorem 1 of Coleman and Bryan (2003) shows that setting $\lambda(0,0) = 0$ makes λ continuous at (0,0), when $q > -1$. This way of determining $\lambda(0,0)$ avoids division by 0.

4.1.1 Determination of Σ and L

From equation (1), it is clear that $\lambda(0,\Sigma) = 0$ for all $\Sigma > 0$. We would like to define Σ so that whenever either X or $Y \neq 0$, $\Sigma > 0$. We would also like $\Sigma(0,0) = 0$. The simplest equation for Σ is:

$$\Sigma(X,Y) = X + Y = |B| + |F| \tag{9}$$

From equation (9) we can determine L to be

$$L(F,B) = \begin{cases} |F-B|\left(|F|+|B|\right)^q & B \text{ or } F \neq 0 \\ 0 & B = F = 0 \end{cases} \tag{10}$$

A desirable property of the loss function is that it rises in $|F - B|$ for a given average absolute percentage difference. The average absolute relative difference is defined as:[29]

$$|F-B|\left(|F|+|B|\right)^{-1} \tag{11}$$

Note that, in this case, $q = -1$. Choosing $q > -1$ makes the loss function rise in $|F| + |B|$, for a given average absolute relative difference. This is also required by Theorem 1 of Coleman and Bryan (2003). We state this as Property 1′:

Property 1′: The loss function defined by equations (5) increases in $|F| + |B|$ for any given average absolute percentage difference. This is assured whenever $q > -1$.

The reader may note that $q = 0$ turns equation (10) into the absolute values of the difference. Thus, values of q between 0 and –1 represent various tradeoffs between the absolute value of the difference and average absolute percentage difference. Consider the product of the rth power of the absolute difference and the sth power of the average absolute relative difference, where $r, s > 0$: $|F-B|^r \times \left(\frac{|F-B|}{|F|+|B|}\right)^s$. By Lie symmetry, this function is isomorphic to the loss function $|F-B|\left(|F|+|B|\right)^{-\frac{r}{r+s}}$. Thus, these intermediate values of q correspond to an infinite number of pairs (r, s) where $q = -r / (r + s)$. Geometrically, the same loss function is generated for all pairs (r, s) lying on the line $s = -(1 - 1/q)\, r$.

4.2 The Signed Loss Function

Again, we create the signed loss function S, which is again simply the loss function L, multiplied by the signum function of the difference:

$$S(F,B) = \begin{cases} |F-B|\left(|F|+|B|\right)^q \operatorname{sgn}(F-B) & B \text{ or } F \neq 0 \\ (F-B)\left(|F|+|B|\right)^q & \\ 0 & B = F = 0 \end{cases} . \tag{12}$$

Using S, one can create different critical values for loss, depending on whether the difference is positive or

[28] It should be noted again that an infinite number of loss functions satisfy Assumptions 1-3. This one is merely the simplest.

[29] This is obtained by taking the average of absolute relative differences formed with B and F in the denominators: $|F-B||B|^{-1}$ and $|F-B||F|^{-1}$ and assuming that $B \approx F$.

negative, similar to Subsection 2.3. Again, one can pick C_+, C_-, $C_+ \neq -C_-$, such that a pair (F, B) is declared an outlier if either $S(F;B) < C_-$ or $S(F;B) > C_+$. Again, the choice of whether to use S and then use asymmetric critical bounds is an empirical matter.[30] However, since S has been developed using strong symmetry assumptions, using asymmetric bounds is probably not worthwhile for detecting outliers. The next Section relies on geometric analysis of S to suggest the best choice for q.

4.3 Choice of Loss Function

The loss functions L and S exhibit wildly different behaviors depending on the value of q. The choice of q requires examination of plots of S for various values of q, $-1 \leq q \leq 0$, to obtain a reasonable loss function.[31] The limiting functions when $q = 0$ and $q = -1$ are of particular interest. $q = 0$ implies that $S(F,B) = F - B$. This defines a plane in $\Re^3$, which is not useful for outlier detection in this paper's framework. Setting $q = -1$ produces some strange behavior. Whenever B and F are of opposite signs, $S(F,B) = \text{sgn } F$. This can be seen by substituting $q = -1$ into equations (12) when B or F is nonzero:

$$S(F,B) = (F - B) / \left(|F| + |B|\right) \tag{13}$$

Noting that $|x| = x$ when $x > 0$ and $|x| = -x$ when $x < 0$, we can examine the behavior of S when B and F are of opposite signs. When $F > 0$ and $B < 0$, equation (13) becomes

$$\begin{aligned} S(F,B) &= (F - B) / \left(|F| + |B|\right) \\ &= \left[|F| - \left(-|B|\right)\right] \Big/ \left(|F| + |B|\right) \\ &= \left(|F| + |B|\right) \Big/ \left(|F| + |B|\right) = 1 = \text{sgn } F \end{aligned} \tag{14}$$

The reader may verify that $S(F,B) = -1 = \text{sgn } F$ when $F < 0$ and $B > 0$. These equalities easily generalize to the cases in which either B or F is zero.

Another problem occurs at the origin when $q = -1$: from the previous paragraph we can observe that S simultaneously acquires the values ± 1, which contradicts the assumption that S is single-valued.[32]

Finally, cusps exist along the axes for every $q < 0$, but are most severe for $q = -1$.[33]

Given all of the anomalies and degeneracies associated with this family of loss functions, the problem is to decide on a value of q which produces reasonable behavior, in his mind. It appears that intermediate choices of q are best behaved: these offer a good compromise between simply taking the difference between F and B ($q = 0$) and the bizarre behavior of S when q approaches -1. In particular, the value $q = -0.5$ shows the best tradeoff of the different attributes. Thus, the recommended unsigned loss function is

$$L(F,B) = \begin{cases} |F - B| \left(|F| + |B|\right)^{-0.5} & B \text{ or } F \neq 0 \\ 0 & B = F = 0 \end{cases} \tag{15}$$

with the corresponding signed loss function

$$S(F,B) = \begin{cases} (F - B) \left(|F| + |B|\right)^{-0.5} & B \text{ or } F \neq 0 \\ 0 & B = F = 0 \end{cases}. \tag{16}$$

Again, note that no subject matter analyst's judgment is used to parametrize these loss functions. Instead, the parametrization is based on an evaluation of the geometry of these functions.

6. Conclusion

This paper has used time as an explicit dimension in constructing loss functions for detecting outliers in panel data. Loss functions put all differences on the same basis so that data ranging several orders of magnitude can be compared. When the data are positive, interaction with the subject matter analyst is necessary to properly parametrize the loss function. When the data can assume any real value, geometric considerations dictate the parametrization of the loss function. Some examples have been provided.

7. Acknowledgements

This paper is based on Coleman et al. (2003) and Coleman and Bryan (2003). I would like to thank Mary H. Mulry for helpful comments and peer review.

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.

References

Barnett, Vic and Lewis, Toby (1994). *Outliers in*

[30] The asymmetry need not be limited to the critical values. The signed loss function can incorporate different values of q, depending on the sign of the difference. However, as Subsection 3.2 shows, there is little latitude in the choice of q.

[31] This is done in Coleman and Bryan (2003). This is a different sort of subjectivity than that of Section 2. There, the coefficient q is determined empirically, often from the data. In this Section, the subjectivity lies in the choice of the form of the loss function.

[32] This argument does not even consider approaching the origin along rays in the positive and negative orthants, which may produce yet other values for S.

[33] These can be seen in Coleman and Bryan (2003, Figures 3-11).

Statistical Data, 3rd edition, John Wiley & Sons, New York.

Coleman, Charles D., (2000). "Evaluating and Optimizing Population Projections Using Loss Functions," *Federal Forecasters Conference 2000: Papers and Proceedings*, Washington: U.S Department of Education, Office of Educational Research and Improvement, 27-32.

Coleman, Charles D., (2002). "Optimizing Population Projections Using Loss Functions When the Base Populations are Subject to Revision," *Federal Forecasters Conference 2002: Papers and Proceedings*, Washington: U.S Department of Education, Office of Educational Research and Improvement, 27-32.

Coleman, Charles D. (2003). "Loss Functions for Assessing the Accuracy of Cross-Sectional Predictions," manuscript, U.S. Census Bureau.

Coleman, Charles D. and Bryan, Thomas (2003). "Loss Functions for Detecting Outliers in Panel Data when the Data May Change Sign," manuscript, U.S. Census Bureau.

Coleman, Charles D., Bryan, Thomas and Devine, Jason (2003). "Loss Functions for Detecting Outliers in Panel Data," manuscript, U.S. Census Bureau.

DuMouchel, William (1999). "Bayesian Data Mining in Large Frequency Tables, With an Application to the FDA Spontaneous Reporting System," *The American Statistician* **53**, 177-188.

Hoaglin, David C. (1983). "Letter Values: A Set of Selected Order Statistics." In, Hoaglin, David C., Frederick Mosteller and John W. Tukey [eds.], *Understanding Robust and Exploratory Data Analysis*, Wiley, New York.

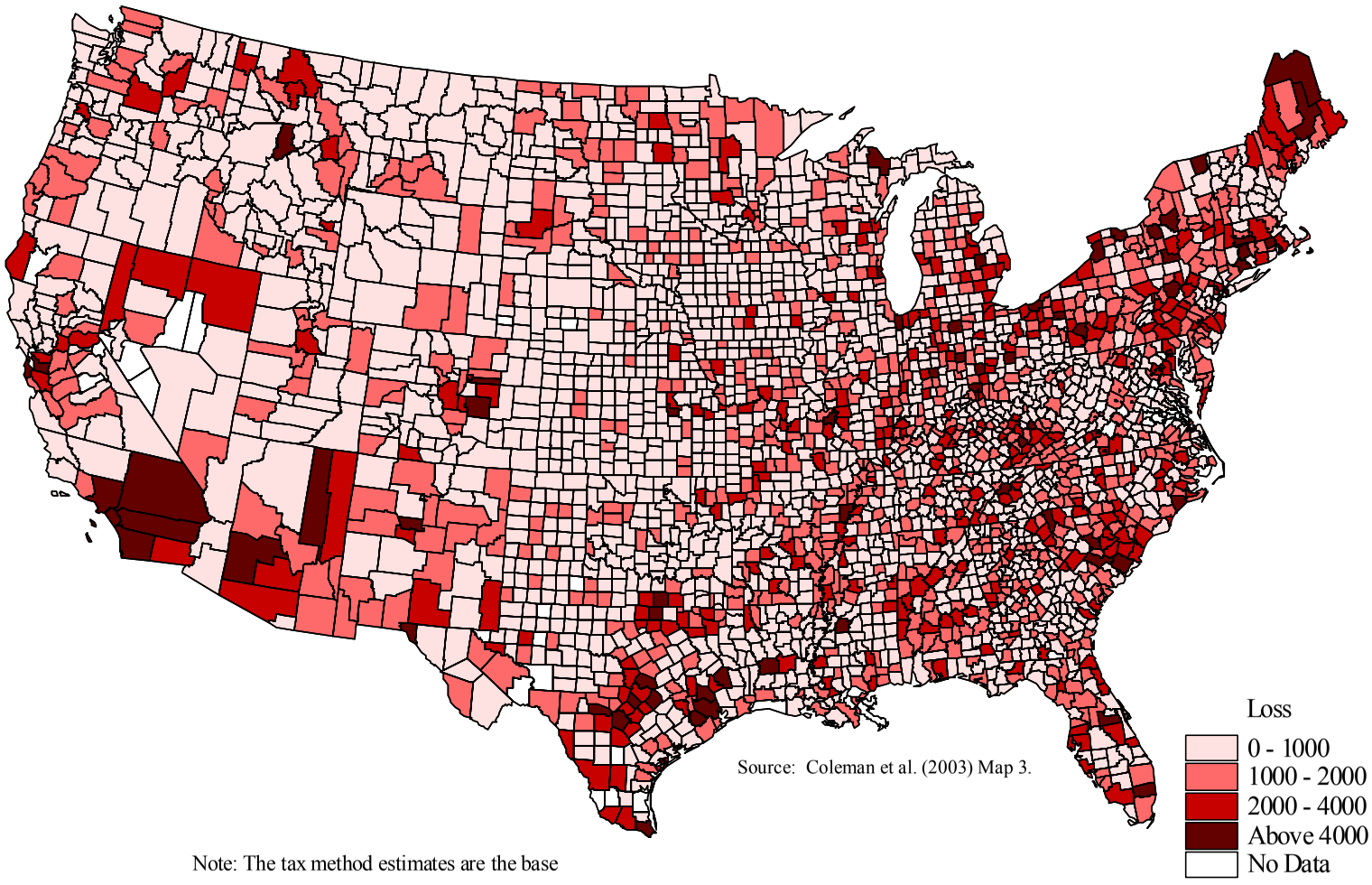
Lindley, D.V. (1953). "Statistical Inference," *Journal of the Royal Statistical Society*, Series B **15**, 30-76.

Rousseeuw, Peter J., Ruts, Ida and Tukey, John W. (1999). "The Bagplot: A Bivariate Boxplot," *The American Statistician* **53**, 382-387.

Rudas, T., Clogg, C. C. and Lindsay, B. G. (1994). "A New Index of Fit Based on Mixture Methods for the Analysis of Contingency Tables," *Journal of the Royal Statistical Society*, Series B **56**, 623-639.

Tukey, John W. (1977). Exploratory Data Analysis, Addison-Wesley, Reading, Massachussetts.

Map 1
Loss Function Values Comparing Two Sets of County Population Estimates



Source: Coleman et al. (2003) Map 3.

Note: The tax method estimates are the base

MARS: AN ALTERNATIVE TO NEURAL NET

Dan Steinberg, N. Scott Cardel, Mikhail Golovnya, Salford Systems

1. Introduction

The recent decade has been characterized by rapid developments in data mining, including extensive growth of ``black box" techniques aimed at improving the predictive accuracy of models, *neural nets* being the most vivid examples. A neural net uses a complicated yet flexible internal structure to make predictions for a response variable. To properly set up and run a neural net, a number of issues need to be resolved: predictor normalization, missing value imputation, large cardinality categorical variable expansion, etc. The model building process itself takes a significant amount of time and is often critical to the successful use of neural nets, as are optimization algorithm selection, initial guess choice, and the presence of irrelevant predictors. The end result is normally hard to understand and may be virtually impossible to interpret. In view of this, the search for alternative predictive modeling techniques devoid of the above-mentioned limitations is crucial.

As far back as the early eighties, soon after publishing the monograph *Classification and Regression Trees* with Breiman, Olshen, and Stone ([1]), Jerome Friedman began to improve some major deficiencies of building trees in a regression context. His work culminated in a comprehensive paper ([2]) that introduced a new technique known as MARS (Multivariate Adaptive Regression Splines). The initial response to this technique was both excited and skeptical. On the one hand, MARS presented a unique and flexible solution to complex regression problems, including automatic discovery of bias-removing transformations, variable selection, missing value support, region-specific interactions, etc. On the other hand, the fact that running MARS on a decent dataset would essentially mean running multiple linear regressions hundreds of thousands of times in a row could easily make it computationally not feasible. That major problem, as well as lack of a solid user-friendly implementation, doomed MARS from being widely used for nearly another decade.

