

A DC Coincident Index in the Age of Big Data

Fitzroy Lee

October 13, 2016

Coincident indexes

- ▶ A **coincident economic index** combines a set of comoving economic series into a single measure of the current state of the economy.
- ▶ Examples of coincident indexes:
 - ▶ The Conference Board (TCB) coincident economic index for the U.S.
 - ▶ The Philadelphia Federal Reserve indexes for each of the 50 states
 - ▶ The Chicago Fed National Activity Index (CFNAI)

Why a DC coincident index?

- ▶ Answer to question: How is the DC economy doing right now?
- ▶ DC GDP available annually and not a good measure for DC
- ▶ Personal income available quarterly with several months lag
- ▶ U.S. indexes do not necessarily help; local business cycle may be different
- ▶ Dozens of economic measures tracked in *DC Economic Indicators* and *DC Economic and Revenue Trends* have different lags and leads relative to the current state of the economy

Big Data challenges

1. **Mixed-frequency:** Economic indicators are available with different frequencies (daily, weekly, monthly, quarterly, each second (e.g., stock prices)).
2. **“Jagged edge” data sets:** Even if all the series are the same frequency economic indicators are released at different dates, leading to missing data for some series in the last period of the data set.
3. **High dimensionality:** Availability of a vast and growing number of data series creates opportunities as well as challenges.

Constructing a coincident index – Then

- ▶ Not model based (not based on a probability distribution)
- ▶ Constructed as a simple average of component economic series
- ▶ Drawbacks include:
 - ▶ No explicit reference to target variable
 - ▶ Fixed weighting scheme
 - ▶ Unable to produce standard errors for leading index
 - ▶ Require frequent revisions
- ▶ Big advantage: simple to construct, explain, and interpret

Constructing a coincident index – Now

- ▶ Two main classes of models provide a statistical framework
 - ▶ Dynamic factor models – Stock and Watson (1989)
 - ▶ Markov switching models
- ▶ A related technique: Nowcasting

Dynamic factor models (DFMs)

$$X_t = \lambda(L)f_t + e_t$$

$$\psi(L)f_t = \eta_t \quad \eta_t \sim \mathcal{N}(0, \sigma_f)$$

$$D(L)e_t = \zeta_t \quad \zeta_t \sim \mathcal{N}(0, \Sigma)$$

where $\Sigma = \sigma^2 I$ and $\sigma^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2)$

- ▶ An unobserved common factor as well as shocks idiosyncratic to each component economic series drive the comovement of a set of economic variables
- ▶ Idiosyncratic shocks are assumed to be uncorrelated across the component variables
- ▶ The coincident economic index is the unobserved common factor estimated from the dynamic factor model

DFMs – expanded notation

$$X_{1t} = \lambda_1 f_t + e_{1t}$$

$$X_{2t} = \lambda_2 f_t + e_{2t}$$

⋮

$$X_{Nt} = \lambda_N f_t + e_{Nt}$$

where $\lambda(L) = (\lambda_1(L), \dots, \lambda_i(L), \dots, \lambda_N(L))'$,

with $\lambda_i(L)f_t = \lambda_{i0}f_t + \lambda_{i1}f_{t-1} + \dots$

Also, $\psi(L)f_t = \eta_t \Leftrightarrow f_t = \psi_1 f_{t-1} + \psi_2 f_{t-2} + \dots + \psi_p f_{t-p} + \eta_t$

DFMs – expanded notation

and $D(L)e_t = \zeta_t \Leftrightarrow$

$$e_{1t} = d_{11}e_{1t-1} + d_{12}e_{1t-2} + \cdots + d_{1q}e_{1t-q} + \zeta_{1t}, \quad \zeta_{1t} \sim \mathcal{N}(0, \sigma_1^2)$$

$$e_{2t} = d_{21}e_{2t-1} + d_{22}e_{2t-2} + \cdots + d_{2q}e_{2t-q} + \zeta_{2t}, \quad \zeta_{2t} \sim \mathcal{N}(0, \sigma_2^2)$$

\vdots

$$e_{Nt} = d_{N1}e_{Nt-1} + d_{N2}e_{Nt-2} + \cdots + d_{Nq}e_{Nt-q} + \zeta_{Nt}, \quad \zeta_{Nt} \sim \mathcal{N}(0, \sigma_n^2)$$

Markov switching models

- ▶ One criticism of the Stock/Watson dynamic factor model based coincident index is that model parameters do not change over the business cycle
- ▶ In the Hamilton (1989) Markov switching model the growth rates of the variables are conditioned on the state of the economy—whether the economy is expanding or contracting
- ▶ Switching between the two regimes is modelled as a Markov process, a stochastic model where the future of a process is based solely on its current state
- ▶ Diebold and Rudebusch (1996) and Kim and Nelson (1998) are two notable examples of Markov switching model variants of the Stock/Watson coincident index
- ▶ This presentation focuses on DFMs

