Big Data Analytics in Economics: What Have We Learned so Far, and Where Should We Go From Here?

Norman R. Swanson and Weiqi Xiong*

June 2017

*Rutgers University

prep. for Workshop on Forecasting at Deutsche Bundesbank, Sept., 2017

Introduction

- Availability of big data at many frequencies, for many variables is a key driving force for applied and theoretical work.
- Methodological and empirical advances have accumulated very quickly in recent years.
- I will discuss a very few of the advances in forecasting due in large part to this phenomenon – model building and model selection methods.

Introduction

- I . Model Building:

- Discuss - Factor Models and Diffusion Indices
  - Principal component analysis
  - Sparse principal component analysis
  - Independent component analysis

- Mention - Mixed Frequency (MF) Indices
  - Hybrid models using MF and diffusion indices
  - Modeling with switching and surveys

Introduction

- Discuss - Machine Learning, Variable Selection, and Shrinkage
  - Bagging
  - Boosting
  - Ridge regression
  - Least angle regression
  - Lasso
  - Elastic net
  - Non-negative garrote

  - Hybrid factor models using above methods

Introduction

- II. Model Selection:

- Loss Function Dependent Tests
  - Pairwise Comparison
  - Data Snooping or Multiple Comparison

- Robust Forecast Comparison
  - Stochastic Dominance Methods
  - Robust to Choice of Loss Function

I. Factor Models and Diffusion Indices

$$X_t = \mu + \Lambda F_t + \xi_t,$$

X_t an $N \times 1$ vector
Λ an $N \times r$ factor loading matrix
μ_0 an $N \times 1$ intercept

F_t is unobserved $r \times 1$ factor vector
ξ_t an error term

$$y_{t+h} = W_t \beta_W + F_t \beta_F + \varepsilon_{t+h}$$

- Above model includes W_t variables; autoregressive structure – key additional variables.

- Allow for random walk, AR, and VAR strawman models.

- Factor model is an approximation. Underlying model may not have a factor structure, but complex and rich covariance structure (e.g. in MC studies) across the X variables lends itself to principal component type shrinkage.

- What about mixed frequency models? Estimation of diffusion indices?

Factor Models and Diffusion Indexes

- What about usefulness of sparseness (SPCA – discussed later) and zero restrictions in factor loadings?

  e.g. Ability to isolate potential control variables for policy analysis. Interpretability remains an issue.

- Armah and Swanson (2010): Factor "proxy" selection →small set of observables as predictors. Parsimonious model selection?

- Key predictors = "variable subset"? Targeted predictors (e.g. Bai and Ng (2007,2008))?

- In Carrasco and Rossi (2016) → factors chosen using cross validation … explicitly considers "target variable". What about also selecting factor loadings based on target variable? i.e. three layers here:

- (i) Traditional approach of using highest eigenvalue factors.

- (ii) Select factors other than highest eigenval. ones, given target variable.

- (iii) Use (ii) and also determine "adjusted" loadings = shrinkage = lasso …

Factor Models and Diffusion Indexes

- Might *lack of sparseness* be of interest?

→ Variables that are *not usually* relevant included, and if these variables "jump" under structural change, then may impose robustness to structural instability
→ Turning point stability of predictions ...

- But sparseness useful to isolate potential control variable … interpretability.

- Again leads to methodology of Bai and Ng, i.e., targeted predictors.

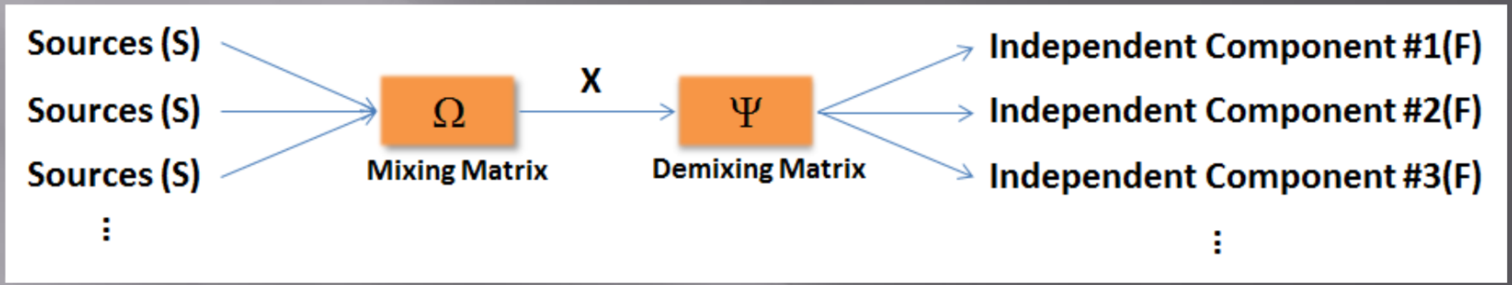
- What about: Couple shrinkage regression approach with factor/loadings shrinkage methods, such as sparse PCA, and include also a set of W targeted "stability predictors", say, or a factor constructed using these stability predictors.

- Kim and Swanson (2014,2016) →SPCA then shrinkage, or shrinkage followed by ICA, SPCA or PCA dimension reduction
  = lasso, elastic net -> get targeted predictors … then construct factors …

- Or directly "shrink" factors to a particular target …

Factor Models and Diffusion Indexes

- Independent Component Analysis
    - Assume the *F* are statistically independent
- .



- As is evident from above figure, ICA exactly the same as PCA, if demixing matrix is the factor loading coefficient matrix associated with PCA.

- In general, ICA yields uncorrelated factors with descending variance => easy "ordering".

- Moreover, those components explaining the largest share of the variance are often assumed to be the "relevant" ones for subsequent use in diffusion index forecasting.

Factor Models and Diffusion Indexes

- For simplicity, consider two observables, $X = (X_1, X_2)$.
- PCA transforms X into uncorrelated components $F = (F_1, F_2)$.
- Joint pdf characterized by $E(F_1 F_2) = E(F_1)E(F_2)$.
- ICA finds a demixing matrix which transforms the observed X into independent components $F^* = (F_1^*, F_2^*)$.
- Joint pdf characterized by $E[F_1^{*p} F_2^{*q}] = E[F_1^{*p}]E[F_2^{*q}]$.

Mention: Mixed Frequency Indexes

- Use multiple frequencies of data?