Recently, however, since Salford Systems released a commercial implementation of MARS built around the original MARS code, it has become possible for a number of users to use MARS on a daily basis. The availability of a GUI interface, along with powerful graphical displays of the results, has made it possible to run MARS on a number of predictive modeling problems within seconds. In addition, minimal data preparation is required. In this paper, we introduce key MARS ideas and, using a simple well-known Boston Housing dataset, we illustrate the basics of reading MARS classic and GUI output.

2. The Modeler's Problem

A typical regression problem could be formulated as making the best possible prediction of some continuous outcome variable y based on an observed set of predictors X and some underlying loss function. In real life, the relationship between y and X is never deterministic. In other words, it is possible to have different observed values of y given the same observed state X. The underlying loss function is introduced to resolve this ambiguity; the best predicted value of y is then defined as the one that minimizes the expected value of the loss function. In particular, this means that the most complete solution to this problem would first involve determining the joint probability distribution of y and X, followed by determining the conditional distribution of y given X, and finally solving the optimization problem that minimizes the conditional expectation of loss given the observed X.

In reality, the complete solution is hardly ever possible to find in closed form. To cope with the problem, a number of simplifying assumptions are usually introduced. One may show that under the least squares loss the best prediction of y is simply the conditional expectation of y given X . The problem can thus be reformulated as

$y = f(X) + \text{noise}$.

Here $f(X)$ represents the conditional mean of y and the noise component is a random variable having zero conditional mean. The problem thus becomes to find the unknown function $f(X)$.

Unfortunately, the above setting is still too general to be used in practice. Two main issues need to be resolved:

- Which predictors constitute vector X? This is essentially the problem of selecting important variables and filtering out irrelevant (with respect to the target) predictors.
- What is the exact form of the functional relationship f?

Note that the second issue raises many difficulties in practice and requires special techniques or simplifying assumptions to be handled properly.

3. Sample Dataset

To illustrate the main regression concepts, we use a classical dataset widely known as the Boston Housing data, available from the *UCI Machine Learning Repository* (originally taken from the StatLib library, which is maintained at Carnegie Mellon University). The dataset first appeared in [5].

The goal of the study was to determine the relationship between quality of life variables and property values based on 506 census tracts in the Boston area in 1970. The target variable in this case was the median value (\texttt{MV}) of owner-occupied homes in each census tract. The quality of life variables included the following items:

NOX	concentration of nitrogen oxides (pphm)
AGE	percent built before 1940
DIS	weighted distance to centers of employment
RM	average number of rooms per house
LSTAT	percent neighborhood 'lower SES'
RAD	accessibility to radial highways
ZN	percent land zoned for lots
CHAS	borders Charles River (0/1)
INDUS	percent non-retail business
TAX	tax rate
PT	pupil-teacher ratio

Figure 1 shows the scatter matrix for a set of selected predictors and the target variable.

One may easily conclude that the underlying relationships among the collected variables are highly non-linear and complex. Thus, standard normality assumptions assumed by many standard techniques may not hold; indeed, they are violated based on any traditional normality test.

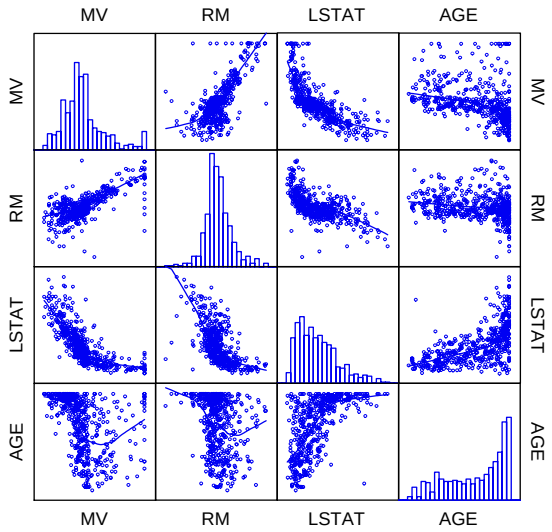


Figure 1. Scatter Matrix for Selected Variables

4. Global Versus Local Modeling

4.1. Global Parametric Modeling

We first turn our attention to classical regression techniques, multiple linear regression being the best-known.

In this setting, one usually assumes that the underlying function $f(X)$ is defined up to a certain fixed number of unknown parameters. It is then possible to estimate the unknown parameters using traditional powerful statistical techniques such as maximum likelihood or least squares estimation.

Multiple linear regression assumes that $f(X)$ is decomposed into an additive sum of known functions with unknown coefficients. For the underlying estimation theory to work, it is also important that some distribution assumptions also hold (constant variance, no autocorrelation, independent predictors, etc.). Those used most widely are standard normality assumptions, but solutions for more "exotic"' distributions also exist. (For example, assuming that the target variable has a Bernoulli distribution with parameter p depending on X, by using a logistic curve one immediately arrives at the theory of classical logistic regression, another common method in the category of global parametric methods).

Consider, for example, a simple linear regression solution to the relationship between MV and LSTAT, as shown in Figure 2.

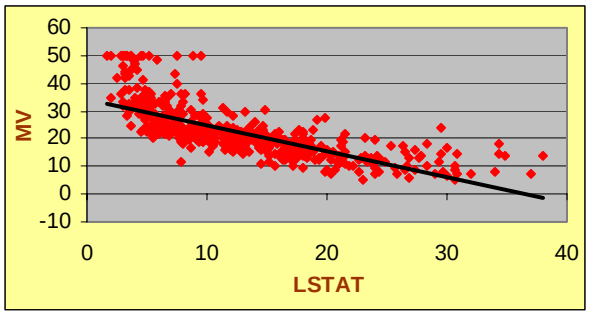


Figure 2. Regressing MV on LSTAT

Because simple linear regression is limited to straight line solutions only, the final prediction is biased: one systematically underestimates MV for low values of LSTAT and high values of LSTAT, and overestimates MV for mid-range values of LSTAT.

Also note that prediction results will, generally speaking, change everywhere, including at low values of LSTAT, if one refits the regression model for a new dataset where y is different for high values of LSTAT. This "globality"' feature of the solution explains why classical techniques are also called *global techniques*.

The main advantages of global techniques are obvious for small datasets, where the scarcity of observations justifies using every available data point to compute parameter estimates.

Note that to change the nature of the relationship presented by the solution, to make it inverse instead of linear, for example, one would have to redefine the regression equation and rebuild the model. This assumes that one can somehow determine (using visual aids, diagnostics, etc.) that $f(X)$ needs to be corrected. While this is feasible in low dimensions, it becomes extremely burdensome and unlikely in higher dimensions, not to say in real world data mining problems.

- To summarize, most classical methods have the following general characteristics:
- Rapid computation but limited flexibility
- Accurate only if the specified model is a reasonable approximation to the true function
- Can work extremely well with small data sets %(only need two points to define a line!)
- All data points influence virtually all aspects of the model

As the dataset size increases, the problem of selecting the proper form of $f(X)$ becomes more and more important because in reality most natural phenomena do not agree with simple theoretical assumptions.

4. 2. Nonparametric Local Modeling

Modern tools tend to learn about the form of $f(X)$ directly, using the data rather than relying on predefined assumptions. The idea is that once enough observed data points are available, one could somehow partition data into smaller regions and then fit separate models for each region meeting some natural boundary conditions along the way. The fact that only data points located in the immediate proximity of the current X are used in making the prediction for y automatically reduces the bias in such models. The flexibility of locally-defined regions also makes the entire process flexible, able to adjust itself based on the available data. This is what is meant by saying that *"an algorithm learns about f(X) from the data."*

Consider, for example, a simple median smooth (a well-known low-dimensional non-parametric localized method) that uses the 10\% data window shown in Figure 3 to solve the same prediction problem as above.

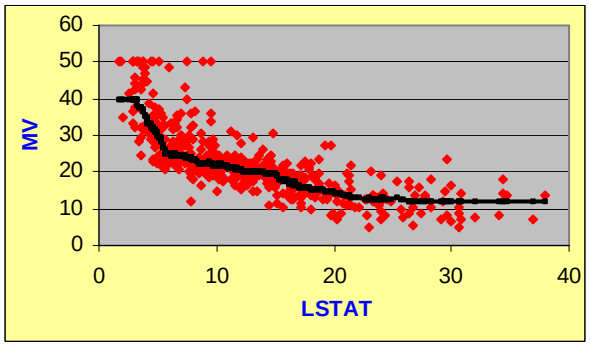


Figure 3. 10% Median Smooth

It is apparent that the solution "adjusted" itself towards capturing the correct unbiased structure of the relationship. By the nature of the median smooth, which takes 10% of the data points around any given value of X and predicts the median of y based solely on this subset of the data, the predictions on the left-hand side are no longer influenced by the observed values on the right-hand side. This property constitutes the local structure of the solution. It is also non-parametric, as no parameters are estimated. Rather the entire prediction is based directly on the observed data values.

This framework gives rise to several new.

First, one has to somehow define local regions. In low dimensions, this is relatively straightforward. For example, one can divide each variable into a few ranges. In high dimensions, however, one will quickly bump into the curse of dimensionality, which manifests itself in an exponential explosion of the number of possible subsets of a space. In addition, dimensionality results in rather unintuitive and unusual properties of the terms "nearness" and "neighborhood." Many

exhaustive search techniques have been developed to overcome dimensionality by dividing the space into local regions based on the observed data. The majority of tree-based non-parametric methods, including CART and MARS, use such an approach.

The second complication deals with defining the extent to which a model needs to be localized. In other words, how far local should one go to produce the best results? This is also known as the problem of overfitting: highly localized models tend to trace random noise patterns and fail to generalize well.

To illustrate this phenomenon, first consider the same median smooth using the 50% window shown in Figure 4.

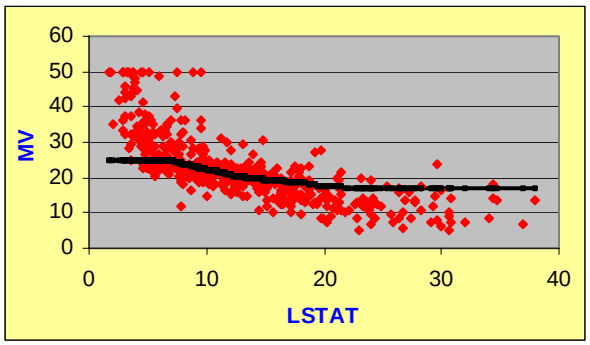


Figure 4. 50% Median Smooth

It is clear that the solution in this case is highly biased but stable (that is, has low variance).

Next, we show a 5% median smooth on the same data (see Figure 5). Now the solution is unbiased but highly unstable (that is, has high variance). The unbiasedness here must be understood in the sense that if one were able to repeat the estimation process a multiple number of times, the predictions on average would agree with the true underlying function *f*. (This forms the basis for a vast number of techniques generally known as *bootstrapping*.)

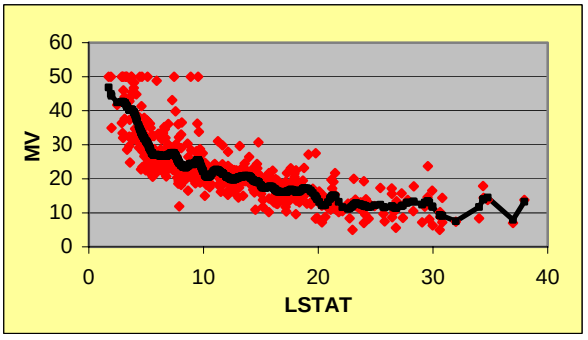


Figure 5. 5% Median Smooth

The phenomenon introduced above is generally known as the *Bias Versus Variance Tradeoff*. As locality of a model grows, the bias of predictions decreases and the variance of predictions increases.

No definitive resolution to this quandary exists, although one could consider different loss functions, which would in turn result in different model localities being optimal. For example, under least squares loss,

$$MSE = E\left(y - \hat{f}(x)\right)^2$$

$$MSE\left(\hat{f}\right) = E_x\left[Var\left(\hat{f}(x)\right) + \left\{f(x) - E\left(\hat{f}(x)\right)\right\}^2\right]$$

$$Variance \quad + \quad Bias^2$$

one would focus on minimizing *the mean squared error* (*MSE*), defined as

There are other types of measures as well. In general, the notion of the optimal model is contingent on the type of measure chosen. The exact theoretical solution in many cases is cumbersome and hard to obtain or use.

In practice, the most useful way of defining an optimal model is usually the following:

- Randomly partition data into three pieces – *Learn*, *Test*, and *Validate*
- Develop a sequence of models with varying localities based on the Learn data
- Check the performance of each model on the Test data, using whatever performance measure is required (MSE, MAE, Top Decile Lift, etc.). The optimal model is by definition the model that performs best on the Test set
- Confirm the performance of the chosen model on an independent Validation set.

Note that often one can skip the validation step because the Test set is not used until very late in the game, resulting in the selection bias introduced being small.

5. MARS Algorithm

5.1. Approximating Functions With Splines

We now turn our attention toward the MARS algorithm. In the previous section, we showed that a multiple linear regression solution is stable and fast to compute , while at the same time being quite inflexible and thus difficult to use for real life modeling tasks. The main question then is whether multiple linear regression can be modified so as to make it more flexible while preserving its robustness.

For example, in our Boston problem, if we could somehow determine that the range of LSTAT could be partitioned into four distinct local regions, then we would be able to fit separate regression lines within

each region, requiring only that they must merge at the end points (boundary conditions). This could result to the solution shown in Figure 6.

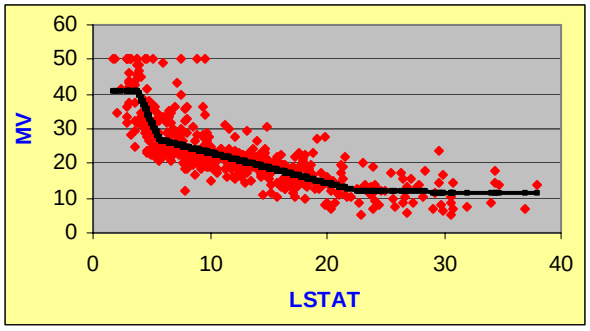


Figure 6. Piece-Wise Linear Regression

Note the advantages of such a solution. It is flexible and yet preserves the "smoothness" of a simple linear regression. The big question that remains is how to define the regions above.

We call the point where two regions join a *knot*. If one needed to introduce just two regions (a single knot), a reasonable approach would be to try all available data points as possible knot locations, fit separate regression lines complying with the boundary condition and then pick the knot location that results in the greatest decrease in the sum of squared errors. Thus, the problem of finding the best single knot location requires approximately N separate regression iterations. Unfortunately, one cannot generalize this process by considering simultaneous placement of K knots at the same time because this would require approximately N^K regression iterations, which becomes unwieldy even when K=2.

Instead, we will add knots sequentially, one by one.

Consider, for example, how the proposed method would handle the flat top function shown in Figure 7.