Estimating DFMs

- ▶ Several ways to estimate DFMs:
 1. Recast as a state space model (SSM) and estimate by MLE and Kalman filter
 2. Extract common factor by principal components analysis (PCA)
 3. Combine 1 and 2 into hybrid PCA/SSM
 4. Estimate SSM and common factor together in a Bayesian framework
- ▶ Stock and Watson (2011) divides the three frequentist approaches into “generations” to emphasize that each approach is the result of successive refinements over time

- ▶ Recast the dynamic factor model as a state space model:

$$X_t = \Lambda F_t + e_t$$

$$F_t = AF_{t-1} + \nu_t, \quad \nu_t \sim \mathcal{N}(0, W_t)$$

$$D(L)e_t = \zeta_t \quad \zeta_t \sim \mathcal{N}(0, \Sigma)$$

$$\text{where } F_t = (f_t, f_{t-1}, f_{t-2}, \dots, f_{t-p})$$

- ▶ Estimate the parameters by MLE and use the Kalman filter and smoother to estimate the factor
- ▶ With normality assumptions, Kalman filter averages across series and time to extract the unobserved factor
 - ▶ Kalman filter: adaptive moving average technique that efficiently extracts a signal from noisy measurements

Features of SSM/MLE

- ▶ Handles missing data like that in mixed frequency and “jagged edge” or “ragged edge” data sets
- ▶ Restricted to low-dimensional (small N) parametric models because MLE estimation of the parameters rely on nonlinear optimization, which becomes computationally intractable with a large number of parameters

PCA

- ▶ PCA is a transformation that converts a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components
- ▶ The first principal component accounts for as much of the variability in the data as possible, and each succeeding component in turn has the highest variance possible while being uncorrelated with the preceding components
- ▶ The PCA estimate of the common factor, $f_t = N^{-1}\hat{\Lambda}X_t$, is a weighted sum of the component indicators, with weights, $\hat{\Lambda}$, derived from a decomposition of the sample covariance matrix of X_t , $\hat{\Sigma}_X = T^{-1} \sum_{t=1}^T X_t X_t'$
- ▶ The PCA estimate of the common factor is the principal component that explains the largest amount of the variation among the indicators

Features of PCA

- ▶ Employing nonparametric cross-sectional averaging, PCA can extract the common factor from large dimensional (large N) data sets
- ▶ More robust and is easier to apply than MLE estimation of a state space model
- ▶ Cannot handle mixed frequency or real-time data
- ▶ Fails to exploit time variations in a data set

Hybrid PCA/SSM

- ▶ Combine the high-dimensional data handling capability and robustness of PCA with the mixed frequency data handling capability and statistical efficiency—ability to exploit variations in the data across series and time—of state space models
- ▶ Implemented in two steps as described in Giannone, Reichlin, and Small (2008)
 1. Extracts an estimate of the common factor \hat{F}_t using PCA
 2. Takes the estimated factor \hat{F}_t from the first step to estimate by regression analysis the parameters Λ, A, V_t, W_t of the state space representation
- ▶ With the state space model now fully parameterized, the Kalman filter/smoothing exploits the time variations of the data series to get an improved estimate of the factor F_t

Features of PCA/SSM

- ▶ High-dimensional data handling
- ▶ Mixed frequency data handling
- ▶ Real-time updating
- ▶ Robust algorithm
- ▶ Statistical efficiency

SSM/Bayesian

- ▶ With distributional assumptions on the model parameters, use Kalman filter and Baye's rule to form joint posterior distribution of parameters and factors
- ▶ Estimate model parameters and factor together using Markov Chain Monte Carlo (MCMC) techniques to sample the conditional posterior distribution for parameters and factors

Features of SSM/Bayesian

- ▶ Handles high-dimensional data by relying on computationally more tractable integration (summation over a range) to obtain parameter and factor estimates
- ▶ Handles mixed frequency because it uses the Kalman filter
- ▶ Cannot update in real-time because when new data arrive must run the MCMC algorithm again to re-estimate parameters and factor together
- ▶ Can estimate nonlinear/non-Gaussian dynamic factor models
- ▶ Allows the incorporation of prior information into the model

Big Data features of the 4 methods

Big Data feature	SSM/MLE	PCA	PCA/SSM	SSM/Bayesian
High-dimensionality		✓	✓	✓
Mixed frequencies	✓		✓	✓
Real-time updating	✓		✓	

High dimensionality: PCA versus Supervised PCA

- ▶ Almost 200 DC economic indicators available on FRED
- ▶ PCA is a fast, efficient dimension-reduction techniques, but could pre-selecting variables improve forecast performance?
- ▶ Supervised PCA: Uses a subset of the predictors selected based on their association with the target variable. (Bair, Hastie, Paul, and Tibshirani, 2006)
- ▶ Bai and Ng (2008) finds improved forecasts for economic time series for supervised PCA over PCA

Variable selection strategies: 'Hard thresholding' versus 'soft thresholding'

- ▶ 'Hard thresholding': uses a statistical test to determine if the i -th predictor is significant without regard for the other predictors being considered.
- ▶ 'Soft thresholding': uses a shrinkage method such as LASSO to select a subset of variables. Sets coefficient of weak predictors to zero. Selected variables are the ones with nonzero coefficients.