  → Pastcasting, nowcasting, forecasting, and "continuous" updating.

- Example: Factor MIDAS used for predicting quarterly data via the use of monthly factors (Marcellino and Schumacher (2010)).

- MIDAS model for forecasting h_q quarters ahead is

- . $$Y_{t_q+h_q} = \beta_0 + \beta_1 B(L^{1/m},\theta)\hat{F}^{(3)}_{t_m} + \varepsilon_{t_q}$$

- . $$B(L^{1/m},\theta) = \sum_{j=0}^{j^{\max}} b(j,\theta)L^{j/m}$$

- . $$b(j,\theta) = \frac{\exp\left(\theta_1 j+\theta_2 j^2\right)}{\sum_{j=0}^{j^{max}} \exp\left(\theta_1 j+\theta_2 j^2\right)}$$ -- Almon distributed lag

$\theta = (\theta_1,\theta_2)$ $\hat{F}_{t_m}$ is a set of monthly factors $\hat{F}^{(3)}_{t_m}$ is skip sampled from the monthly factor, $\hat{F}_{t_m}$

Machine Learning, Variable Selection, and Shrinkage

Sparseness not present in ridge regression, but may be useful for interpretation of factors. Key idea is to be able to (uniquely) estimate regression coefficients when number of variables > sample size.

- Optimization Problems that treat such multicollinearity.

Ridge (Hoerl): $\min \|y - X\beta\|^2 \quad \text{s.t. } \|\beta\|_2 = \sum_{j=1}^{p} \beta_j^2 \le \alpha.$

Lasso (Tibshirani): $\min \|y - X\beta\|^2 \quad \text{s.t. } \|\beta\|_1 = \sum_{j=1}^{p} |\beta_j| \le \alpha.$

Elastic Net (Zou, Hastie): $\min \|y - X\beta\|^2 \quad \text{s.t. } \|\beta\|_2 = \sum_{j=1}^{p} |\beta_j| \le \alpha_1 \text{ and } \sum_{j=1}^{p} \beta_j^2 \le \alpha_2.$

- Ridge the original → but lasso (least absolute shrinkage and selection operator) shrinks some parameters all the way to zero.

- Elastic net (Zou and Hastie (2005)) combines the two.

- If do not care about sparsity, how about neural nets as an alternative? Overfitting matters – how big an issue in factor analysis w/o sparseness, in the sense of PEER?

- Circling Back → Consider SPCA (Zou, Hastie and Tibshirani (2006)), which adds the sparseness feature of lasso (elastic net) to PCA.

How? Reformulate PCA as a regression-type optimization problem, and then impose the lasso (elastic net = double shrinkage) .

Consider penalized regression form of the optimization problems outlined above.

$$\widehat{\beta}_{lasso} = \arg\min_{\beta}\left\{\|y - \Sigma_{j=1}^{N} X_j \beta_j\|^2 + \lambda_1 \Sigma_{j=1}^{N} |\beta_j|\right\}.$$

$$\widehat{\beta}_{elastic\ net} = (1+\lambda_2)\arg\min_{\beta}\left\{\|y - \Sigma_{j=1}^{N} X_j \beta_j\|^2 + \lambda_1 \Sigma_{j=1}^{N} |\beta_j| + \lambda_2 \Sigma_{j=1}^{N} \beta_j^2\right\}.$$

- 2-stage SPCA? Replace y with F -> ridge is PCA then add *L1-norm* penalty.

- Constraint: Lasso can select at most T of N variables, when N>T in PCA construction.

- Economic interpretability of factors. Couple SPCA for factors with further targeted (on predictor variable) penalized regression?

L1 Versus L2 Regularization

Recalling that the L1 norm does not necessarily lead to sparsity, but the L1 regularization term (the penalty) on the weights/coefficients in the model does.

L2-norm (e.g. least squares regression)	L1-norm (e.g. LAD regression)
Not so robust to outliers	Robust
Stable solution	Unstable solution for small data perturbations
Unique solution	Possibly multiple solutions
Non-sparsity	Sparsity
Computational efficiency (anal. soln)	Comput. inefficiency (what if non-sparse?)

Model Selection (loss function dependent)

- Diebold and Mariano (1995), White (2000), Chao, Corradi and Swanson (2001), Clark and McCracken (2001,2013), Corradi and Swanson (2006) …

- Key Question: Should We Utilize Loss Function Specific Measures, or Not?

- .

$$H_0 : E(g(u_{0,t+h}) - g(u_{1,t+h})) = 0$$

$$H_A : E(g(u_{0,t+h}) - g(u_{1,t+h})) \neq 0$$

- ***Pairwise Accuracy*** $$DM_P = \frac{\bar{d}_t}{\hat{\sigma}_{\bar{d}_t}} \overset{d}{\rightarrow} N(0,1),$$

$$\bar{d}_t = \frac{1}{P}\sum_{t=R+1}^{T} d_t,\ d_t = g(\hat{u}_{0,t+h}) - g(\hat{u}_{1,t+h}), \text{ and } \hat{\sigma}_{\bar{d}_t} = \frac{\hat{\sigma}_{d_t}}{\sqrt{P}}.$$

- ***Causality*** $$m_P = P^{-1/2}\sum_{t=R+1}^{T} \hat{u}_{0,t+h} X_t$$

- ***Big Data*** $$S_P = \max_{k=1,\ldots,m} DM_P(1,k)$$

Model Selection (non loss function dependent)

Linton, Maasoumi and Whang (2005), Jin, Corradi, Swanson (2017)

Stochastic Dominance Methods

- *General Loss Forecast Superiority <-> 1st Order Stochastic Dominance*

$$u_1 \succeq_G u_2 \; iff \; E(L(u_1)) \le E(L(u_2)), \; \forall L \in \mathcal{L}_G$$

- *Convex Loss Forecast Superiority <-> 2nd Order Stochastic Dominance*

$$u_1 \succeq_C u_2 \; iff \; E(L(u_1)) \le E(L(u_2)), \; \forall L \in \mathcal{L}_C$$

- *Implementation:* $E(L(u_1)) \le E(L(u_2))$ *for all* L *iff* $G(x) \le 0,$

$$G(x) = (F_2(x) - F_1(x)) sgn(x)$$

$$C(x) = \int_{-\infty}^{x} (F_1(t) - F_2(t)) dt 1(x < 0) + \int_{x}^{\infty} (F_2(t) - F_1(t)) dt 1(x \ge 0)$$