This functional shape is well known for its resistance to default global parametric methods. For example, running a simple linear regression on such data would produce no result because there is no linear trend. On the other hand, the sequential knot placement procedure will generate the sequence of events shown in Figure 8.

The first knot is placed somewhere in the middle of the flat part, a location that minimizes the resulting sum of squared errors. The resulting angular shape, albeit primitive and simple, already reflects the general tendency of the relationship to increase initially, followed by a reversal.

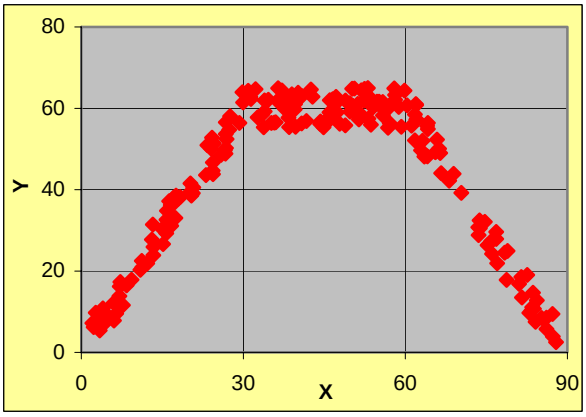


Figure 7. Flat Top Function

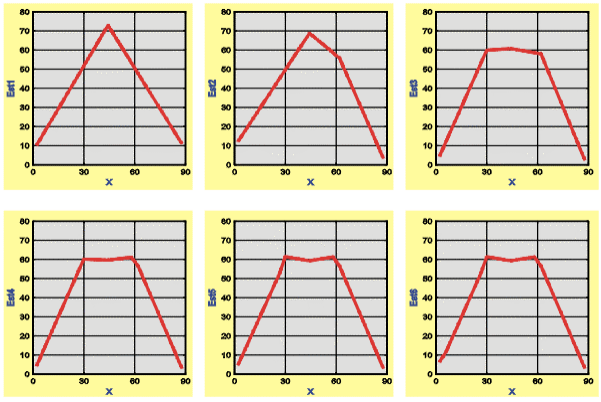


Figure 8. Sequential Knot Placement

Adding the second and third knots essentially recovers the true shape of the function. The process need not stop here, however. More knots can be added to further refine the relationship, as shown in Figure 8.

One natural consequence seen in this experiment is that three knots (three steps) are required to completely uncover the true shape whereas the simultaneous knot placement process would require only two knots. On the other hand, our process required computing only N*3 regressions, whereas the simultaneous knot placement would require N^2 regressions, a substantially larger number of calculations. The price we pay is the presence of the initial knot in the middle, which turns out to be redundant. Note that having this knot is unavoidable as it is the first one introduced.

One could now refine this process by introducing a follow-up knot deletion step. Once a certain number of knots have been introduced, we start removing knots sequentially (one at a time), such that the resulting sum of squared errors increases the least, essentially eliminating the redundant knots that are no longer needed. For the flat top function example, the backwards elimination process would first remove all

nonessential knots while keeping the two real ones until the end of the process.

5.2. Basis Functions

The geometric construction above basically illustrates how MARS works. However, a formal definition of the MARS algorithm requires introducing analytical machinery called *basis functions*.

Consider the following class of functions, or basis transformations of variable X:

Direct transformation – Max(0, X-c)

Mirror transformation – Max(0, c-X)

Note that these "hockey-stick" transformations are piece-wise linear with c being the knot location that defines where the slope changes from 0 to 1. These transformations are also continuous because of the shift in the amount of c that occurs at the knot location. The direct transformation effectively eliminates the variance of X to the left from the knot by truncating values to zero, whereas the mirror transformation eliminates the variance to the right from the knot. This means that if one could use these transformed Xs as terms in a multiple linear regression, the estimates of the coefficient would be based only on smaller sub-regions of data. Hence, such a technique would fit the above definition of the localized techniques.

Figure 9 demonstrates fitting a standard regression using a basis transformation of variable INDUS with c=4.

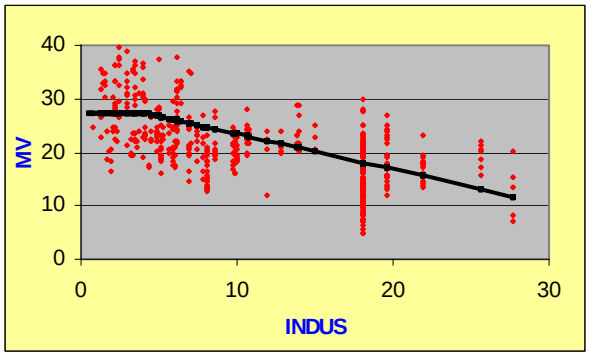


Figure 9. Fit Using One Basis Function

The solid black line is obtained by plotting the following regression solution

MV = 27.395 - 0.659*(INDUS -4)_+.

Here the slope and intercept were obtained by a running simple multiple linear regression using the transformed variable INDUS.

If we now add another transformation and run a multiple linear regression using the two basis functions based on INDUS, with knots set at 4 and 8, the result is the model shown in Figure 10.

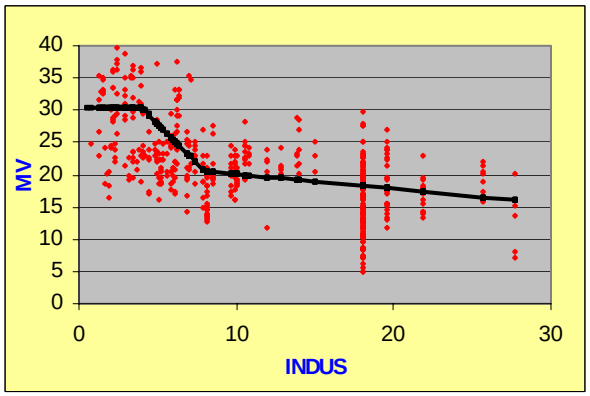


Figure 10. Fit Using Two Basis Functions

The corresponding equation is

$$MV = 30.290 - 2.439*(INDUS - 4)_+ + 2.215*(INDUS-8)_+$$

Note that the resulting solution automatically meets the continuity boundary conditions. This is no surprise. Because the individual basis functions were constructed to be continuous, the resulting linear combination will always be continuous. The boundary conditions are thus met automatically! Another consequence of such special construction is that the interpretation of the regression coefficients is no longer trivial. The –2.439 coefficient in front of the "leftmost" basis function reflects the slope of the line, whereas the 2.215 coefficient of the second basis function must be interpreted as an adjustment to the previous slope, thus yielding the –0.224 overall slope of the last segment of the line.

Having only direct basis functions is obviously not enough to restore the flat top function introduced earlier. Indeed, as one may see from the previous plots, no matter what the coefficients in the resulting equation are, the leftmost line segment is bound to be horizontal (having zero slope). The mirror basis function is needed to eliminate this limitation.

The following solution is constructed using two previous direct basis functions and one mirror basis function with c=4:

$$MV= 29.433 + 0.925*(4 - INDUS)_+ -2.180*(INDUS-4)_+ +1.939*(INDUS-8)_+$$

Notice from Figure 11 that the left-most line segment has a negative slope of –0.925 even though the corresponding mirror basis function has a positive coefficient. The mirror function itself is negatively related to the predictor on which it is based.

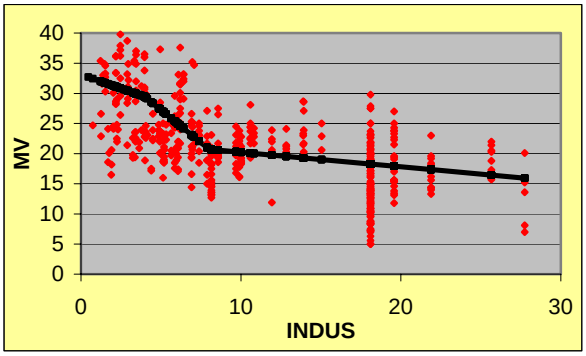


Figure 11. Fit Using Three Basis Functions

5.3. MARS Algorithm

We are now in a position to introduce the final MARS algorithm:

- Start with a simple intercept-only model.
- Search for the variable-knot combination that improves the model the most.
  - Technically this means trying all possible pairs of basis functions one at a time.
  - Improvement is measured in terms of MSE.
  - Adding a basis function pair will never increase MSE.
- Repeat the process of adding basis functions until the largest model is built (a user-controlled parameter).
- Proceed with backwards elimination.
  - Remove basis functions one at a time such that at each step MSE increases the least.
  - At the end of this step one will have a sequence of models of varying sizes (varying locality).
- Choose the optimal model by using MARS' built-in measure of *generalized cross validation* (*GCV*) or by conducting an external evaluation based on a Test sample.

Note that MARS can handle categorical variables by trying dummy indicator variables (categorical analogue of basis functions) based on all possible combinations of levels.

MARS also naturally handles interactions by allowing products of basis functions. This natural extension has the striking and powerful feature of detecting region-specific interactions (because individual basis functions are defined by only a subset of the data). These products are far more useful and flexible than traditional global interactions usually considered in multiple linear regression, and also are obtained automatically!

MARS has a built-in ability to handle missing values by introducing special types of basis functions that are missing value indicators. The resulting MARS model "switches" among different basis functions depending on whether the underlying predictors are available or not on a case-by-case basis.

6. Running MARS on the Boston Housing Data

We begin our modeling session by opening the dataset (Figure 12).

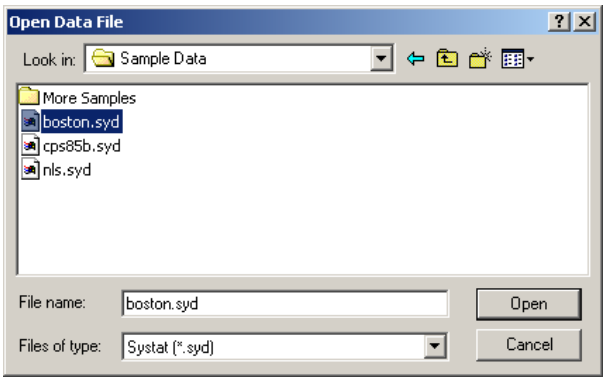


Figure 12. Opening Up Data

We then use the resulting *Model Setup* window to set up our run by selecting the target variable, potential predictors, categorical variables, etc. (Figure 13)

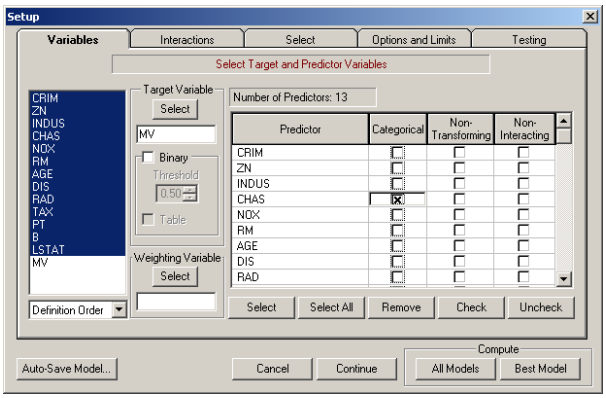


Figure 13. Setting Up Model

A special *Options and Limits* tab allows a user to specify important MARS controls such as the maximum number of basis functions, interactions level, processing speed, etc. (Figure 14)

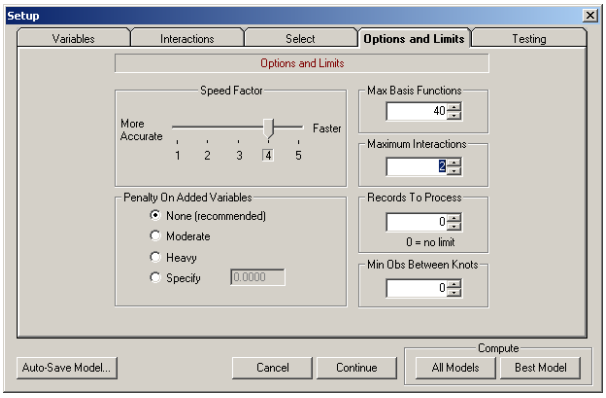


Figure 14. Setting Up MARS Parameters

After clicking the **[All Models]** button, MARS displays a run progress window followed by results windows.

The first window, known as the selector window (Figure 15), displays the complete list of models developed by MARS and allows the user to select a model of any size for a closer look.

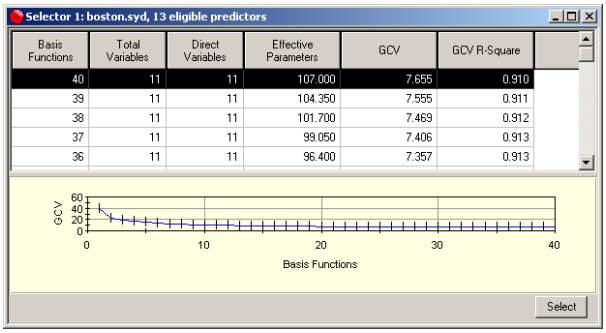


Figure 15. Selector Window

The second window (Figure 16) shows details of the optimal model (based on the GCV measure).

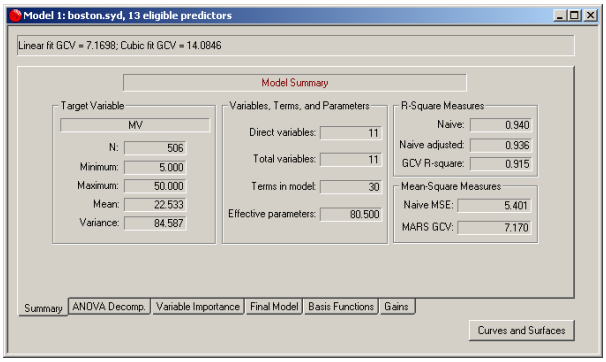


Figure 16. Model Results Window

According to MARS, the optimal model has 30 basis functions based on 11 predictors and results in an adjusted R-squared of about 93%.

The model itself can be viewed under the *Basis Functions* tab (Figure 17).