Hard thresholding

Considers the bivariate relation between y_{t+h} and X_{it} after controlling for other predictors W_t (such as lags of y_t) since a simple autoregressive forecast is always available as an alternative forecasting procedure.

Hard thresholding procedure

The details (Bai and Ng 2008):

1. For each $i = 1, \dots, N$, perform a regression of y_t^h on W_{t-h} . W_{t-h} includes a constant and lags of y_t . Let F_i denote the F -statistic associated with X_{it-h} .
2. Obtain a ranking of the marginal predictive power of X_{it} by sorting $|F_1|, |F_2|, \dots, |F_N|$ in descending order.
3. Let k_α^* be the number of series whose $|F_i|$ exceeds a threshold significance level, α ;
4. Let $x_t(\alpha) = (x_{[1t]}, \dots, x_{[k_\alpha^* t]})$ corresponding set of k_α^* targeted predictors. Estimate f_t from $x_t(\alpha)$ by the method of principal components to yield \hat{f}_t .

Soft thresholding

- ▶ Hard thresholding sensitive to small changes in the data because of the discreteness of the decision rule.
- ▶ Also, because it selects predictors one at a time, it cannot take into account the information in other predictors.
- ▶ May select variables that are too 'similar, reducing the benefits of model averaging the effectiveness of which derives from pooling over variables that bear distinct information from each other.
- ▶ 'Soft thresholding methods perform subset selection and shrinkage simultaneously.

LASSO

- ▶ LASSO (Least Absolute Shrinkage Selection Operator) uses penalized regressions to drop uninformative regressors.
- ▶ Let $RSS(\alpha, \beta)$ be the sum of squared residuals of the regression y_t on *all* the regressors, W_{it} and X_{it} . The LASSO estimator is the solution to:

$$\min_{\alpha, \beta} RSS \text{ subject to } \sum_{j=1}^N |\beta_j| \leq c$$

where the parameter $c \geq 0$ controls the degree of shrinkage.

- ▶ The LASSO penalty ($|\beta_j|$), an L_1 penalty, unlike the L_2 penalty (β^2) of a ridge regression, sets some coefficients exactly to zero.

Data

- ▶ Almost 150 DC economic indicators from FRED spanning various periods from 1948 until latest month with a mix of weekly, monthly, and quarterly series
- ▶ Most variables are logged, differenced, then normalized (exceptions are rates, like the unemployment rates, which are not logged or differenced, but still normalized)
- ▶ Signs are reversed for variables with an inverse relationship with the target (like unemployment rate)

All Indicators

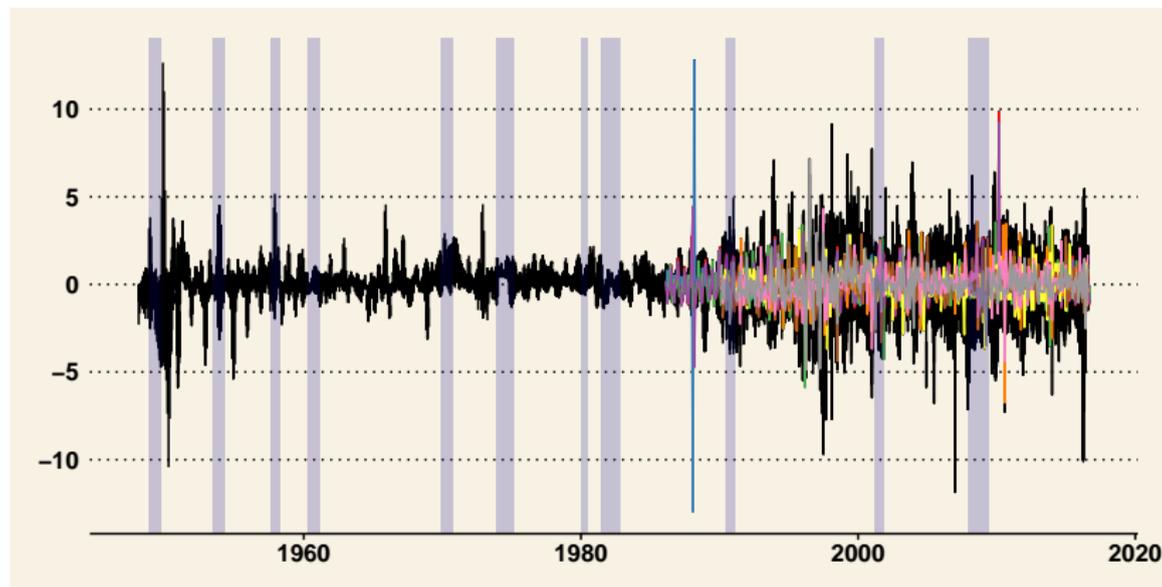


Figure 1:DC Economic Indicators (logged, differenced, standardized)