Model Selection (non loss function dependent)

Linton, Maasoumi and Whang (2005), Jin, Corradi, Swanson (2017)

- $F_k(x) = P(u_{k,t} \leq x)$

$$\overline{F}_{k,n}(x) = P^{-1}\sum_{t=R}^{T} 1(u_{k,t} \leq x)$$

$$TG_n^{+} = \max_{k=2,..,m} \sup_{x\in\mathcal{X}^{+}} \sqrt{n}\, G_{k,n}(x) and TG_n^{-} = \max_{k=2,..,m} \sup_{x\in\mathcal{X}^{-}} \sqrt{n}\, G_{k,n}(x)$$

$$TC_n^{+} = \max_{k=2,..,m} \sup_{x\in\mathcal{X}^{+}} \sqrt{n}\, C_{k,n}(x) \text{ and } TC_n^{-} = \max_{k=2,..,m} \sup_{x\in\mathcal{X}^{-}} \sqrt{n}\, C_{k,n}(x),$$

$$G_{k,n}(x) = (\overline{F}_{k,n}(x) - \overline{F}_{1,n}(x))sgn(x)$$

$$C_{k,n}(x) = \left\{\int_{-\infty}^{x} (\overline{F}_{1,n}(s) - \overline{F}_{k,n}(s))ds1(x < 0) + \int_{x}^{\infty} (\overline{F}_{k,n}(s) - \overline{F}_{1,n}(s))ds1(x \geq 0)\right\}$$

Empirical Illustration: Forecasting the Term Structure

- ***The Models***

- *Dynamic Nelson Siegel – a 'small data' model (Diebold and Li (2006)) with time decay parameter, maturity parameter; and level, slope and curvature 'factors' (i.e., the betas), so that factor loading on level factor is one, etc.*

- *Slope factor increase -> slope of curve increases as short rates increases more than long rates in this case …*

$$\mathbf{y}_t(\tau) = \beta_{1,t} + \beta_{2,t}\left[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau}\right] + \beta_{3,t}\left[\frac{1-\exp(-\lambda_t\tau)}{\lambda_t\tau} - \exp(-\lambda_t\tau)\right] + \varepsilon_t$$

- *Dimension Reduction (Big Data) Models*

$$\mathbf{y}_{t+h}(\tau) = \beta' W_t + \alpha_2' F_t^b + \alpha_2' F_t^s + \varepsilon_{t+h}$$

- *Strawman Econometric Models*

$$\mathbf{y}_{t+h}(\tau) = \beta' W_t + \varepsilon_{t+h}$$

Empirical Illustration: Forecasting the Term Structure

- Use zero coupon U.S. Treasury yield curve, monthly, 1982-2016;
  Gurkaynak, Sack, and Wright ((2006).

- Target variables are 1,2,3,5,10 year maturity yields.

- Forecast horizons are h=1,3,12 months.

- Prediction subsamples 1992-99, ‘2000-07, 2008-16, recession/expansion.

- Small data panel has N=10, T=415.

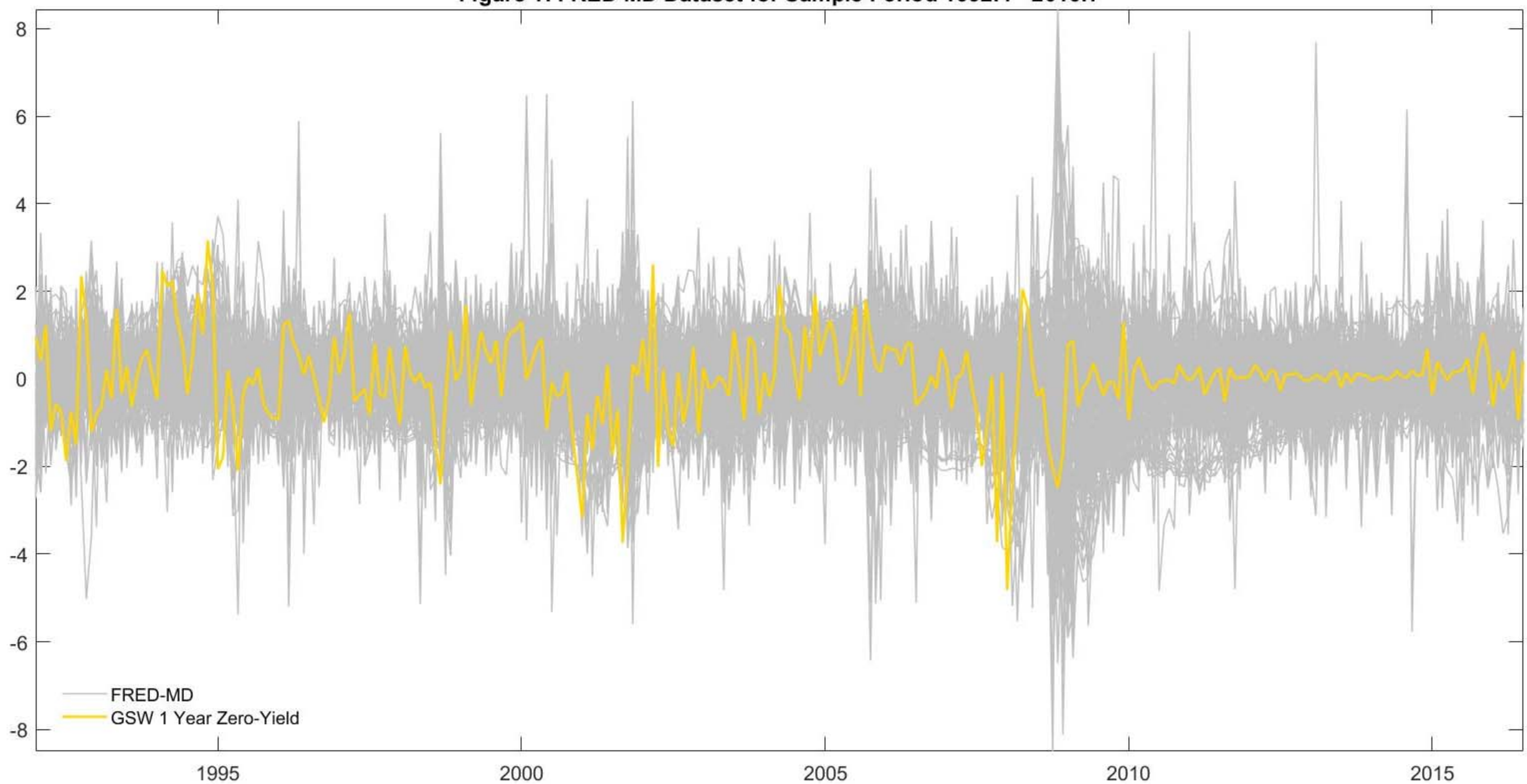
- Big data panel uses FRED-MD dataset with 103 macroeconomic variables.