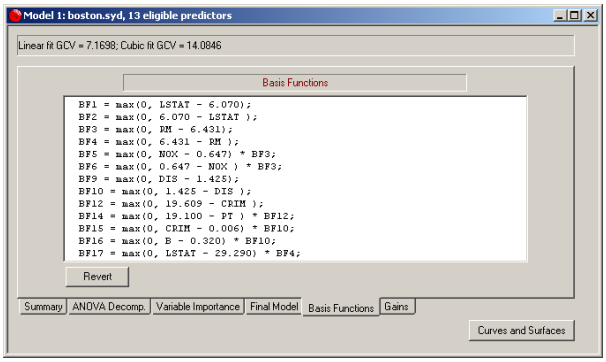


Figure 17. Basis Functions

Here is the same model obtained by cutting and pasting from the above tab:

```
BF1 = max(0, LSTAT - 6.070);
BF2 = max(0, 6.070 - LSTAT );
BF3 = max(0, RM - 6.431);
BF4 = max(0, 6.431 - RM );
BF5 = max(0, NOX - 0.647) * BF3;
BF6 = max(0, 0.647 - NOX ) * BF3;
BF9 = max(0, DIS - 1.425);
BF10 = max(0, 1.425 - DIS );
BF12 = max(0, 19.609 - CRIM );
BF14 = max(0, 19.100 - PT ) * BF12;
BF15 = max(0, CRIM - 0.006) * BF10;
BF16 = max(0, B - 0.320) * BF10;
BF17 = max(0, LSTAT - 29.290) * BF4;
BF19 = max(0, NOX - 0.693) * BF1;
BF20 = max(0, 0.693 - NOX ) * BF1;
BF21 = max(0, AGE - 18.800) * BF3;
BF22 = max(0, 18.800 - AGE ) * BF3;
BF23 = max(0, NOX - 0.770) * BF4;
BF24 = max(0, 0.770 - NOX ) * BF4;
BF25 = max(0, TAX - 233.000);
BF26 = max(0, 233.000 - TAX );
BF27 = max(0, RAD - 3.000);
BF28 = max(0, 3.000 - RAD );
BF29 = max(0, AGE - 98.800);
BF30 = max(0, 98.800 - AGE );
BF31 = max(0, B - 232.600) * BF30;
BF32 = max(0, 232.600 - B ) * BF30;
BF33 = max(0, PT - 18.600) * BF26;
BF35 = max(0, INDUS - 10.810) * BF2;
BF38 = max(0, 5.631 - RM ) * BF29;
BF39 = max(0, NOX - 0.385) * BF10;
BF40 = max(0, B - 0.320) * BF2;

Y = 17.648 - 0.638 * BF1 + 22.851 * BF2 + 17.494 * BF3
          + 1.725 * BF4 - 181.448 * BF5 - 16.229 * BF6
          - 0.693 * BF9 + 655.693 * BF10 + 0.265 * BF12
          + 0.030 * BF14 - 7.229 * BF15 - 1.003 * BF16
          + 0.476 * BF17 + 2.146 * BF19 + 2.662 * BF20
          - 0.097 * BF21 - 0.430 * BF22 - 55.515 * BF23
          - 18.060 * BF24 - 0.006 * BF25 + 0.173 * BF26
          + 0.200 * BF27 - 0.968 * BF28 + .335598E-03 * BF31
          - .732026E-03 * BF32 - 0.224 * BF33 + 0.281 * BF35
          + 3.804 * BF38 - 530.688 * BF39 - 0.056 * BF40;
```

Note that the model representation is completely compatible with SAS and can be easily converted into any other language.

Even though the model looks horrendously complicated, it reveals a very simple structure. First, it

defines the hockey stick transformations of the original variables. One could treat the resulting basis functions simply as new variables that are introduced into our dataset. In addition, note how interactions are handled by defining a product of two basis functions (see, for example, BF14 and BF15).

The final prediction equation is simply a multiple linear regression around the transformed variables. Because it is essentially a linear combination of individual terms, it can be decomposed easily into individual parts based on one or two original predictors. These in turn could be plotted to reflect individual contributions. The plots are available by pressing the **[Curves and Surfaces]** button.

Figure 18 shows joint contribution of RM and NOX towards MV.

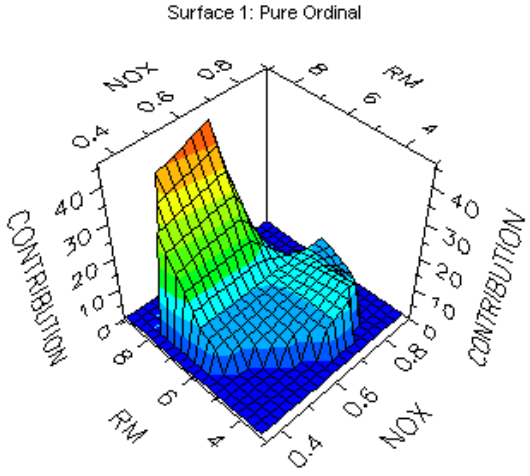


Figure 18. 3D Contribution Plot

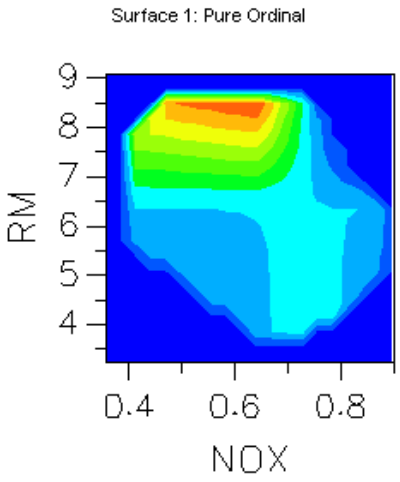


Figure 19. Heat Contribution Map

This plot can be rotated to get a better view or can be presented as a heat map (Figure 19).

One can immediately conclude that a substantial jump in house values occurs when the average number of rooms exceeds 6 or 7 (a rather unexpected result, as one might expect a gradual increase in house values as the number of rooms increases). It is also clear from this plot that higher pollution levels only exist in areas with somewhat smaller houses. The negative contribution of the pollution levels once they exceed 0.7\% is also apparent from this plot.

Similarly interesting conclusions are obtained from analyzing the interaction between the distance to downtown and the crime rate (Figure 20).

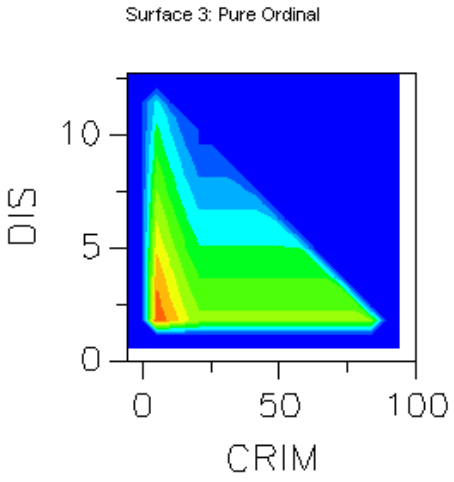


Figure 20. Heat Contribution Map

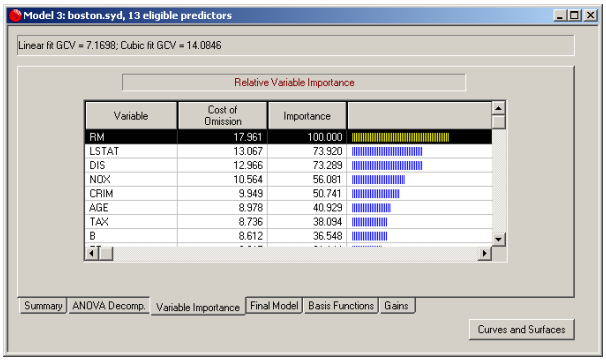


Figure 21. Variable Importance

Clearly, the highest crime rates are associated with certain neighborhoods within the downtown area. As one moves further away, crime rates tend to decrease. Not surprisingly, downtown areas with low crime rates combine with a dramatic spike in home values, an effect that eventually disappears as one moves away into the suburbs.

MARS also offers *Variable Importance* (Figure 21) another useful way to look at our model. The importance of each predictor is computed based on how much the resulting MSE increases once this predictor is removed from the model (also called *the cost of omission*). Naturally, the most important predictor results in the greatest increase; the remaining predictors are scaled appropriately on a 100\% scale. According to our model, house sizes, poverty rates, and distance to downtown are the most important factors associated with varying house values, followed by pollution rate, crime rate, etc.

7. Conclusion

MARS is a powerful modern technique that allows us to find a non-parametric non-linear multiple linear regression solution quickly, efficiently and automatically. The final model can be viewed as a piece-wise linear approximation to the underlying non-linear function. The approximation itself is based on combining first order splines and their products, also known as basis functions. The end result contains not only a final model that can be used to make future predictions effectively, but also a collection of useful displays, including variable importance scores and contribution plots. While running a multiple linear regression on the Boston Housing data results in an R-squared of about 73% and limited insights into the nature of the relationships, running MARS on the same data quickly produces a far richer model with interesting localized insights and superior accuracy, with an R-squared of about 93%.

Due to various internal computational tricks, MARS effectively runs hundreds of thousands of multiple linear regression tries in a reasonable amount of time. (For example, the total running time of MARS on the Boston Housing data was about 12 seconds, while a single multiple linear regression round took about 4 seconds to complete.)

Given the natural ease of use, powerful presentation of the results, ability to interpret the final model, and fast running times, MARS provides a powerful alternative to other modern non-parametric non-linear techniques such as Neural Nets.

References

[1] Leo Breiman, Jerome~H. Friedman, Richard~A. Olshen, Charles~J. Stone, *Classification and Regression Trees*, Wadsworth \& Brooks, Pacific Grove, California (1984).

[2] Jerome~H. Friedman, *Multivariate adaptive regression splines (with discussion)*, Annals of Statistics, 19 (1991), 1–141.

[3] Jerome~H. Friedman and B.~W. Silverman (1989), *Flexible parsimonious smoothing and additive modeling (with discussion)*. TECHNOMETRICS, 31 (1989), 3–39.

[4] R. D. De Veaux D. C. Psichogiosand L. H. Ungar *A Comparison of Two Nonparametric Estimation Schemes: Mars and Neutral Networks*, Computers Chemical Engineering, Vol.17 (1993), No.8.

[5] D. Harrisonand D. Rubinfeld, *Hedonic Housing Prices \& Demand For Clean Air*. Journal of Environmental Economics and Management, v5 (1978), 81–102.

FORECASTING WAITING-TIME FOR HEALTH SERVICES: CAPTURING THE NONLINEAR DYNAMICS IMPLIED BY A CONSTRAINED DECISION MAKER

Trond Jorgensen, Ph.D., Trond.Jorgensen@Altarum.org

Health Solutions Division, Altarum Institute

Introduction

When attempting to forecast a time-series of operational performance metrics, how can we take into account the existence of an intelligent agent who is attempting to control the process? In this paper, we study this problem in the context of forecasting waiting-time for healthcare services.

Throughout this paper, when we are referring to an "intelligent agent" or "rational decision maker", we are referring to a decision maker who makes optimal decisions given the information that is available. Mathematical models that prescribe the optimal decision for many management problems, such as optimally allocating resources under uncertainty, are available in the field of Operations Research.

Whereas Operations Research models are founded within a philosophy of *rational* modeling (i.e. based on logical principles or axioms), time-series analysis is carried out in an *empirical* tradition where the whiteness and magnitude of residuals determine whether a model is classified as acceptable, or not.

In this paper, we suggest that when attempting to forecast a time-series where a rational decision maker attempts to control the observed process, this prior knowledge can be used to select a more appropriate mathematical structure based on the theory of Operations Research. The end result is a class of forecasting models that can be justified both from an empirical perspective and a rational perspective.

Existing literature

In the field of Operations Research, models are often formulated as mathematical optimization problems [1]. Note that such decision models are usually intended as prescriptive tools, in contrast to descriptive and predictive tools. In general, decision models are intended to support a decision maker faced with a complex problem or a problem that requires careful analysis before a decision is made, for example, when the stakes are high [2].

The rationale for optimization models can sometimes be linked directly to a set of axioms for rational decision makers. From these axioms, it is possible to prove the existence of a utility function that measures the decision maker's subjective preferences and attitude towards taking risk. Based on this well-founded theory, it can be derived that it is optimal for a rational decision maker to maximize expected utility (as defined by the utility function) [3].

Alternatively, Operations Research models that are not based directly on utility theory are often based on established economic principles (such as profit maximization) or based on stated management objectives. For example, in a project management setting, the project manager may choose to minimize project duration by optimally allocating resources [4].

By definition, forecasting models are not prescriptive and, therefore, are usually considered as a completely separate area of interest than prescriptive modeling in Operations Research. However, in this paper we will combine the two areas.

In traditional time-series analysis, linear state-space models are often applied, including the well-known special cases such as ARMA and ARIMA models [5].

Recall that a common discrete-time formulation of a linear state-space model is:

$$x(k+1) = Ax(k) + \varepsilon(k)$$

$$y(k) = Hx(k) + z(k)$$

where $x(k)$ is the state-vector at time k and $y(k)$ is the observation vector at time k. The matrices A and H are known in advance (and may be generalized to be time-varying), while the vectors $\varepsilon(k)$ and $z(k)$ are white noise processes. From a mathematical point of view, the state-space representation above is a set of (stochastic) difference equations. Formulating a similar continuous-time version leads to a set of (stochastic) differential equations.

In the field of nonlinear forecasting, more general nonlinear models are also suggested, such as neural networks. In particular, by applying new insight in the mathematical field of nonlinear dynamics (often referred to as *Chaos theory*) new nonlinear forecasting models have been suggested [6], [7]. From a strictly empirical perspective, any mathematical model that passes the well-defined residual analysis, model selection theory and cross-validation techniques (see e.g. [8]) is deemed acceptable.

The area of study sometimes referred to as *Control theory* is related to both predictive time-series analysis as well as prescriptive Operations Research. In particular, traditional control theory demonstrates how any process that can be described accurately by a linear state-space model can be controlled if an unconstrained additive input can be applied to the system. Control theory has found many important applications in engineering, such as missile control and robotic control [9].

A central theme in traditional Control theory is the use of *linear feedback*. The theory around this concept determines the conditions that must be met for linear state-space models to be appropriate for forecasting systems involving rational decision makers.

The following theorem (on page 278 in [10]) defines and justifies the use of linear feedback. Generalizations of this result with more extensive objective functions and that add stochastic noise terms and discrete-time versions give very similar results and are omitted for clarity. E.g., see [11].

Theorem 1: Optimal regulator

Let the process under study be modeled using the following deterministic continuous-time linear state-space model:

$$\frac{dx}{dt} = Ax + Bu$$

where $x = x(t)$ is the state-vector describing the system over time t and $u = u(t)$ is the input vector that the decision maker can control over time t and where A and B are constant matrices. The decision maker wants to regulate the system such that $x(t)$ remains as close as possible to the origin. This is modeled by minimizing the following quadratic cost function:

$$C = \int_0^\infty (x^T Q x + u^T R u) dt$$

where Q and R are known matrices. Given this system model, the optimal decision is to choose:

$$u = Kx$$

where the matrix $K = -R^{-1}B^T P$ is a constant matrix and where P is the unique matrix that satisfies the so-called algebraic Riccatti equation:

$$PBR^{-1}B^T P - A^T P - PA - Q = 0$$

Note that, in Theorem 1, the optimal decision u is based on the current state x. This is referred to as *feedback control*. Since the optimal control, in this case, is a linear function of the state vector ($u = Kx$), it is referred to as *linear feedback*.

As a matter of terminology, we refer to the system as *closed-loop* when the decision maker applies feedback control. Based on the model in Theorem 1, it is important to realize that, in this case, the closed-loop system remains linear:

$$\frac{dx}{dt} = Ax + Bu = Ax + BKx = (A + BK)x$$

Under the assumptions of Theorem 1, this linear closed-loop model gives a rational justification for the use of linear state-space models for modeling closed-loop systems as advocated in traditional time-series analysis textbooks.