Index: All Indicators

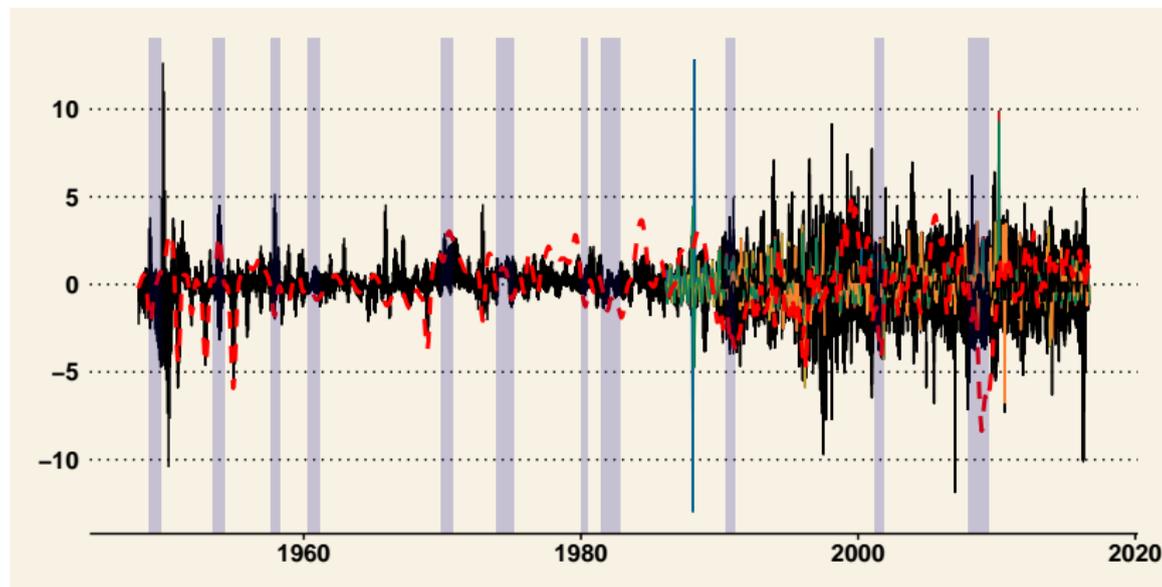


Figure 2: Index constructed from all indicators

Compared to DC personal income

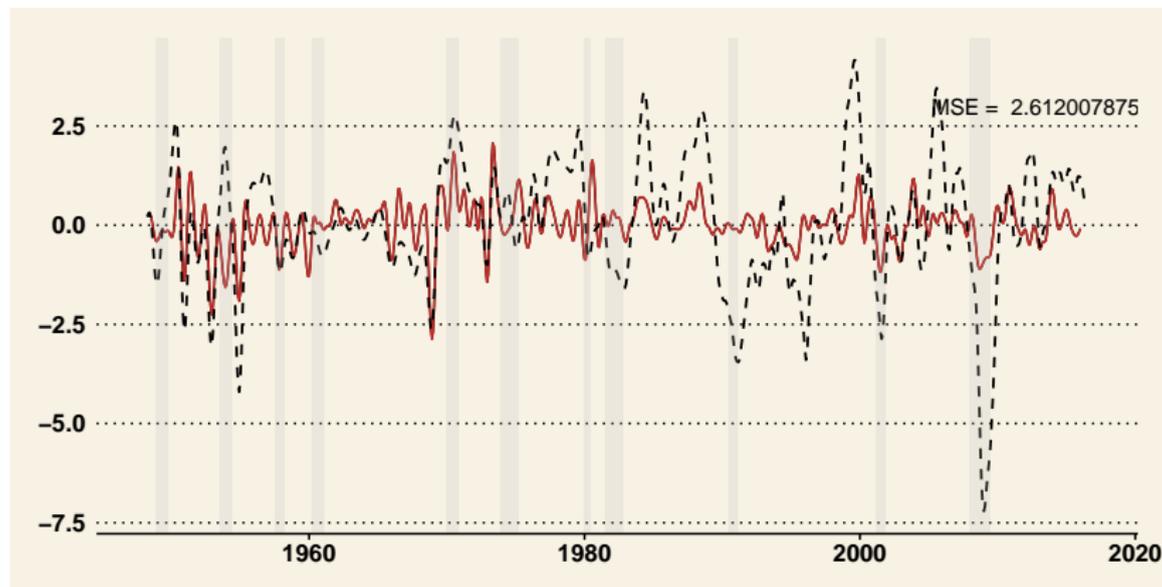


Figure 3: Change in Index vs Change in DC personal income

Selecting variables: Hard thresholding

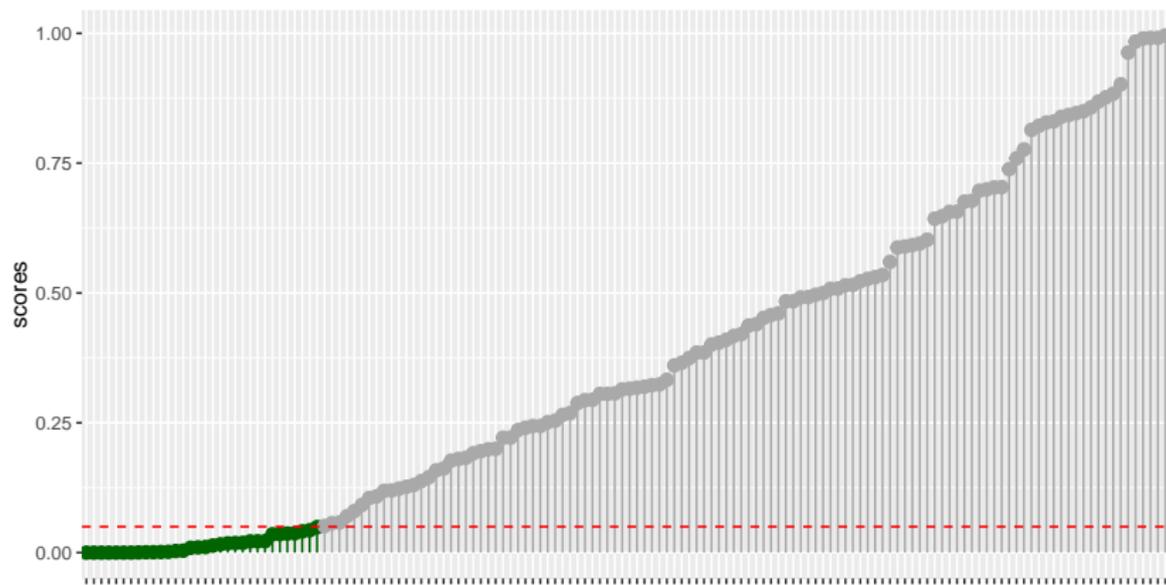


Figure 4: Selected indicators for threshold = 0.05

First 16 selected variables: Hard thresholding

Selected indicators	NA
DC.Continued.Claims	DC.State.Unemployment.B
DC.Initial.Claims	DC.Dividends..Interest.and.