- Predictions constructed in real-time, and estimations are based on rolling windows.

- Model Selection: MSFE and DM Tests.

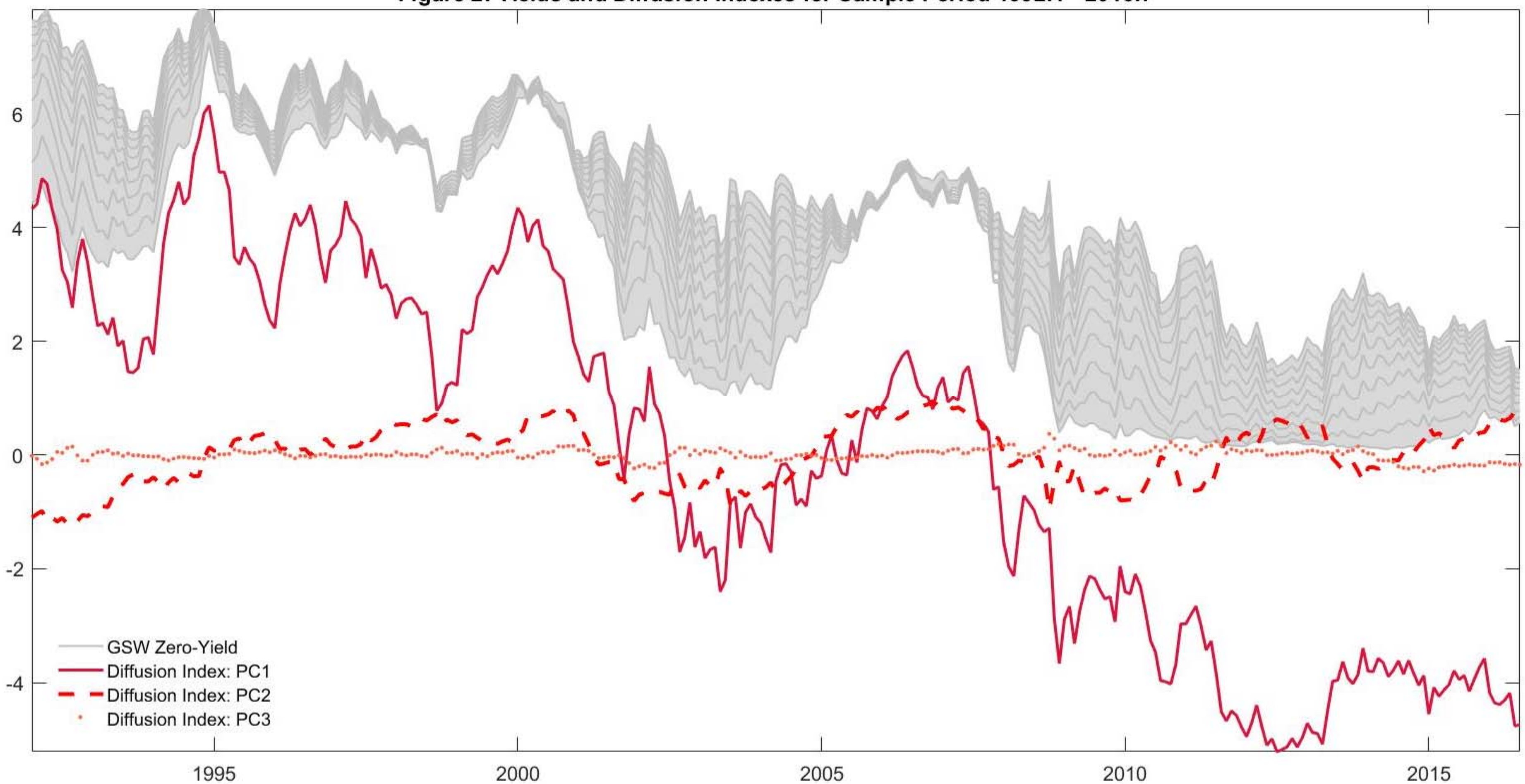
Empirical Illustration: Forecasting the Term Structure