However, the main point that we make in this paper is to point out that, while applicable in many situations (as demonstrated by the literature), one of the important (but implicit) assumptions in Theorem 1 is often unrealistic in closed-loop situations related to managerial processes: The

literature related to the study of managerial decision problems (i.e. Management Science or Operations Research) shows that such problems can be recognized by their many complex and interrelated constraints. Thus, for the class of processes that are controlled by the management in an organization, the implicit assumption in Theorem 1 that the input u can be chosen without any constraints is unrealistic.

Furthermore, it is well known within the field of Control Theory that it is significantly harder to find analytical solutions when adding constraints. Only general analytical principles are available (such as Pontryagin's maximum principle [10]). Also, since the applications of Control theory traditionally have been mostly within the field of engineering, the need for quick controller implementations in hardware have supported research and use of analytical solutions (i.e. closed-form solutions) rather than numerical techniques (i.e. open-form solutions) that tend to be much slower to calculate. In contrast, in the field of Operations Research, such severe time-constraints on computing are usually not present and the main focus has been on numerical techniques (e.g. see [12] and [13]).

Based on these observations, in the next section, we give a simple example where we demonstrate the nonlinearity that arises when a decision maker tries to keep waiting-time under control. In later sections, we also show how this class of nonlinear models can be used for forecasting.

It should be noted that our focus is on micro-level problems with only a single decision maker affecting the system under study. In contrast, an extensive overview of existing macro economic models (i.e. that seek to take into account the aggregate effect of a large number of decision makers) can be found in [14].

Describing a managerial process affected by an intelligent agent

Having realized from the previous section that linear state-space models (such as ARMA and ARIMA) are not appropriate for capturing the decisions made by decision makers (or agents) in managerial processes, we look to the field of Operations Research to attempt to capture the mathematical structure that is commonly used to model managerial decisions.

A reoccurring theme in the Operations Research literature is the use of convex optimization models. From a modeling perspective, a large number of managerial decision problems have been formulated accurately using convex models. Also, from a practical point of view, convex models have proved feasible and practical to solve numerically.

A simple deterministic convex model has a quadratic objective function and linear constraints and takes the form:

$$\min \frac{1}{2} x^T H x + v^T x$$

subject to:

$$Ax = b$$
$$Cx \le d$$

As a special case, note that it is easy to formulate a discrete-time optimization model with quadratic objective function, constraints and with a dynamic system model similar to the state-space model in Theorem 1. For example:

$$\min \sum_{k=0}^{N} {x_k}^T Q x_k + g^T x_k$$

subject to:

$$x_{k+1} = Ax_k + Bu_k \text{ for all } k$$
$$u_k \le Cx_k + v \text{ for all } k$$
$$u_k \ge Dx_k + w \text{ for all } k$$
$$Fx_k \ge 0 \text{ for all } k$$

Note that this is a deterministic model. Alternatively, we can use a multi-stage stochastic model.

Even though this model can be classified as a convex optimization problem and has a unique solution (if Q is positive semi definite), in the next section we demonstrate by an example that u_k can be a non-convex function in x_k (due to the constraints). Thus, the closed-loop system is also non-convex and non-linear. Also, note that it is difficult to write down an analytical expression for u_k other than simply stating that u_k is the solution of an optimization problem (as done above).

However, the overall decision problem is convex and can be readily solved numerically. Thus, from a practical point of view, the optimization model can be readily evaluated and used for forecasting purposes.

A simple waiting-time model

As a concrete example of the previous discussion, consider the waiting-time for receiving a particular health service at a particular facility and suppose a manager (i.e. the decision maker) balances waiting-time and staffing costs by hiring and firing personnel. The actions of the decision maker depend on the current state of the process (i.e. the observed waiting-time and the current staffing level), but the underlying structure of the decision is known in advance and can be described using an optimization model.

Before we can formulate the optimization model, let us introduce some notation:

w_k : The mean waiting-time in time-period k.
b_k : The number of patients waiting for a service in time-period *k* (i.e. the backlog).
d_k : The number of new patients arriving in time-period *k* (i.e. the demand).
v_k : The number of patients treated in period *k*.
s_k : The number of critical staff (Full Time Equivalents) available in time-period *k*.
u_k : Hiring decision in time-period *k* (i.e. the number of staff hired, or fired if negative).

We know for certain that the backlog is given by:

$$b_{k+1} = b_k + d_k - v_k \quad (1)$$

Similarly, the staff-level is defined by:

$$s_{k+1} = s_k + u_k \geq 0 \quad (2)$$

Suppose that the dynamics of personnel efficiency is not of interest and that we model efficiency as a constant (c_2):

$$v_k = c_2 s_k \quad (3)$$

Also, suppose there are practical limits for hiring and firing (e.g. due to budget constraints) and, more precisely, that changes in staffing are limited by certain percentages (c_3 and c_4) of the existing staffing level.

$$u_k \leq c_3 s_k \quad (4)$$
$$u_k \geq -c_4 s_k \quad (5)$$

For simplicity, the decision maker plans for a particular fixed percentage growth in demand:

$$d_{k+1} = \frac{1}{\alpha} d_k \quad (6)$$

where α is a strictly positive constant.

Note that, so far, we have not explicitly involved waiting-time w_k , only the backlog b_k . However, since we are not intending to model the short-term dynamics in the healthcare queue, we assume that the queue is approximately in steady-state such that Little's formula from queuing theory applies (see e.g. [2]):

$$w_k = \frac{b_k}{d_k} \quad (7)$$

Little's formula motivates the following objective function where the decision maker balances the goals of having the waiting-time close to a target value $\overline{w}$ while, at the same time, minimizing staffing cost:

$$\min \sum_{k=1}^{N} \left(\frac{b_k}{d_k} - \overline{w} \right)^2 + c_1 \frac{s_k}{d_k} \quad (8)$$

We can simplify the formulation of this optimization problem by introducing the following transformations and making Equation (6) redundant:

$$g_k = c_2 \frac{s_k}{d_k} \quad (9)$$
$$h_k = c_2 \frac{u_k}{d_k} \quad (10)$$
$$c = \frac{c_1}{c_2} \quad (11)$$

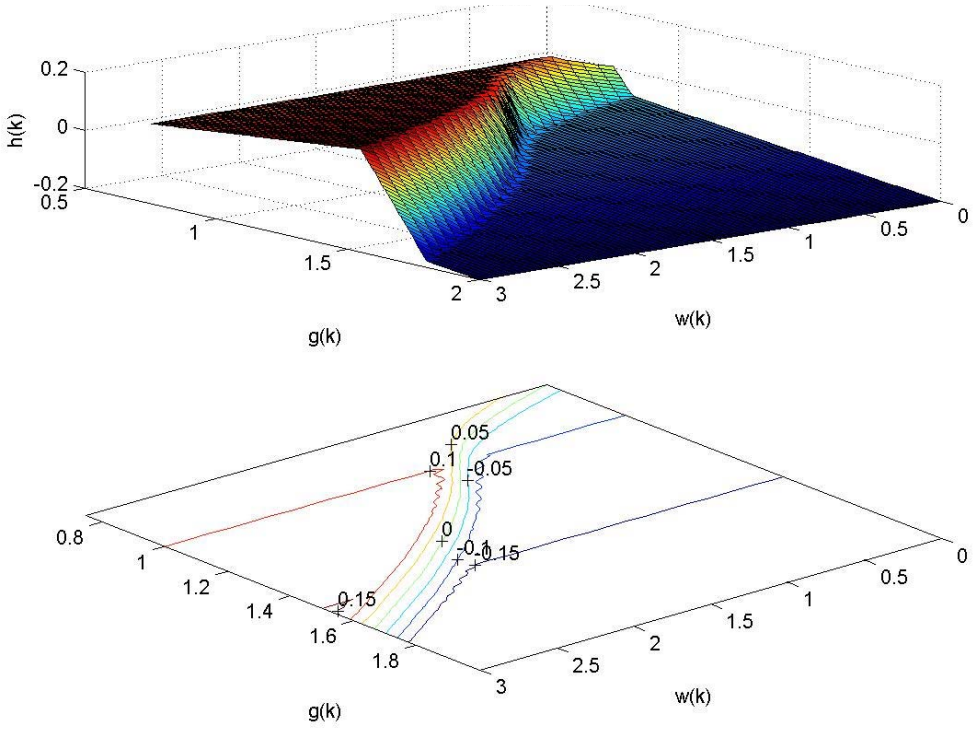
In Equations (1)-(5), we divide by d_k and apply Equations (6), (7), (9) and (10). Similarly, we reformulate the objective function (8). This leads to the following quadratic optimization problem:

$$\min \sum_{k=1}^{N} (w_k - \overline{w})^2 + c g_k \quad (12)$$

subject to:

$$w_{k+1} = \alpha(w_k + 1 - g_k) \ , \ w_0 \text{ known} \quad (13)$$
$$g_{k+1} = \alpha(g_k + h_k) \ , \ g_0 \text{ known} \quad (14)$$

Figure 1. The optimal nonlinear feedback function for optimal hiring of personnel as a function of current staffing level and waiting-time



(15) $h_k \le c_3 g_k$

(16) $h_k \ge -c_4 g_k$

(17) $g_k \ge 0$

(Note that h_{N-1} and g_N are included for notational simplicity only, but are irrelevant and could have been deleted from the model.)

Note that h_k is the decision-variable (over time) that determines hiring (u_k). In particular, note that the structure of the optimization model (12)-(17) is similar to the quadratic model example in the previous section and that it differs from the model in Theorem 1 especially due to the introduction of constraints. In contrast to Theorem 1, it is particularly important to notice that the optimal hiring decision is a nonlinear function of the current state and that this causes nonlinear closed-loop dynamics.

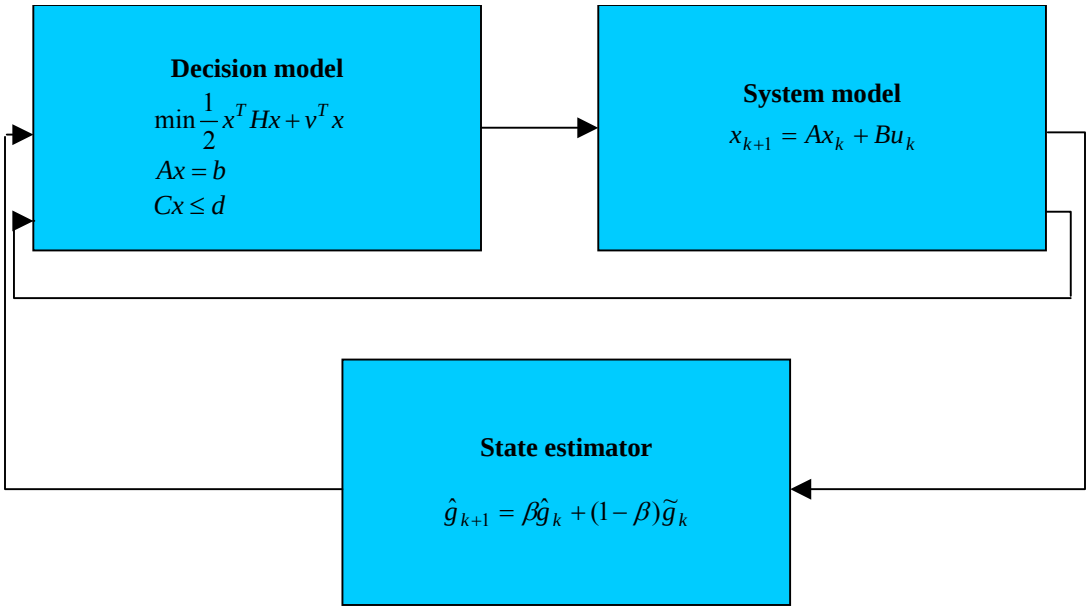
To illustrate this important point, for a particular choice of parameters $(\alpha, c, c_3, c_4, \overline{w})$, we evaluated the optimal decision $h_1 = h_1(w_0, g_0)$ for different values of w_0 and g_0 using the software package MATLAB (by The MathWorks, Inc.) and its optimization toolbox. This corresponds to the decision that would be used if the decision maker reapplies the optimization model reactively at each time-step based on the current state. This could also be classified as a sliding horizon approach since we assume a finite horizon N in the calculations. The result is shown in Figure 1. We see immediately that unless the dynamics of w_k and g_k occurs within a very small area, a linear approximation of the optimal feedback (as implied by Theorem 1) does not fit well at all.

Based on the above, rather than relying on the traditional (implicit) assumption of linear feedback when analyzing waiting-time (which leads to linear state-space forecasting models), we find it more natural to use feedback that has the shape as shown in Figure 1 (which is determined by the parameters $\alpha, c, c_3, c_4, \overline{w}$).

Estimating actual staffing level based on observations of waiting-time

In the simple waiting-time model in the previous section, the manager bases the staffing decision on two factors: current staff level and current waiting-

Figure 2. The structure of the closed-loop forecasting model



time. However, suppose we have been given only a series of past waiting-times and want to forecast future waiting-time. In other words, we do not have any observations of the staffing level. To get around this problem, we estimated the staffing state-variable g_k and used the estimated staff-level $\hat{g}_k$ in the evaluation of the nonlinear feedback function.

We applied a simple adaptive approach based on the fact that Equation (13) suggests a relationship between the transformed staffing g_k and waiting-time w_k:

$$\tilde{g}_k = \max(0, -c_5 w_k + \alpha(w_{k-1} + h_{k-1} + 1))$$
$$\hat{g}_{k+1} = \beta \hat{g}_k + (1-\beta)\tilde{g}_k$$

The maximum operator ensures that the estimate is positive. Note that two new parameters have been introduced: β is a smoothing parameter and c_5 is a parameter that captures the noise characteristics. In addition, we also need the initial parameter g_0 to be used in the beginning of the time-series. These three new parameters are estimated simultaneously with the other five unknown parameters in the decision model in the previous section ($\alpha, c, c_3, c_4, \overline{w}$). Thus, the total number of parameters to be estimated based on the series of waiting-time observations is eight.

An alternative approach to estimating the staffing would be to apply Kalman filtering (see e.g. [5]). However, this method would require estimation of a larger number of parameters.