DC.Insured.Unemployment.Rate	DC.Other.Labor.Income
DC.Employment..Other.Services	DC.Transfers.less.State.Un
DC.Employment..Legal.Services	DC.Earnings..Administrativ
DC.Employment..Religious.Civic.Prof..Orgs.	DC.Earnings..Finance.and.I
DC.Net.Earnings.by.Place.of.Residence	DC.Earnings..Professional.a
DC.Earnings.by.Place.of.Work	DC.Earnings..Retail.Trade

Index of indicators selected by hard thresholding

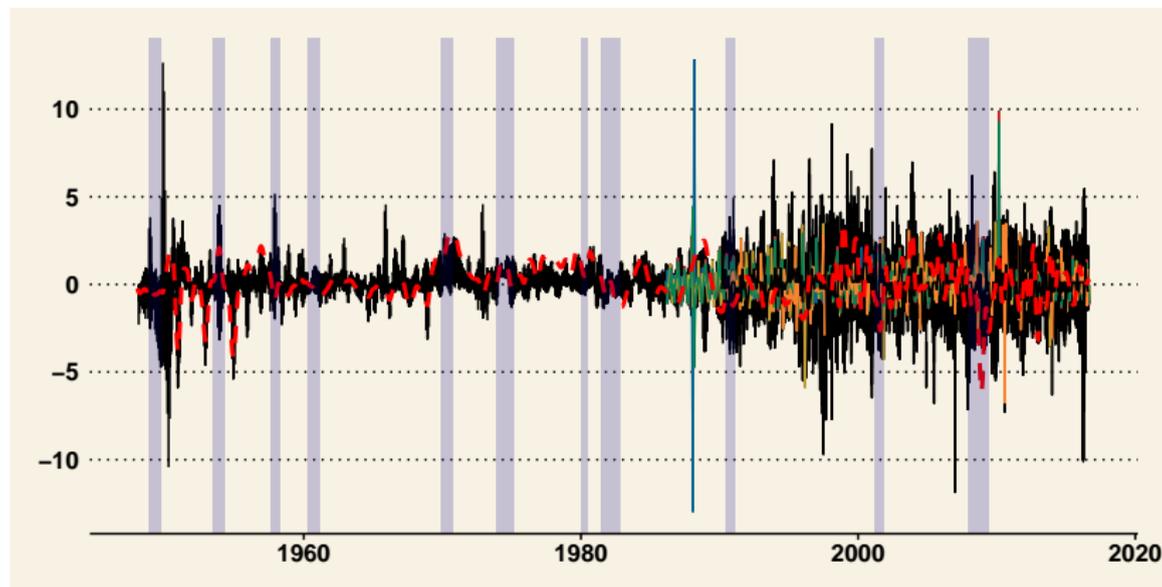


Figure 5: Index constructed from selected indicators (hard threshold)