Figure 1: FRED MD Dataset for Sample Period 1992:1 - 2016:7



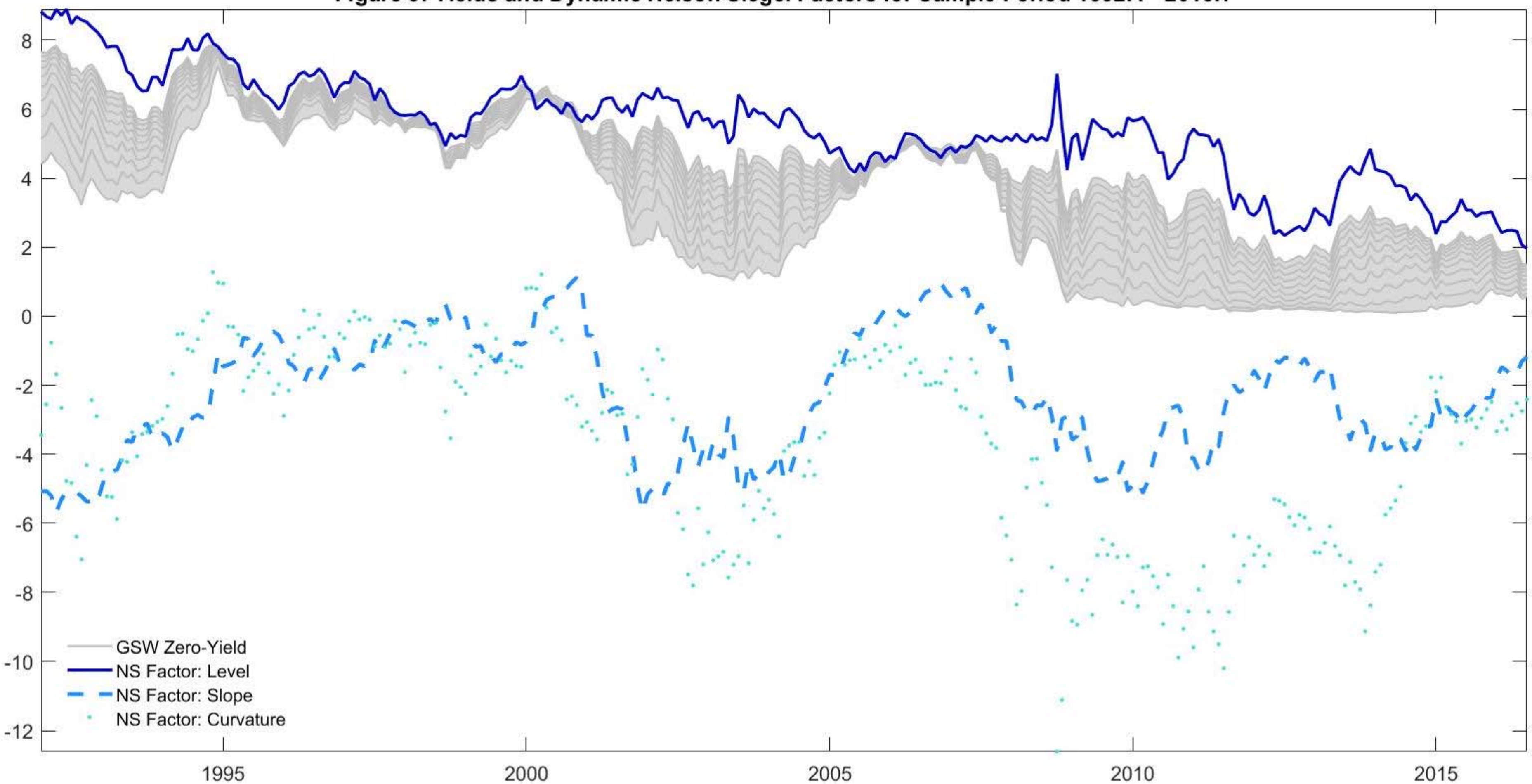
Empirical Illustration: Forecasting the Term Structure

Figure 2: Yields and Diffusion Indexes for Sample Period 1992:1 - 2016:7

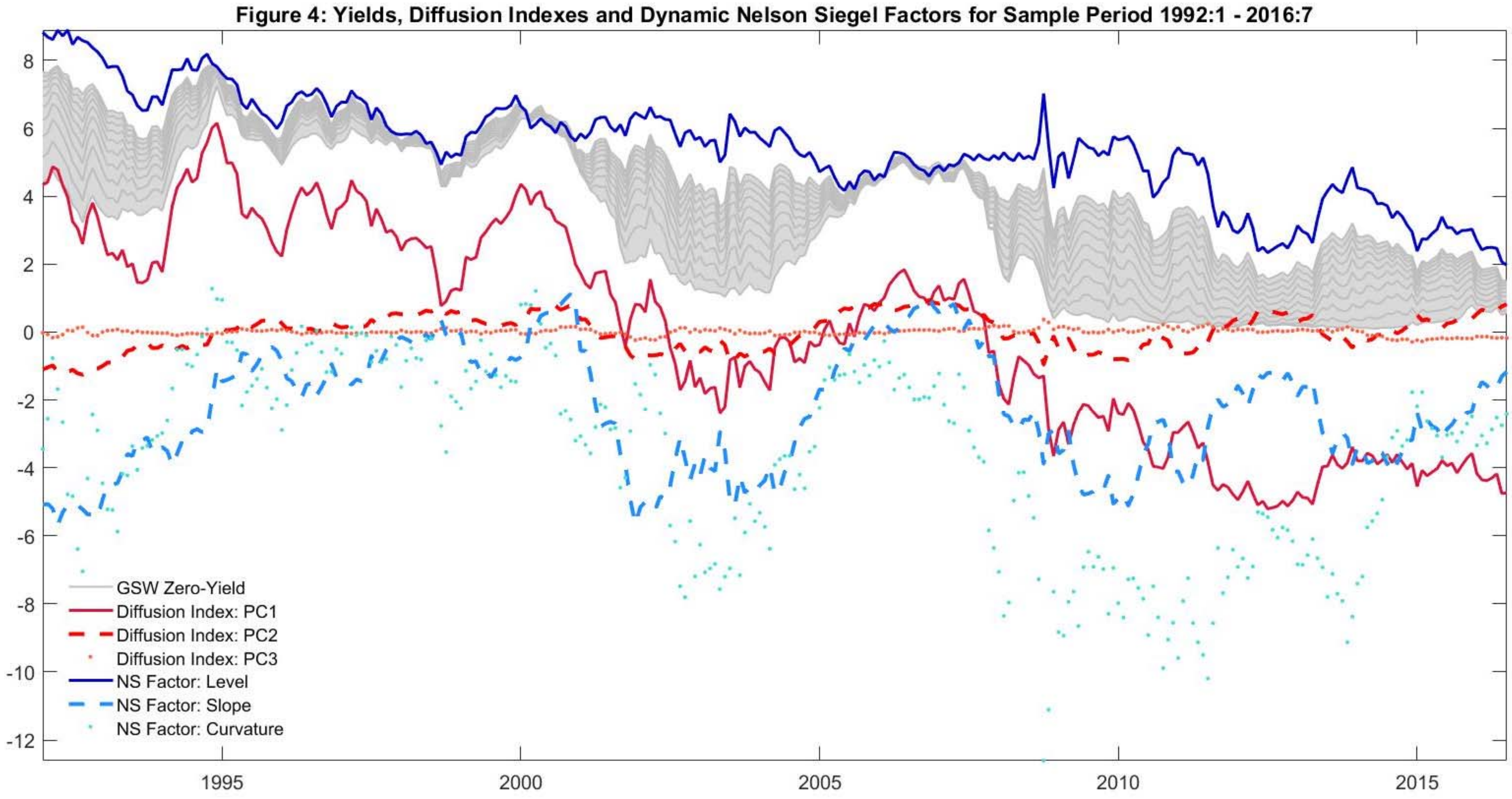


Empirical Illustration: Forecasting the Term Structure

Figure 3: Yields and Dynamic Nelson Siegel Factors for Sample Period 1992:1 - 2016:7