Empirically estimating the unknown model parameters

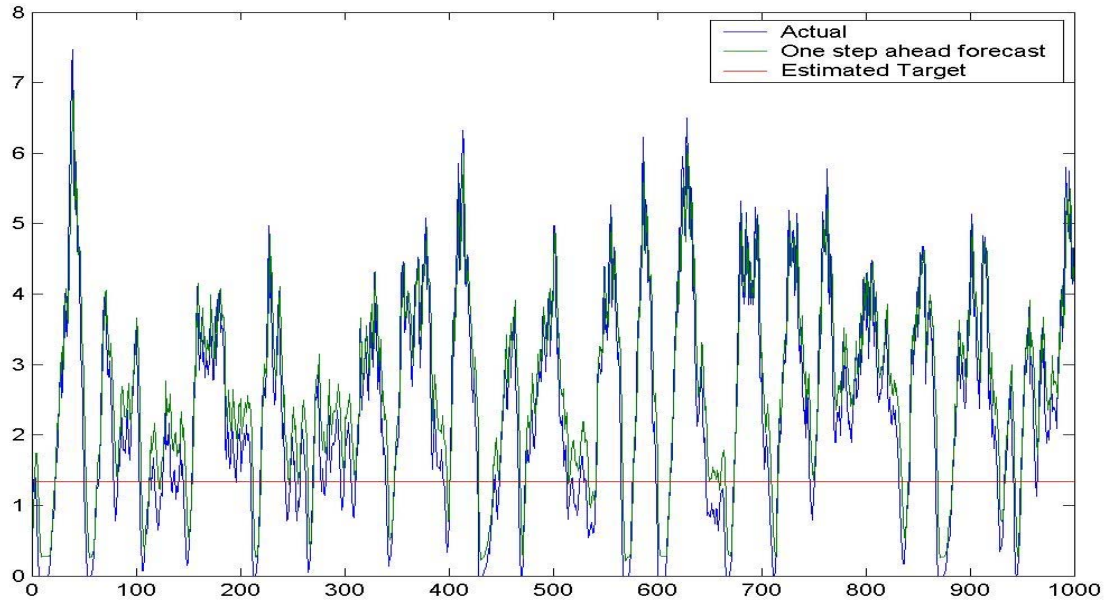
The complete forecasting model consists of three components: The decision model, the dynamic system model and the state-estimator. See Figure 2.

By estimating the staffing level (as explained in the previous section) and reapplying the quadratic optimization model (12)-(17) for each new observation of actual waiting-time, we can evaluate a one-step-ahead prediction.

By comparing with the actual waiting-time series, we can readily evaluate the residuals. Based on this, we estimated the eight unknown parameters ($\alpha, \beta, c, c_3, c_4, c_5, \overline{w}, g_0$) using an implementation of the nonlinear least-squares method available in MATLAB (called lsqnonlin).

From a mathematical point of view, it is notable that we do not have a closed-form analytical representation of the function being minimized.

Figure 3. Simulated waiting-time and a one-step-ahead prediction



We have a numerical representation, however. Similarly, we have no analytical expression of its gradient.

In addition, since evaluating any feasible solution involves solving a quadratic optimization problem, the process of searching for the parameters using the nonlinear least-squares approach is computationally intensive.

Similar as with other nonlinear estimation approaches, since we have not shown that the estimation problem is convex, we are not guaranteed to find the global optimum, only a local optimum.

Comparison with existing forecasting models

Due to limited availability of waiting-time data, we developed a simulator based on incorporating noise into the model equation for demand. By adding a noise term to the parameter α, demand was simulated as realizations of geometric Brownian motion.

Given a simulated realization of waiting-time (only), we attempted to estimate the parameters that generated the series. Since the simple decision model is deterministic (and not stochastic) and since information is lost since the staffing information is not used during the estimation and since the nonlinear least-square estimation process may have local minima, we could not hope to recover the exact set of original parameters.

However, when the noise-level was sufficiently high, we did observe that our nonlinear forecasting model outperformed linear dynamical models (including models of high dimension). An example is given in Figure 3.

Surprisingly, for a low level of noise, linear models fit the data set better than our nonlinear model. We attribute this result to the limitations of the proposed deterministic model. Based on this, an area for future research is to use a stochastic dynamic decision model (rather than a deterministic decision model). This entails a significantly higher computational cost, however.

We also compared the model using a real observed series of waiting-times. In this case, we found that a 78-order linear dynamical model fits the data set better than the nonlinear model. However, recall that the nonlinear model has only eight parameters (or degrees of freedom) while the linear model has 78. Another area for future research is to allow additional degrees of freedom in the nonlinear model.

In addition to the same limitations of the model as for the simulated data set, another possibility is that the observed waiting-time stays in the vicinity of a stable equilibrium that can be well approximated by a linear model. However, it is an interesting topic for future research to find data sets where the rational decision model possibly fits the data set better than traditional models.

Conclusions

Recall that if an intelligent agent (i.e. a rational decision maker in the sense of Operations Research) controls a process, then (by definition) it is theoretically possible to formulate a model of the actions of the decision maker using a decision model based on the principles of Operations Research.

We have suggested one such decision model formulation for waiting-time in healthcare and shown how the unknown parameters can be estimated. In particular, our model generalizes the closed-loop linear state-space model approach and takes into account that the decision maker's actions are constrained.

Based on the above, if a rational decision maker is present, we expect that the presented model is more accurate for prediction purposes than traditional models (in both a rational sense as well as an empirical sense). However, in the single available waiting-time data set, we did not observe that the presented model outperforms traditional models. However, this has been observed in simulations.

Additional data is needed to determine whether this means that there was no rational decision maker present (in this particular case), or that it is due to limitations in the forecasting model. To rule out the latter possibility, the use of stochastic dynamic decision models is an interesting area for future research.

In addition to serving as a forecasting model, the proposed model also has the potential to serve as an indicator whether a rational decision maker is present, or not. Such an indicator is useful for evaluating the value that can be added by introducing a decision support model.

References

[1] Model Building in Mathematical Programming, Third Ed., H.P. Williams, John Wiley & Sons, Chichester, 1993.

[2] Operations Research: Principles and Practices, Second Ed., A Ravindran, D. T. Phillips, J. J. Solberg, John Wiley & Sons, New York, 1987.

[3] Decision Theory: An Introduction to the Mathematics of Rationality, S. French, John Wiley & Sons, New York, 1988.

[4] Project Scheduling as a Stochastic Dynamic Decision Problem, T. Jorgensen, Doctoral dissertation, ISBN 82-471-0425-3, ISSN 0802-3271, Norwegian University of Science and Technology, Trondheim, Norway, 1999.

[5] Time Series Analysis: Forecasting and Control, Third Ed., G. P. Box, G. M. Jenkins, G. C. Reinsel, Prentice-Hall, London, 1994.

[6] Nonlinear Modeling and forecasting, M. Casdagli, S. Eubank (editors), Proceedings of the workshop on nonlinear modeling and forecasting held Sept. 1990 in Santa Fe, New Mexico, Proceedings Vol. XII, Santa Fe Institute, Addison Wesley, Redwood City, California, 1992.

[7] Time Series Prediction: Forecasting the future and Understanding the past, A. S. Weigend, N. A. Gershenfeld (editors), Proceedings of the NATO Advanced Research Workshop on Comparative Time Series Analysis held May 1992 in Santa Fe, New Mexico, Proceedings Vol. XV, Santa Fe Institute, Addison Wesley, Reading, Massachusetts, 1994.

[8] Computer Intensive Statistical Methods: Validation, Model Selection and Bootstrap, J. S. Urban Hjorth, Chapman & Hall/CRC, Boca Raton, 1999.

[9] Feedback Control Systems, Second Ed., C. L. Phillips, R. D. Harbor, Prentice-Hall, London, 1991.

[10] Introduction to Mathematical Control theory, Second Ed., S. Barnett, R. G. Cameron, Clarendon Press, Oxford, 1985.

[11] Optimal Control: Basics and beyond, P. Whittle, John Wiley & Sons, Chichester, 1996.

[12] Linear and Nonlinear Programming, S. G. Nash, A. Sofer, McGraw-Hill, New York, 1996.

[13] Stochastic Programming, P. Kall, S. W. Wallace, John Wiley & Sons, Chichester, 1995.

[14] The Past, Present, and Future of Macroeconomic Forecasting, Francis X. Diebold, Journal of Economic Perspectives, 12, 175-192.

Macroeconomic Issues

Chair: Norman C. Saunders, Bureau of Labor Statistics, U.S. Department of Labor

Fair-Weather Forecasters? An Assessment of the Private Sector's Macroeconomic Forecasts

Prakash Loungani, International Monetary Fund

This paper provides evidence on the properties of private sector macroeconomic forecasts using data for a sample of over 60 countries for the period 1989 to 2002. For forecasts of output growth, the main finding is that recessions and crises are almost never predicted in advance, leading to poor overall accuracy of forecasts. Recoveries tend to be forecast better, but double-dip recessions or periods of prolonged weakness are poorly forecast. The difficulty in forecasting turning points infects other forecasts, such as those of the current account balance and the government fiscal balance: the extent of reversals in these balances is typically underestimated. Inflation forecasts have tended to fare better than those of other macroeconomic variables.

Trade Liberalization in the Global Cotton Market

Stephen MacDonald, Leslie Meyer, and Agapi Somwaru
Economic Research Service, U.S. Department of Agriculture

World cotton prices fell to nearly unprecedented levels during the 2001/02 marketing year, causing distress to cotton producers and exporters worldwide. With the Doha Development Agenda's negotiations underway, discussion about the impact of trade barriers on the cotton sectors of developing countries has become more intense. A static computable general equilibrium (CGE) model forecasts that removing cotton tariffs and other trade barriers by all countries increases global welfare, but only slightly. Global welfare improves with liberalization, and the welfare of developing countries in aggregate is also forecast to improve. However, while some developing countries benefit, not all developing countries are forecast to see welfare gains.

The Theory of Forecasting: Dynamics Applied to Economics

Foster Morrison and Nancy L. Morrison, Turtle Hollow Associates, Inc.

Some forecasting methodologies seem to be ad hoc methods of extrapolation, with little or no connection to economic or other theories. The primary reason is that economics was formulated before computers made it possible to construct useful dynamical models. Results-oriented forecasters have developed techniques that can be implemented with the computing resources available. The dynamics of forecasting models is noise-driven, damped linear systems, which is appropriate for applying economic principles to the real world. A phase plane model of the business cycle illustrates the application of aggregation, data selection, and other dynamical principles to modeling, forecasting, and analysis.

TRADE LIBERALIZATION IN THE GLOBAL COTTON MARKET

Stephen MacDonald, Leslie Meyer, and Agapi Somwaru
Economic Research Service (ERS)
U.S. Department of Agriculture (USDA)

Abstract

World cotton prices fell to nearly unprecedented levels during the 2001/02 marketing year, causing distress to cotton producers and exporters worldwide. In a number of developing countries highly dependent on cotton for export earnings or where cotton is the primary cash crop, this distress was particularly acute. Global trade barriers to cotton are widespread, leading to some concern about the relationship between these trade barriers and global welfare. In particular, with the Doha Development Agenda's negotiations underway, discussion about the impact of trade barriers on the cotton sectors of developing countries has become more intense. A static computable general equilibrium (CGE) model finds that removing cotton tariffs and other trade barriers to cotton by all countries increases global welfare but only slightly. Global welfare improves with liberalization, and the welfare of developing countries in aggregate also improves. However, while some developing countries demonstrably benefit, not all developing countries see welfare gains. In addition to welfare, removing all global cotton trade barriers increases world trade in cotton.

Introduction

World commodity prices have been relatively low since the late 1990's, but while prices
of some major field crops recovered in 2002, the price of cotton remained well below recent averages (Figure 1). At the beginning of the 2002/03 marketing year, the world price of cotton was 33 percent below its 1990-94 average. In contrast, soybean prices were only 4 percent lower, while corn prices were 5 percent above their 1990-94 average and wheat prices were 24 percent higher (MacDonald and Meyer, 2002). Trade barriers to cotton are widespread and their removal would be expected to improve global welfare.

This study uses an 18 commodity, 40 country/region global computable general equilibrium model to assess world impacts of reducing global trade barriers to cotton and their effects on global welfare and trade. The model is static in its specification and uses the GTAP database, version 5.2. Aggregation was designed to account for major cotton producing countries—such as the United States and China—as well as major cotton producing regions—such as the European Union and sub-Saharan Africa.

Background

From the late 1990's to 2001, commodity markets were affected by the macroeconomic environment, particularly by the strong U.S. dollar, and a slowing world economy. According to the International Monetary Fund, world economic growth averaged 3.9 percent annually during 1994-97 but slowed with the Asian financial crisis and declined to 2.2 percent in 2001. On the other hand, as global commodity stocks shrank in 2002, prices of most major field crops recovered. Cotton prices, however, lagged the rebound for other commodities.

The impact of falling cotton prices was felt particularly acutely in a number of sub-Saharan African and Central Asian countries. A number of countries depend on cotton for a significant share of their export earnings, with Burkina Faso the most dependent in this sense, with cotton exports averaging almost 60 percent of its total exports during 1997-99. Chad, Mali, Benin, and Uzbekistan all had cotton accounting for at least 40 percent of export earnings. While cotton's share of exports is lower for other West African producers, in many cases cotton is one of the primary cash crops for farmers, giving it a large role in these countries' agricultural sectors (Levin, 2000 and Badiane et al 2002). During the 2001/02 marketing year, the vulnerability of low-income African countries became a concern, with advocacy groups such as OXFAM publicizing the plight of cotton farmers in these countries (Oxfam, 2002).

Global Trade Barriers for Cotton

Applied tariffs on cotton are often low, given cotton's role as an input for textile and apparel production. According to UNCTAD's TRAINS database, global applied cotton tariffs weighted by imports averaged only 2 percent. Bound tariffs are higher, with an import-weighted average of 21 percent (AMAD

database). In recent years various countries, Brazil and India for example, have raised their cotton import tariffs, suggesting that the applied tariffs in UNCTAD's database might not be appropriate for a longer-run analysis. The GTAP database embodies levels of protection higher than simple applied tariffs, taking into account other trade barriers (such as sanitary and phyto-sanitary measures) and giving a better measure of cotton trade policy. The average global level of import protection for cotton in the GTAP database is 5 percent.

This study assumes global export subsidies are negligible. In the GTAP database export subsidies are negligible, in part because it captures the policies in 1997 as the base year. China began subsidizing cotton exports in 1998/99, and none of the countries with subsidy reduction obligations have used subsidies in recent years. Only 4 WTO members have export subsidy reduction obligations under the URAA (Brazil, Colombia, Israel, and South Africa) and none have subsidized since 1995. China subsidized cotton exports in the years before it joined the WTO, sometimes explicitly, and other times in the form of tax rebates and reductions targeted at Xinjiang, the province supplying virtually all of China's cotton exports. As part of China's WTO accession, it agreed to eliminate all export subsidies for a wide range of commodities, including cotton.

Previous Research

Previous research on the impact of trade liberalization on the global cotton sector has been limited. Westcott and Price (2001) used USDA's FAPSIM model in an analysis of the impact of removing the U.S. marketing loan programs for all commodities. However, this study did not account for trade liberalization and did not estimate welfare effects.

The International Monetary Fund (IMF, 2002) used a computable general equilibrium model to simulate the impact of removing distortions for all commodities. This study, and a more recent version of the same analysis by Tokarick (2003), found negative welfare effects for some developing countries. Food-importing developing countries had welfare reductions in some cases although globally welfare improved by $100 billion, or 0.3 percent. However, these studies included policies liberalization in other commodities as well as cotton.

Likewise, USDA's Market and Trade Economics Division (MTED, 2001) used a CGE framework to examine the impact of removing all global policy distortions in agriculture, and found global welfare improved by about 0.2 percent, slightly less than the IMF and Tokarick.

The ICAC (2002) focused on cotton, but did not examine trade liberalization at all. Like Westcott and Price, the ICAC did not include welfare analysis.