Index (hard threshold) compared to DC personal income

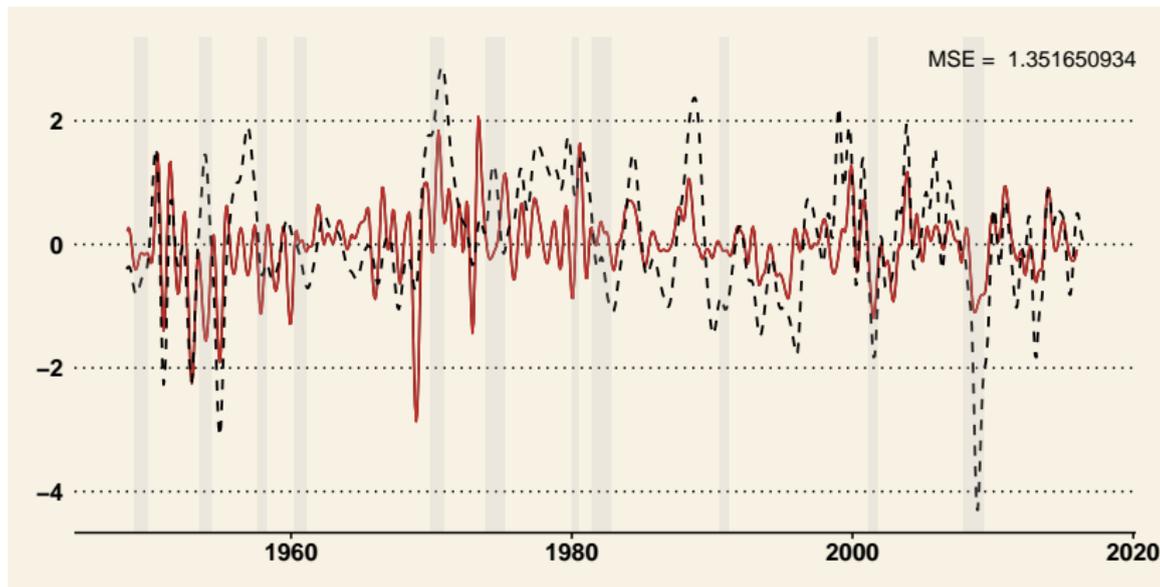


Figure 6: Index vs DC personal income

Selecting variables: Soft thresholding

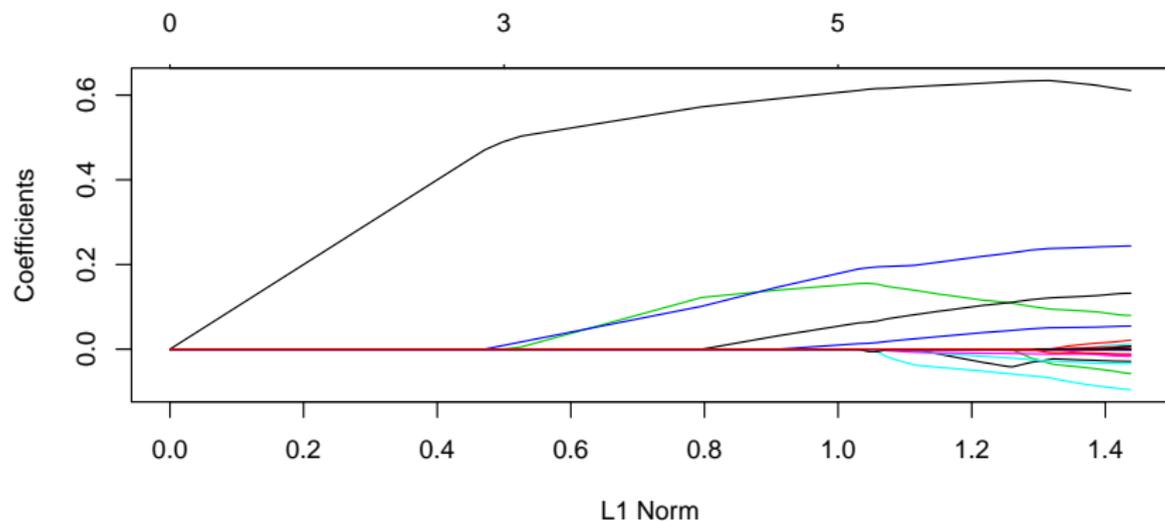


Figure 7: How coefficient estimates vary with shrinkage parameter

Using cross validation to select shrinkage parameter

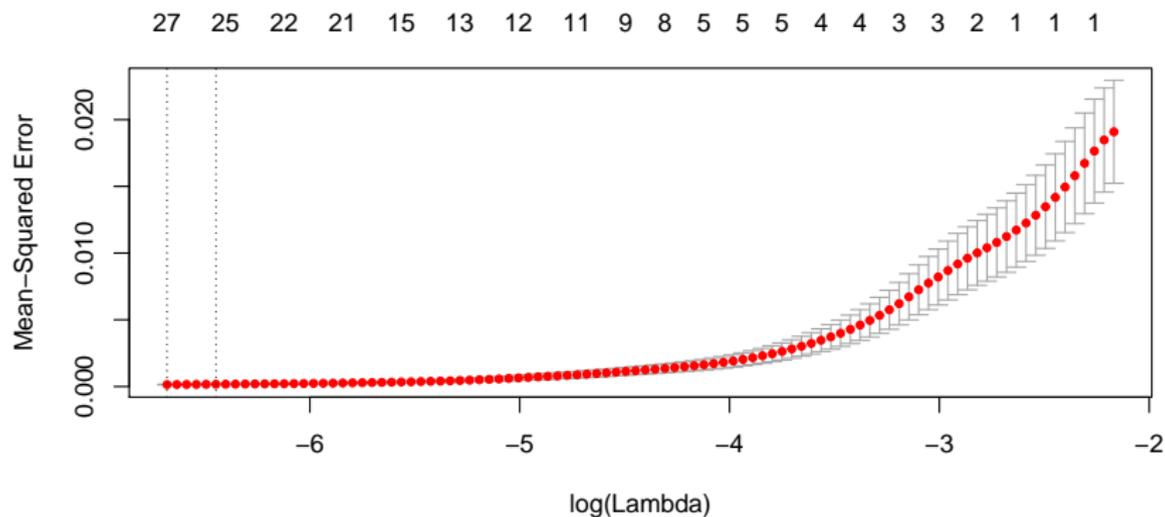


Figure 8:LASSO Coefficient Shrinking

Selected variables: Soft thresholding

Selected indicators

NA

DC.Employment..Information

DC.State.Unemployment.

DC.Employment..Transport.Utilities