Empirical Illustration: Forecasting the Term Structure



Empirical Illustration: Forecasting the Term Structure

- **The Models**

AR(1)	Autoregressive model with one lag
VAR(1)	Five-dimensional vector autoregressive model with one lag
VAR(1)+FB1	VAR(1) model with one principle component added, principle component anal
VAR(1)+FB2	VAR(1) model with two principle components added, principle component ana
AR(SIC)	Autoregressive model with lag(s) selected by the Schwarz information criterion
VAR(SIC)	Five-dimensional vector autoregressive model with lag(s) selected by the Schw
VAR(SIC)+FB1	VAR(SIC) model with one principle component added, principle component a
VAR(SIC)+FB2	VAR(SIC) model with two principle components added, principle component

Empirical Illustration: Forecasting the Term Structure

DNS(1)	Dynamic Nelson-Siegel (DNS) model with underlying AR(1) factor specific
DNS(2)	DNS model with underlying AR(1) factor specifications fitted with six-dim
DNS(3)	DNS model with underlying AR(1) factor specifications fitted with four-di
DNS(4)	DNS model with underlying VAR(1) factor specifications fitted with ten-di
DNS(5)	DNS model with underlying VAR(1) factor specifications fitted with six-di
DNS(6)	DNS model with underlying VAR(1) factor specifications fitted with four-d
DNS(1)+FB1	DNS(1) model with one principle component added, principle component a
DNS(2)+FB1	DNS(2) model with one principle component added, principle component a
DNS(3)+FB1	DNS(3) model with one principle component added, principle component a
DNS(4)+FB1	DNS(4) model with one principle component added, principle component a
DNS(5)+FB1	DNS(5) model with one principle component added, principle component a
DNS(6)+FB1	DNS(6) model with one principle component added, principle component a
DNS(1)+FB2	DNS(1) model with two principle components added, principle component
DNS(2)+FB2	DNS(2) model with two principle components added, principle component
DNS(3)+FB2	DNS(3) model with two principle components added, principle component
DNS(4)+FB2	DNS(4) model with two principle components added, principle component
DNS(5)+FB2	DNS(5) model with two principle components added, principle component
DNS(6)+FB2	DNS(6) model with two principle components added, principle component

Empirical Illustration: Forecasting the Term Structure

DNS(1)+MAC	DNS(1) model with three key macroeconomic variables added: manufacturing cap
DNS(2)+MAC	DNS(2) model with three key macroeconomic variables added: manufacturing cap
DNS(3)+MAC	DNS(3) model with three key macroeconomic variables added: manufacturing cap
DNS(4)+MAC	DNS(4) model with three key macroeconomic variables added: manufacturing cap
DNS(5)+MAC	DNS(5) model with three key macroeconomic variables added: manufacturing cap
DNS(6)+MAC	DNS(6) model with three key macroeconomic variables added: manufacturing cap
DIF(1)	Diffusion index model with one principle component estimator based on all ten-di
DIF(2)	Diffusion index model with two principle component estimators based on all ten-d
DIF(3)	Diffusion index model with three principle component estimators based on all ten-
DIF(4)	Diffusion index model with one principle component estimator based on all 103 m
DIF(5)	Diffusion index model with two principle component estimators based on all 103 n
DIF(6)	Diffusion index model with three principle component estimators based on all 103
DIF(1)+FB1	DIF(1) model with one principle component added, principle component analysis
DIF(2)+FB1	DIF(2) model with one principle component added, principle component analysis
DIF(3)+FB1	DIF(3) model with one principle component added, principle component analysis
DIF(1)+FB2	DIF(1) model with two principle components added, principle component analysis
DIF(2)+FB2	DIF(2) model with two principle components added, principle component analysis
DIF(3)+FB2	DIF(3) model with two principle components added, principle component analysis

Empirical Illustration: Forecasting the Term Structure

- **MSFE-Best Models**

Forecast Sample	Maturity / Horizon	3 Months	1 Year	3 Years	5 Years	10 Years
1992:1-1999:12 Subsample 1'	1 Step	**DNS(6)+FB2**	**DNS(5)+FB2**	**DNS(5)+FB2**	**DIF(1)**	**AR(SIC)**
		DNS(6)+FB1	DNS(4)+FB2	DNS(4)+FB2	DNS(4)+FB2	DNS(2)+FB1
		DNS(4)+FB1	DNS(5)+FB1	DNS(6)+FB2	DNS(4)+FB1	DNS(1)+FB1
	3 Step	**DNS(4)+FB1**	**DNS(1)+FB1**	**DNS(1)+FB1**	**DNS(6)+FB1**	**AR(SIC)**
		DNS(5)+FB1	DNS(3)+FB1	DNS(3)+FB1	DNS(5)+FB1	DNS(6)+FB1
		DNS(6)+FB1	DNS(2)+FB1	DNS(2)+FB1	DNS(4)+FB1	DNS(5)+FB1
	12 Step	**DNS(3)**	**DNS(3)**	**DNS(3)**	**DNS(3)**	**DNS(1)**
		DNS(1)	DNS(1)	DNS(1)	DNS(1)	DNS(2)
		DNS(2)	DNS(2)	DNS(2)	DNS(2)	AR(SIC)

Empirical Illustration: Forecasting the Term Structure

	1 Step	**DNS(6)+FB2**	**DNS(2)+FB2**	**DNS(3)+FB2**	**DNS(1)+FB2**	**DNS(2)+FB2**
		DNS(3)+FB2	DNS(1)+FB2	DNS(1)+FB2	DNS(2)+FB2	DNS(1)+FB2
		DNS(5)+FB2	DNS(3)+FB2	DNS(2)+FB2	DNS(3)+FB2	DNS(3)+FB2
2000:1-2007:12 'Subsample 2'	3 Step	**DNS(3)+FB1**	**DNS(3)+FB1**	**DNS(3)+FB1**	**DNS(1)+FB1**	**DNS(2)+FB1**
		DNS(2)+FB1	DNS(1)+FB1	DNS(1)+FB1	DNS(2)+FB1	DNS(1)+FB1
		DNS(1)+FB1	DNS(2)+FB1	DNS(2)+FB1	DNS(3)+FB1	DNS(3)+FB1
	12 Step	**DNS(3)+FB1**	**VAR(1)**	**VAR(1)**	**VAR(1)**	**AR(SIC)**
		DNS(2)+FB1	VAR(SIC)	VAR(SIC)	VAR(SIC)	VAR(SIC)
		DNS(1)+FB1	AR(SIC)	AR(SIC)	VAR(SIC)+FB1	VAR(1)