Methodology

In this study the liberalization of world cotton markets is simulated in a CGE framework with the assumption that all tariffs are set zero and all non-tariff barriers are removed. All export subsidies are removed, but, as noted earlier, export subsidies in the GTAP database 5.2 are negligible. The model includes 18 commodities and 40 countries/regions. The model is static in its specification and uses the GTAP database, version 5.2.

Results

Removing all global import barriers to cotton trade raises global welfare (measured by Equivalent Variation, see ERS, 2001 for discussion). Global welfare increases by 0.03 percent, with welfare improving in both developed and less developed countries (Table 1). Welfare improves more for less developed countries than for developed countries, rising 0.05 percent. World cotton trade rises 9 percent, although there is little effect on other commodities.

Examining the welfare changes for specific countries provides some insight into the sources of welfare gains and losses (Table 2). Generally speaking, exporting countries enjoy welfare gains.

The world's largest exporter is the United States, but cotton trade is a very small part of the economy; therefore, its welfare gain is slight, 0.04 percent. The largest welfare gains are achieved by developing country cotton exporters. The Former Soviet Union is an aggregation of importing Russia and exporting Central Asian countries like Uzbekistan (the GTAP database does not disaggregate these regions), and is the world's second largest exporter. The Former Soviet Union's welfare increases by more than the United States. West Africa is the third largest exporter in the world, and the one most economically dependent on cotton exports.

Welfare gains in West and Southern Africa range from 2 to 3 percent, far exceeding changes experienced in any other region. This is consistent with the expectations of many analysts, but indicates

the dangers of generalization. This analysis strongly suggests that many sub-Saharan African countries would experience substantial economic gains if world cotton markets were liberalized, and these are some of the world's poorest countries. However, many other countries, including developing countries, would actually suffer setbacks due to liberalization according to this analysis, and in aggregate the gains to developing countries would be small, about 0.05 percent.

China's welfare declines, as does welfare in a number of Latin American countries. Two examples, Peru and Central America are included in the table. Japan and Korea also suffer very slight declines in welfare. This is a very heterogeneous group of countries, and generalizations about sources of welfare losses are difficult to draw, but it is worth noting that not all developing countries benefit. In most cases, it probably reflects distortions in other sectors of their economies that impede realization of the gains from global liberalization. As has been widely noted (e.g. IMF 2002) developing countries have many trade barriers among themselves, and a variety of preferential trading relationships. This further complicates generalizations about the impact of cotton trade liberalization. While global trade increases by a non-trivial amount, and welfare increases slightly, there are cases among both importing and exporting countries where welfare declines.

Conclusions

While some observers have suggested that barriers to cotton trade have large welfare implications, other analyses, including this study, suggest otherwise. Liberalizing world cotton trade has only small welfare effects, and not all developing countries benefit. However, global trade increases by a non-trivial amount, and certain developing countries experience more pronounced welfare gains.

References

American Embassy, Beijing, *Cotton: Annual Report*. Various years.

Badiane, O., Ghura, D., Goreux, L., and Masson, P., "Cotton Sector Strategies in West and Central Africa," World Bank Policy Research Working Paper 2867. July 2002.

Economic Research Service, Market and Trade Economics Division, *The Road Ahead: Agricultural Policy Reform in the WTO—Summary Report*, Agricultural Economic Report No. 797, U.S. Department of Agriculture, Economic Research Service. January 2001.

International Cotton Advisory Committee, *Production and Trade Policies Affecting the Cotton Industry*. July 2002.

International Monetary Fund, "How Do Industrial Country Agricultural Policies Affect Developing Countries?, " *World Economic Outlook*. September 2002.

Levin, A., "Francophone West Africa Cotton Update," *2000 Proceedings Beltwide Cotton Conferences*, National Cotton Council. 2002.

MacDonald, S. and Meyer, L., "Price Recovery Elusive for Cotton," *Agricultural Outlook*, U.S. Department of Agriculture, Economic Research Service. November 2002.

MacDonald, S. "The New Agricultural Trade Negotiations: Background and Issues for the U.S. Cotton Sector," *Cotton and Wool Situation and Outlook Yearbook*, CWS-2000, U.S. Department of Agriculture, Economic Research Service. November 2000.

Meyer, L., and MacDonald, S., *Cotton: Background and Issues for Farm Legislation*, CWS-0601-01, U.S. Department of Agriculture, Economic Research Service. July 2001.

Oxfam, *Cultivating Poverty*, Oxfam Briefing Paper 30. 2002.

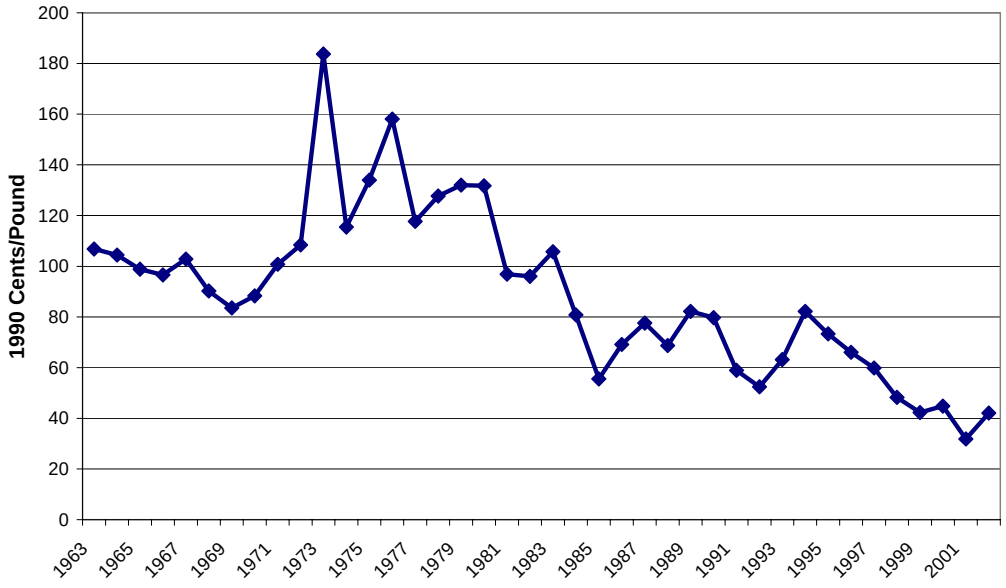
Tokarick, S., "Measuring the Impact of Distortions in Agricultural Trade in Partial and General Equilibrium," Eastern Economics Association meetings. February 2003.

Westcott, P. and Price, M., *Analysis of the U.S. Commodity Loan Program with Marketing Loan Provisions*, ERS Agricultural Economic Report No. 801, U.S. Department of Agriculture, Economic Research Service. April 2001.

Table 1. Changes Due to Cotton Trade Liberalization:	
Percent	
Global Welfare	0.03
Developed Country Welfare	0.03
Less Developed Country Welfare	0.05
World Cotton Trade	9.08

Table 2. Welfare Changes Due to Cotton Trade Liberalization Percent	
Central America	-.11
USA	.04
EU	.03
Southeast Asia	.19
China	-1.37
India	.13
Former Soviet Union	.22
Peru	-.36
West Africa	1.87
Southern Africa	3.31

Figure 1. World Cotton Price, 1963-2002



THE THEORY OF FORECASTING: DYNAMICS APPLIED TO ECONOMICS

Foster Morrison and Nancy L. Morrison
Turtle Hollow Associates, Inc.
PO Box 3639
Gaithersburg, MD 20885-3639 USA
Phone: 301-762-5652 Fax: 301-762-2044
Email: turtlehollowinc@cs.com

1. The Trouble with Economics

Economics has more in common with mathematics than it does with the experimental sciences, such as physics, chemistry, and biology. Economic theory tends to be axiomatic and deductive rather than empirical and inductive. Philosophy and ideology play major roles in economics, something it has in common with other social sciences.

Contrary to what lay people and even most scientists believe, mathematics is not eternal verities, but a cultural entity that depends strongly on philosophy. Euclidean geometry comprises most people's exposure to mathematical logic. But long before the time of Euclid, many cultures developed sophisticated mathematical methods that were highly empirical.

Calculus and most other parts of what comprise applied mathematics were developed on basically Euclidean axioms starting in the days of Newton and Leibniz. It was not until late in the 19th century that mathematicians found Euclid to be wanting, especially after the development of non-Euclidean geometry.

Georg Cantor offered a new axiomatic basis for mathematics through the introduction of set theory. Geometric axioms were replaced by abstract sets, whose elements (or members) might be tangible objects (apples or oranges) as well as abstractions (numbers or triangles). Set theory also produced a logical quagmire far worse than Euclidean geometry.

In the 1960s Errett Bishop (Bishop, 1975; Bishop and Bridges, 1985) developed a mathematical logic based on the arithmetic of the rational numbers rather than geometry or set theory. Other academic mathematicians extended the theory from the realm of abstract analysis into modern algebra. This was a sterling achievement, but it has had no impact whatsoever on science or technology, or for that matter, on most university departments of mathematics.

While Bishop was laboring to reformulate the basics of mathematics, a well-known inventor, Jay Forrester (1961), was attempting to rework economic analysis. Using the large mainframe computers that his magnetic memory donuts had made possible, he introduced differential equation modeling into the analysis of management systems, economics, and the environment. His research created a major controversy when his models demonstrated that exponential growth was not sustainable.

Bishop's theory, which he called constructive mathematics, is little known because it contributes almost nothing to science or technology except a level of rigor few people can appreciate. Forrester's models have had almost no impact because they did not provide predictions any better than those of econometric models, time series methods or *ad hoc* forecasts (Wils, 1988).

The system dynamics approach, as Forrester called it, failed to recognize that even fairly small systems of nonlinear differential equations can be highly unstable. This basic fact has been glorified with the name "chaos theory," although chaotic behavior is rare and difficult to identify. Once lauded as a "new science," its net contribution has been as illusory as that of system dynamics.

The nonsustainability of exponential growth is best demonstrated by looking at a table of the positive, integral powers of 2. A large, complex computer model only confuses the issue. The fact that predictability is the exception rather than the rule is readily apparent by looking at typical data sets, such as the stock market indices plotted in *The Wall Street Journal*. No sophisticated mathematics is needed to make that obvious, nor is any likely to provide results good enough for day traders.

2. Forecasting and Technical Analysis

Investors and business analysts have developed pragmatic methods for forecasting. Technical analysis involves looking for patterns in the data, mainly in hopes of spotting trend reversals. Some methods are simple and empirical, while others are elaborate and border on the mystical.

Forecasting has taken on a specific meaning and become an identifiable discipline practiced in the business world, academe and government. It is "mathematical" in that it utilizes numerical data to generate numerical predictions. But it is not "scientific" since the methods are largely empirical.

The basic forecasting method is linear prediction, where the data and forecasts are for discrete, uniformly spaced times

$$x_{n+1} = a_1 x_n + a_2 x_{n-1} + \ldots + a_m x_{n-m+1} \quad (1)$$

where x is the data series being forecast and a_i, $i = 1, \ldots, m$, the elements of a set of constant coefficients. Various forecasting methodologies are merely differing ways to select the set of coefficients

$$\mathbb{A} = \{a_i, i = 1, m\} \quad (2)$$

and its size, m. Alternative techniques include linear difference equations, statistical time series analysis, and *ad hoc* methods. Multivariate forecasts require that x be generalized to a vector **x** and the a_i, to a matrix **A**. When more than one previous value in a given time series are used in the forecast, the structure of **A** must be tailored accordingly.

All generally accepted forecasting methodologies are of the form

$$\mathbf{x}_i = \mathbf{A}\mathbf{x}_{i-1} \quad (3)$$

where the particular properties are expressed in the structure of the matrix **A** (Makridakis *et al*., 1998). Hence almost all forecasting models can be analyzed using linear algebra, as it is applied to difference equations. The primary result is that all the eigenvalues of **A**, some or all of which may be complex numbers, must have an absolute value of less than one. This, in turn, implies that the value of **x** decays to **0** as i increases, which means for all times farther in the future.

3. The Dynamics of Forecasting

Is (3) the model for the one or more time series being predicted? In general, no. Least squares is used to fit (3) to the data, but the residuals are assumed to be "noise" in the data rather than errors in the observations. There also may be errors in the observations, but these are thought to be small to the point of being negligible for econometric time series. Hence, the economic model used by the forecasting profession is

$$\mathbf{x}_i = \mathbf{A}\mathbf{x}_{i-1} + \mathbf{n}_{i-1} \quad (4)$$

where **n** is a vector time series with components consisting of "random" noise. The actual statistical properties of **n** may be less than perfectly "random," which itself is not easy to define (Kac, 1983, 1984).

The model (3) is the linearized version of the basic assumption of economics, that all markets tend to move toward equilibrium (**x** = **0**). Model (4) brings in a well-know reality factor: they never get there.

What are the causes of noise? For one thing, there are far too many variables in the complex system comprising the exchange economy, the biosphere, and whatever astronomical effects may be significant. Many of the variables have discrete numerical values. Others may change often or even continuously within the sampling intervals.

Models (3) and (4) do not allow for population or economic growth, so in many cases some or all of the components of **x** are detrended logarithms of the actual data series. Polynomials or other functions of time may be used as trend models, but low pass filters are more stable for extrapolation. The problems encountered with moving averages can be avoided by using the ramp filter, which was designed to have the weighted average associated with the end of the interval rather than the middle (Morrison and Morrison, 1997).

Model (4) can be generalized to produce more advanced or more flexible forecasting methodologies. Deviations of the components of **n** from being stationary random noise could be exploited to improve forecasts. The difference equation could be replaced by a differential equation for data that have varying or inconsistent sampling rates. Nonlinear terms are also easier to add to ODEs (ordinary differential equations). Whether any of these added capabilities really improves a given forecast can be ascertained only by the attempt.

4. Designing a Forecasting Model

Perhaps the most important part of designing a forecasting model is choosing which variables to use. Rather than assuming that one wants to forecast a particular time series, the best way to start is by asking what decisions have to be made.

Suppose a business person wants to know whether to add capacity. If the economy is growing, this might be a good idea. If not, the money required should not be spent or borrowed until there is a reasonable likelihood that the capacity will be needed. Public officials need the same information when determining fiscal policy or setting interest rates.

A forecast of GDP (Gross Domestic Product) would seem to satisfy these needs. The first challenge is that GDP is very difficult to measure. Totaling up every transaction in an economy is impossible. A lot of data is collected and a lot of assumptions have to be made. The US Department of Commerce publishes a figure for each quarter, with two revisions over the following two months. Wholesale revisions of the numbers may be made at later dates.

The exact value of GDP is of no great interest, however. It is the change every quarter that really is crucial. So two numbers of similar size have to be subtracted, which in the worst possible case leaves no significant digits whatsoever.

Why can growth rates be assumed to have any validity at all? For one thing, the process of GDP determination is highly standardized. Hence it can be assumed that all the measurement biases are fairly consistent from quarter to quarter. All the errors are undoubtedly correlated to a high degree so that the growth rates obtained are adequately reliable. Other econometric data tend to support the estimated growth rates.