DC.Dividends..Interest.and

DC.Hourly.Earnings..Trade.Transport.Utilities

DC.Other.Labor.Income

DC.Hourly.Earnings..Leisure.and.Hospitality

DC.Transfers.less.State.U

DC.Weekly.Earnings..Private.Service

DC.Earnings..Administrati

DC.Weekly.Earnings..Professional.Business

DC.Earnings..Finance.and

DC.Net.Earnings.by.Place.of.Residence

DC.Earnings..Professional

DC.Earnings.by.Place.of.Work

DC.Earnings..Retail.Trade

Index of indicators selected by soft thresholding

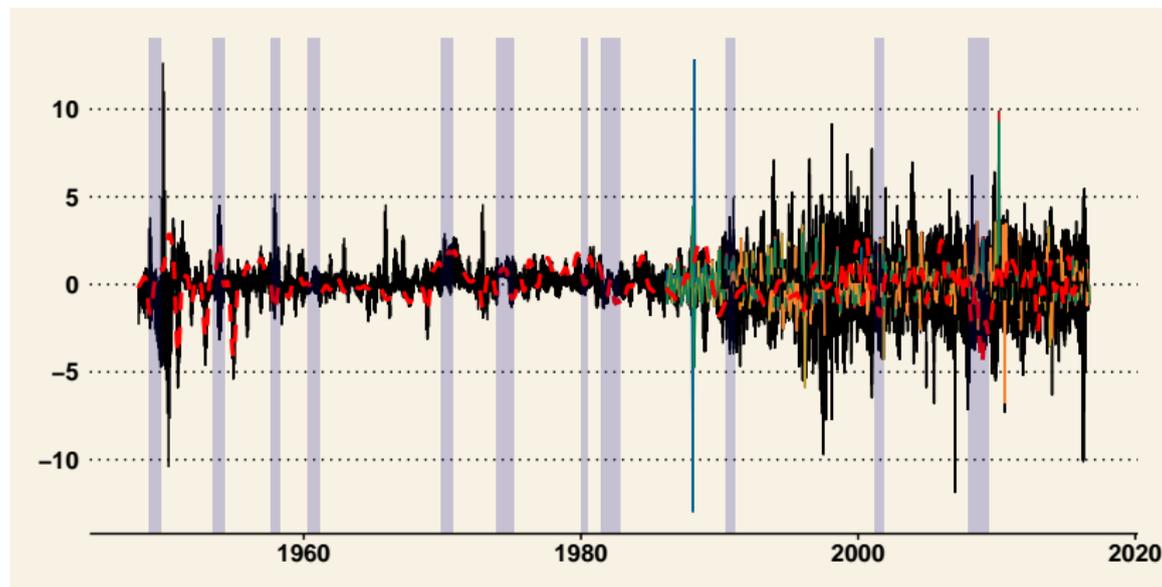


Figure 9: Index constructed from selected indicators (soft threshold)