Empirical Illustration: Forecasting the Term Structure

	1 Step	**AR(1)**	**AR(1)**	**AR(1)**	**AR(SIC)**	**DNS(2)**
		DIF(3)	DNS(3)	AR(SIC)	AR(1)	DNS(1)
		AR(SIC)	AR(SIC)	DNS(3)+MAC	DIF(1)+FB2	DNS(2)+MAC
2008:1-2016:7	3 Step	**DNS(6)+FB1**	**DNS(5)+FB1**	**AR(SIC)**	**AR(SIC)**	**DNS(5)+MAC**
'Subsample 3'		VAR(1)+FB2	DNS(4)+FB1	VAR(1)+FB2	VAR(1)+FB2	DNS(5)
		VAR(SIC)+FB2	DNS(5)+FB2	VAR(SIC)+FB2	VAR(SIC)+FB2	DNS(4)+MAC
	12 Step	**AR(SIC)**	**DNS(6)+MAC**	**AR(SIC)**	**VAR(1)**	**DNS(5)**
		VAR(SIC)	DNS(6)	DNS(6)+MAC	VAR(SIC)	DNS(4)
		VAR(1)	DNS(4)+MAC	DNS(6)	AR(SIC)	DNS(5)+MAC

Empirical Illustration: Forecasting the Term Structure

	1 Step	**DNS(6)+FB1**	**DNS(5)+FB1**	**DNS(4)+FB1**	**AR(1)**	**DNS(2)**
		VAR(SIC)+FB1	DNS(5)+FB2	DNS(5)+FB1	AR(SIC)	DNS(1)
		VAR(1)+FB1	DNS(4)+FB1	DNS(6)+FB1	VAR(SIC)+FB1	DNS(2)+MAC
1992:1-2016:7 'Subsample 4'	3 Step	**DNS(6)+FB1**	**DNS(5)+FB1**	**DNS(6)+FB1**	**AR(SIC)**	**DNS(5)+FB1**
		DNS(5)+FB1	DNS(4)+FB1	DNS(4)+FB1	DNS(4)+FB1	AR(SIC)
		DNS(4)+FB1	DNS(6)+FB1	DNS(5)+FB1	DNS(6)+FB1	DNS(4)+FB1
	12 Step	**AR(SIC)**	**AR(SIC)**	**AR(SIC)**	**AR(SIC)**	**AR(SIC)**
		VAR(1)	VAR(1)	VAR(1)	VAR(1)	DNS(5)
		VAR(SIC)	VAR(SIC)	VAR(SIC)	VAR(SIC)	DNS(4)

Empirical Illustration: Forecasting the Term Structure

		DNS(4)+FB1	**VAR(SIC)+FB1**	**VAR(1)**	**DNS(3)+FB2**	**DIF(2)+FB2**
	1 Step	DNS(5)+FB1	VAR(1)+FB1	VAR(SIC)	DNS(2)+FB2	DNS(3)
		VAR(SIC)+FB1	DNS(2)+MAC	DNS(1)+MAC	VAR(SIC)	DNS(2)
		DNS(6)+FB1	**DNS(6)+FB1**	**DNS(6)+FB1**	**DNS(2)+FB1**	**DNS(3)+FB1**
Recession	3 Step	DNS(6)+MAC	DNS(4)+FB1	DNS(4)+FB1	DNS(3)+FB1	DNS(2)
		DNS(1)+FB1	DNS(6)+MAC	DNS(5)+FB1	DNS(1)+FB1	DNS(1)
		DNS(3)+FB1	**DNS(3)+FB1**	**DNS(3)+FB1**	**DNS(1)+FB1**	**VAR(1)**
	12 Step	DNS(2)+FB1	DNS(1)+FB1	DNS(1)+FB1	DNS(2)+FB1	VAR(SIC)
		DNS(1)+FB1	DNS(2)+FB1	DNS(2)+FB1	DNS(3)+FB1	DNS(5)

Empirical Illustration: Forecasting the Term Structure

		DNS(6)+FB2	**DNS(5)+FB2**	**DNS(6)+FB2**	**AR(1)**	**DNS(2)+FB2**
	1 Step	DNS(6)+FB1	DNS(4)+FB2	DNS(4)+FB2	AR(SIC)	DNS(1)+FB2
		VAR(1)+FB1	DNS(3)+FB2	DNS(5)+FB2	DIF(1)	DNS(2)+FB1
		DNS(5)+FB1	**DNS(5)+FB1**	**AR(SIC)**	**AR(SIC)**	**DNS(5)+FB1**
Expansion	3 Step	DNS(6)+FB1	DNS(4)+FB1	DNS(5)+FB1	DNS(4)+FB1	DNS(4)+FB1
		DNS(4)+FB1	AR(SIC)	DNS(4)+FB1	DNS(6)+FB1	AR(SIC)
		DNS(4)+MAC	**AR(SIC)**	**AR(SIC)**	**AR(SIC)**	**AR(SIC)**
	12 Step	DNS(5)+MAC	DNS(5)+MAC	DNS(5)+MAC	VAR(1)+FB1	DNS(6)
		DNS(6)+MAC	DNS(4)+MAC	DNS(6)+MAC	VAR(SIC)+FB1	DNS(4)

Empirical Illustration: Forecasting the Term Structure

- **Forecast Combination (1-Month Ahead)**

Model	De
All	Average of all forty four forecast models
FB	Average of twenty five models that contain principle component(s)
FS	Average of nineteen non-FB type models
Econometrics	Average of all eight AR and VAR type models
DNS	Average of all twenty two DNS type models
DI	Average of twelve diffusion index type models

	All	0.911	0.971	0.984	1.061	1.042
	FB	0.958	1.011	1.011	1.074	1.082
1992:1-2016:7	FS	1.002	1.022	1.025	1.094	1.022
'Subsample 4'	Econometrics	**0.839*** **	**0.922***	**0.947**	**0.980**	1.053
	DNS	0.922	0.932*	0.951	1.114	**0.991**
	DIF	1.257	1.286	1.215	1.127	1.329

Empirical Illustration: Forecasting the Term Structure

- Subsamples 1 and 2: DNS+FB models usually win, including 17 of 20 maturity/horizon permutations.