Some economists have developed large-scale econometric models, most of which are proprietary and available only to subscribers paying substantial fees. An alternative approach is to try to build the most cost-effective model rather than the best one possible. In any event, it is always best to start with a minimal model and then add more variables and complexity on a marginal basis.

5. The Role of Chaos Theory

James Gleick (1987), other popular writers, and many scientists have touted the great advances to be made by applying chaos theory and nonlinear dynamics to a host of problems. Few, if any, of these benefits have materialized. Why? Because the specialists in various fields, such as business and economic forecasting, already knew that prediction was difficult to impossible. These theories merely explain why and do not offer any solutions to the problems.

What causes chaos? In the well-known Lorenz model, the culprits are unstable equilibrium points. Another important property is that the solutions to the 3 ODEs cannot escape from the vicinity of the unstable equilibria, but keep returning to them again and again (Morrison, 1991).

You do not have to be an expert on ODEs, or even know calculus, to understand chaotic behavior. A simple example is provided by an old-fashioned pin ball machine. All the many bumpers in the machines act just like the unstable equilibria in the Lorenz equations. The direction in which they deflect the ball is extremely sensitive to the path along which the ball approaches. The error of prediction of the trajectory of the ball is magnified enormously by each collision with a bumper. But the ball is contained by the sides of the machine so that eventually it drops through a hole and out of play.

The exchange economy works more like a pin ball machine than a carefully crafted chronometer. Consumers are fickle. Fads sweep through various markets. Advertising agencies attempt to manipulate or create fads. Factories make too much or too little. Stores buy too much or too little. Banks lend too much or too little. Government fiscal policies are set by many agendas, and long-term stability is not dominant among them.

What causes "random" noise? On the molecular scale, Brownian motion is chaotic, a 3-dimensional version of the pin ball machine. But in the aggregate, Brownian motion is stochastic. In other words, a long string of chaotic events looks "random," especially when the sampling rate is much longer than the time between chaotic events. In the exchange economy, every transaction is a chaotic event. But what economists can measure is only aggregates of large numbers of transactions.

When designing a model of a dynamic system it is best to create a qualitative theory first. Next is the simplest possible quantitative model. In general, the largest and most complex model that is feasible will be vastly smaller and simpler than the system under consideration. The resources for data collection are always limited; the budget for development is barely adequate and, even if the data were perfect, the computations could not be exact.

6. Selecting the Number of Variables

What should be the number of distinct variables in **x** in (3) and (4)? (The rules of matrix algebra will make the dimension of **x** larger than this if more than the one previous value is being used in the forecasting methodology.) Ideally all the variables should be independent. If any one can be represented by a linear combination of the others, it should be omitted.

This criterion can be based on statistics, approximation theory, or both. Correlation and linear dependence are the same things computationally. To make this clear one must think of the various time series as vectors, with each observation as a different component, thus

$$\mathbf{y} = [y(t_n), y(t_{n-1}), \ldots, y(t_1)]^T \qquad (5)$$

The dimension of such vectors, n, can become quite large, so it may be wise to look at segments of the data as well as the entire span for which it is available.

The designers of economic indicators chose 3 as the number of variables to be used. Data series were categorized as leading, coincident, lagging, and irrelevant (or uncorrelated). However, there was no effort expended in creating a 3-variable econometric model. Instead the data series were used to create indices that might be helpful in judging where in the business cycle the economy happened to be (*Handbook*, 1984).

7. Aggregation

A model need not be constructed using observable data series. Orbit computations have been done with Keplerian elements when perturbation methods are used and with Cartesian coördinates for numerical integrations. These are cases of selecting variables to facilitate complicated computational procedures.

Weighted averages are commonly used in economic and market analysis. Stock market averages, such as the S&P 500 and the several Dow indices are familiar to many. The indices of leading, coincident, and lagging indicators developed by the US Department of Commerce use methods somewhat more elaborate than weighted averages (*Handbook*, 1984).

What is the benefit of using weighted averages or other sorts of indices? If the "noise" parts of the various original data series are uncorrelated (or weakly correlated), then a weighted average (or other index) may have a better signal-to-noise ratio. In fact, maximizing the signal-to-noise ratio can provide a design criterion for the average (or index).

The art of modeling consists of creating the best feasible model from available data sets. Only in a few special cases in celestial mechanics and other physical sciences can one achieve a level of predictability that might be called "deterministic." For things like economic systems it is necessary to determine the optimal size for the dimension of the state vector, **x**, as in (4). Then the predictability may be improved by aggregation.

Chaos and randomness set a limit to the predictability of all dynamic systems, even the classical example of planetary orbits. In some cases there are limits to variability. The economy can grow only so fast, though it would be difficult to establish a rigorous upper bound. On the other hand, economic decline can be quite rapid.

The most conspicuous current example of the limits to predictability is afforded by the global warming controversy. No one can determine with sufficient precision how much of the warming is being caused by human activities, such as the burning of fossil fuels, and how much might be due to geophysical or astronomical causes. The debate is being driven more by ideology and short-term economic interests than science. Climatology may never become precise enough to answer the question to the satisfaction of all, so an alternative approach, such as risk analysis, may be in order.

8. A Phase Plane Model of the Business Cycle

A less than ideal example is provided by a phase plane model of the business cycle (Morrison and Morrison, 1997, 2001, 2002). Where this model is seemingly deficient is in the fact that the predictions are generated by the forecasting model (4) rather than an econometric model. [In practice, the forecasts are being made by a *linear filtering* model, which assumes the same dynamics as (4), but obtains the coefficients a_i using autocorrelations generated by FFTs.]

Actually, a forecasting model is a discrete, linear econometric model. The model (3) is merely the dynamical version of supply and demand. One classic example is the "cobweb model" (Luenberger, 1979; Morrison, 1991). By replacing (4) (or the linear filtering predictor) with ODEs and adding a few nonlinear terms, something that aspires to the status of econometric would be obtained. Whether any improvements in performance would be more than marginal cannot be determined in advance.

The number of variables, 3, was selected by the developers of the indices of indicators. They also provided the aggregation rules. These procedures have been altered over time and are now maintained by The Conference Board in New York City.

The frequent revisions offer the advantage of constant improvements, but the penalty is that it is correspondingly difficult to compare the analysis of past decades with what is going on today. The GDP numbers share this difficulty. However, our informal analysis indicates that the phase angles in the business cycle model vary little from revision to revision and that the relative changes in the radial coördinates enjoy the same kind of stability. For example, one recent revision made the radii about half their previous values, but the shapes of the cycle graphs changed imperceptibly.

The original cycle models were created using only the leading and coincident indices (Morrison and Morrison, 1997). The detrended lagging index was plotted in the third dimension (z-axis) and the graph was obtained by projection onto the plane that best fit the set of points. This did alter the shape of the graphs noticeably, but the qualitative features were unchanged (Morrison and Morrison, 2001).

The graph in Figure 1 differs little from that presented at the 2002 Federal Forecasters Conference (Morrison and Morrison, 2002), except that more data has been added. The model is still languishing in the fourth quadrant recovery zone, despite modest GDP growth.

Worth noting is that the model was not especially robust during the technology "bubble" of the 1990s. The activity in the expansionary first quadrant was rather weak. What is the model saying? Perhaps that the large sums of money being invested were not forming capital, but merely paying the current expenses

of companies doomed to fail. The modest real expansion has been followed by an equally modest recession. This may be a benefit of not having excessive capital formation, but there seems to be a possibility that the USA will endure the sort of stagnation now being suffered by Japan.

While "high tech" has not grown to the extent hoped for, long established industries have been migrating to low-wage countries. This certainly is one reason that growth has ceased in Japan and now the phenomenon has spread to the USA and parts of the European Union, especially Germany.

Difficulties with further economic growth are also demonstrated by traffic problems in the Washington, DC area. Previous economic and population growth have far exceeded the expansion of the highway system. Mass transit, such as the Metrorail system, has grown less effective as places of employment have decentralized. Poor planning is partly to blame, but the scale of growth has just become too large to manage.

The so-called "war on terrorism" and the expanded expenditures for the occupation of Iraq, Afghanistan, and who knows what else will be placing severe strains on the federal budget. With little or no real growth, all these factors add up to a difficult future for at least the next 10 years. If this scenario is indeed the one that plays out, then analysts will want to be looking at the trends as well as the business cycle.

7. References

Bishop, E., The crisis in contemporary mathematics, *Historia Math*., *2*, 507-517, 1975.

Bishop, E. and D. Bridges, *Constructive Analysis*, Springer-Verlag, Berlin, 1985.

Forrester, J.W., *Industrial Dynamics*, The M.I.T. Press, Cambridge, MA 1961.

Gleick, J., *Chaos: Making a New Science*, Viking Penguin, New York, 1987.

Handbook of Cyclical Indicators: a Supplement to the Business Conditions Digest, 1984, US Dept. of Commerce, Bureau of Economic Analysis, Washington, DC.

Kac, M., What is random?, *American Scientist*, *71*, 405-406, 1983.

Kac, M., More on randomness, *American Scientist*, *72*, 282-283, 1984.

Luenberger, D.G., *Introduction to Dynamic Systems: Theory, Models and Applications*, Wiley, New York, 1979.

Makridakis, S., S.C. Wheelwright and R.J. Hyndman, *Forecasting: Methods and Applications*, third edition, John Wiley & Sons, Inc., New York, 1998.

Morrison, F., *The Art of Modeling Dynamic Systems: Forecasting for Chaos, Randomness, and Determinism*, Wiley-Interscience, New York, 1991.

Morrison, F. & N.L. Morrison, A phase plane model of the business cycle, *The 8th Federal Forecasters Conf. 1996 & The 7th Federal Forecasters Conf. - 1994: Combined Papers & Proceedings*, US Dept. of Education, NCES 97-341, pp. 93-112, Debra E. Gerald, editor, Washington, DC, 1997.

Morrison, F. & N.L. Morrison, An improved phase plane model of the business cycle, *The 11th Federal Forecasters Conf. - 2000*, US Dept. of Education, NCES 2001-036, pp. 183-191, Debra E. Gerald, editor, Washington, DC, 2001.

Morrison, F. & N.L. Morrison, Forecasting the business cycle with polar coördinates, *The 12th Federal Forecasters Conf. - 2002*, pp. 185-191, Debra E. Gerald, editor, Washington, DC, 2002.

Wils, W., *Overview of System Dynamics the World Over*, Croon DeVries, Parkweg 55, 3603 AB Maarssen, Netherlands, 1988.

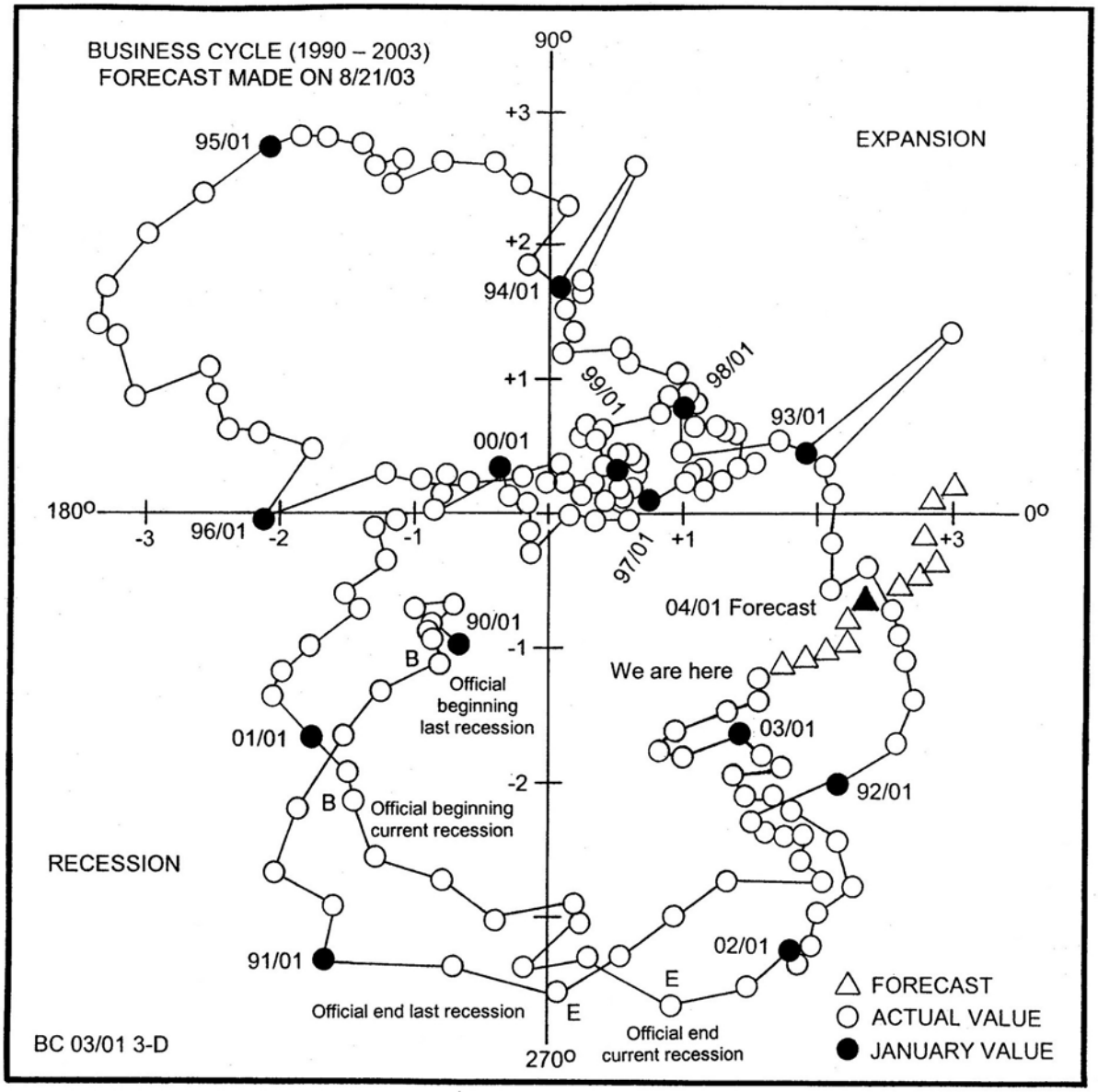


Figure 1. The business cycle model is a phase plane plot of a weighted mean of the detrended leading and detrended lagging indicators as x-coordinate and detrended coincident indicator as y-coordinate. Normal cycles follow a counterclockwise roughly circular path with occasional stalls and reversals. Time is indicated along the cycle path. The data have a 2-month lag. Expansions occur between 0° and 90° and recessions between 180° and 270°. Other angles denote transition (90°-180°) and recovery (270°-360°=0°) periods. An "official" (NBER) beginning of a recession is indicated by a label "B" and an end by "E".

The polar coordinate forecast (Δ - triangle) moves more rapidly than a Cartesian alternative (Morrison and Morrison, 2002). However, this does not necessarily mean that short-term forecasts using polar coordinates will be the more accurate ones. The divergence of two forecasts can provide some room for intuitive judgment.

FEDERAL
FORECASTERS
CONFERENCE