Index (soft threshold) compared to DC personal income

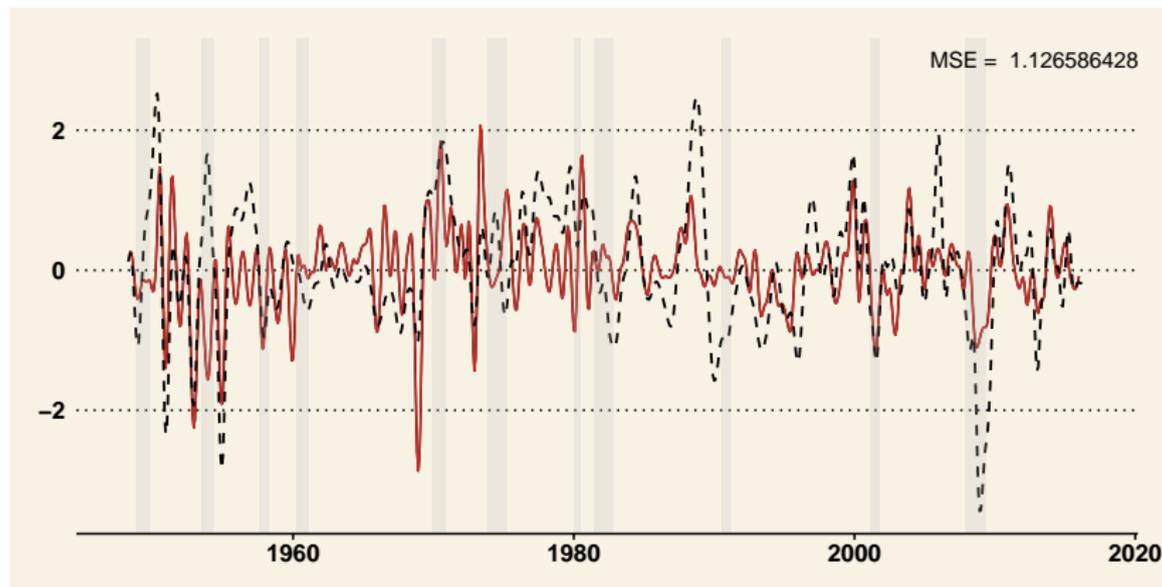


Figure 10: Change in Index vs Change in DC personal income

Index compared to DC personal income in levels

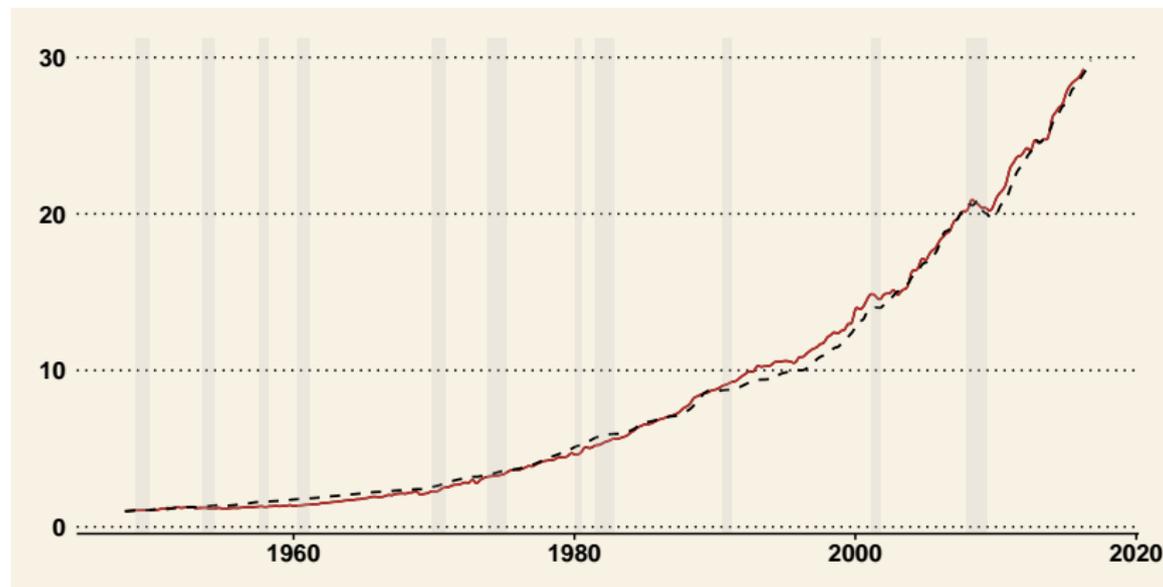


Figure 11: Index vs DC personal income in levels

Closing thoughts and next steps

- ▶ Compare to using all the available indicators selecting a subset of variables improved fit of the index to personal income
- ▶ “Soft thresholding” improved the fit compared to “hard thresholding”
- ▶ Not really that surprising since here we are measuring the in sample performance
- ▶ Next steps
 - ▶ Evaluate the out of sample performance
 - ▶ Build leading index