- Subsample 3: DNS+FB wins in only 2 of 10 cases for h=1 and 3, across maturities. Post Great Recession confusion?

- Entire Sample Period: For h=1,3, DNS+FB wins 7 of 10 times.

- Evidence for h=12 much more mixed, AR, VAR, and pure DNS often ‘win.’

- 1 or 2 factors always ‘best’.

Empirical Illustration: Forecasting the Term Structure

- AND the `best’ models are almost always significantly better than AR(1) straw-man model.

- DNS model ‘winners’ are used ‘vector’ variety. DNS factors do not evolve independently of one another, when predicting.

- Thus, DNS factors best predicted using other DNS factors AND big data diffusion indexes.

- DNS+FB evidence even stronger for recession subsample: DNS+FB wins in 13 of 15 horizon/maturity permutations.

- NOT so for expansion subsample: DNS+FB wins in 7 of 15 permutaitons.

Empirical Illustration: Forecasting the Term Structure

- Forecast Combination is not optimal in our experiments.

- Combinations fail to win in 15 of 20 permutations, for h=1, across all subsamples.

- Combinations fail to win in 18 of 20 permutations, for h=3, across all subsamples and bond maturities.

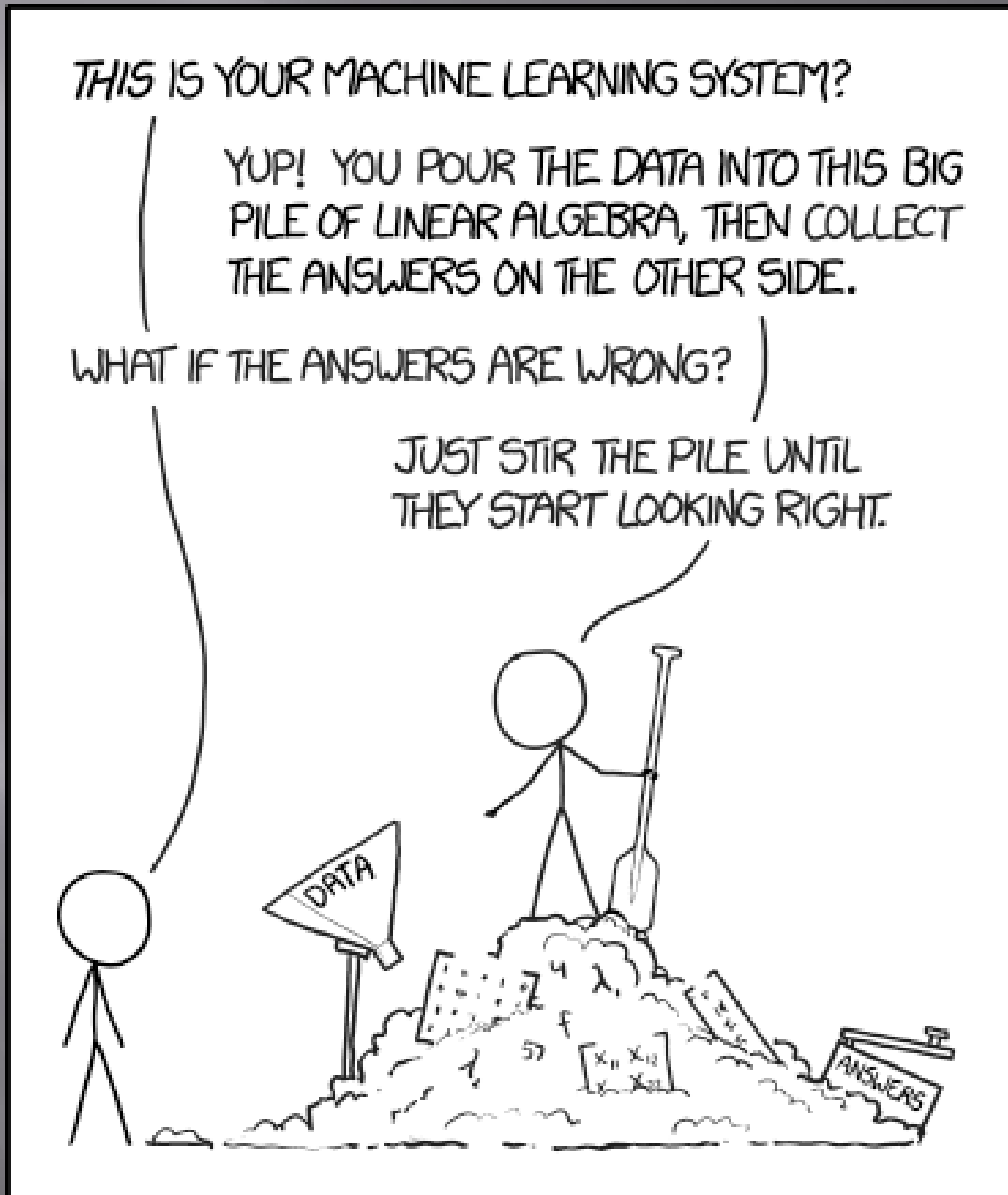
- Combinations fail to win in 17 of 20 permutations, for h=12, across all subsamples and bond maturities.

Concluding Remarks

- Big (and wide) data analysis is a burgeoning area of research, and many interesting methodological advances remain to discovered and also empirically analyzed.

- Not only do we have more data than ever to propel this empirical research, but we also have many useful new tools, ranging from data shrinkage methods to varieties of latent factor modelling, with which to work.

- Thank You!!!!


THIS IS YOUR MACHINE LEARNING SYSTEM?
YUP! YOU POUR THE DATA INTO THIS BIG PILE OF LINEAR ALGEBRA, THEN COLLECT THE ANSWERS ON THE OTHER SIDE.
WHAT IF THE ANSWERS ARE WRONG?
JUST STIR THE PILE UNTIL THEY START LOOKING RIGHT.
DATA
ANSWERS