Mining Big Data Using Parsimonious Factor and Shrinkage Methods

Hyun Hak Kim[1] and Norman Swanson[2]
[1]Bank of Korea and [2]Rutgers University

ECB Workshop on using Big Data for Forecasting and Statistics

Apr 7th, 2014

Outline

Introduction

- Statistical algorithms for large-scale data
  - Classification - decision trees, support vector machine
  - Clustering - K-means clustering, hierachical clustering
  - Regression - neural networks, principal component analysis
  - Sequence labeling - hidden markov models, kalman filters
  - Ensemble learning - bagging, boosting
- Overall, those are finding useful set of variables, that is, **reducing dimesion**

Original 50 components 25 components 2 components

Introduction

- Forecasting applications
  - Macroeconomic forecasting
    - Diffusion Index Model
  - Microeconomic forecasting
    - Consumer's behavior
    - Credit scoring
  - Non-economic forecasting
    - Server traffic analysis - high frequent possible
    - Social network analysis
    - Gene dependence analysis

- Our generic forecasting equation is:

$$Y_{t+h} = W_t \beta_W + F_t \beta_F + \varepsilon_{t+h}, \tag{1}$$

where h is the forecast horizon,

- W_t is a $1 \times s$ vector (possibly including lags of Y), and
- F_t is a $1 \times r$ vector of factors, extracted from F.
- The parameters, β_W and β_F are defined conformably, and
- ε_{t+h} is a disturbance term.

- Let X_{tj} be the observed datum for the $j-$th cross-sectional unit at time t, for $t = 1, ..., T$ and $j = 1, ..., N$. Recall that we shall consider the following model:

$$X_{tj} = \Lambda_j' F_t + e_{tj}, \tag{2}$$

where

- F_t is a $r \times 1$ vector of common factors,
- Λ_j is an $r \times 1$ vector of factor loadings associated with F_t, and
- e_{tj} is the idiosyncratic component of X_{tj}.
- The product $F_t \Lambda_j'$ is called the common component of X_{tj}.

- Shortcomings of Principal Component Analysis (PCA)
  - Since the factors by PCA are a linear combination of all variables,
  - It is not easy to interpret factors
- New approaches: parsimonious factor construction methods
  - Independent Component Analysis
  - Sparse Principal Component Analysis
- New factor construction methods
  - Parsimonious model improves predictive accuracy
  - Interpret factors

- Data set we examine in this paper
  - consists of 144 macro and financial monthly time series from 1960:01–2009:05
  - are what various papers including Stock and Watson (2002) used to investigate the usefulness of factor analysis in the context of forecasting
  - are used to construct factors fixed no matter what we forecast
  - are transformed accordingly to induce stationarity
- Categorize into 7 groups
  - Production, employment, housing, interest rate, price and the miscellaneous

Factor Construction Methods

Three types of factor construction methods are considered

1. Principal Component Analysis
2. Independent Component Analysis
3. Sparse Principal Component Analysis

Factor Construction Method I - PCA

Revisit Principal Component Analysis

- Consider the linear combinations

$$P_i = \mathbf{a}_i' \mathbf{X} = a_{i1} X_1 + a_{i2} X_2 + \cdots + a_{iN} X_N \tag{3}$$

- then we obtain

$$Var(P_i) = \mathbf{a}_i' \Sigma \mathbf{a}_i \qquad i = 1, 2, ..., N \tag{4}$$

$$Cov(P_i, P_k) = \mathbf{a}_i' \Sigma \mathbf{a}_k \qquad i \text{ and } k = 1, 2, ..., N \tag{5}$$

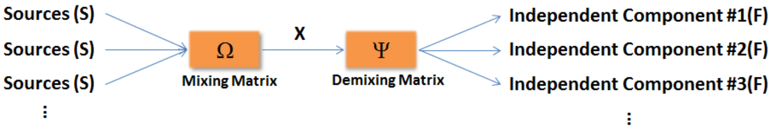
where Σ be the covariance matrix associated with the random vector $\mathbf{X} = [X_1, X_2, ..., X_N]$.

- The principal component are those uncorrelated linear combinations $P_1, P_2, ...P_N$ whose variance in (4) are as large as possible.

Factor Construction Method II - ICA

Introduction of Independent Component Analysis (ICA)

- The starting point for ICA is the very simple assumptions that the components, F, are statistically independent
- The key is the measurement of this *independence* between components
- It begins with statistical independent source data, S, which are mixed according to Ω ; and X which is observed, is a mixture of S weighted by Ω.



Schematic representation of ICA

- For simplicity, consider two observables, $X = (X_1, X_2)$. PCA finds uncorrelated components $F = (F_1, F_2)$, which have a joint probability density function, $p_F(F)$ with

$$E(F_1 F_2) = E(F_1) E(F_2). \tag{6}$$

- On the other hand, ICA finds independent components $F^* = (F_1^*, F_2^*)$, which have a joint pdf $p_{F^*}(F^*)$ with

$$E\left[F_1^{*p} F_2^{*q}\right] = E\left[F_1^{*p}\right] E\left[F_2^{*q}\right], \tag{7}$$

for every positive integer value of p and q.
- That is, independent components work for any moments.

- One often uses a modified version of entropy, so called negentropy, N,where:

$$N(F) = H(F_{gauss}) - H(F), \tag{8}$$

where F_{gauss} is a Gaussian random variable with the same covariance matrix as F.

- This negentropy, $N(\cdot)$, as a measure of nongaussianity, is zero for a Gaussian variable and always nonnegative.

- Simple version of this approximation use only one nonquadratic function, G, leading to:

$$N(F) \propto \left[E\{G(F)\} - E\{G(v)\}\right]^2 . \qquad (9)$$

- If we pick non-fast growing G, we may have more robust estimators. [Hyvärinen and Oja, 2000] suggest two Gs, and they show that these functions yield good approximations:

$$G_1(y) = \frac{1}{a_1} \log\cosh a_1 y \qquad (10)$$

and

$$G_2(y) = -\exp\left(-u^2/2\right), \qquad (11)$$

where $1 \leq a_1 \leq 2$ is some suitable constant.

Factor Construction Method III - SPCA

Introduction of Sparse Principal Component Analysis (SPCA)

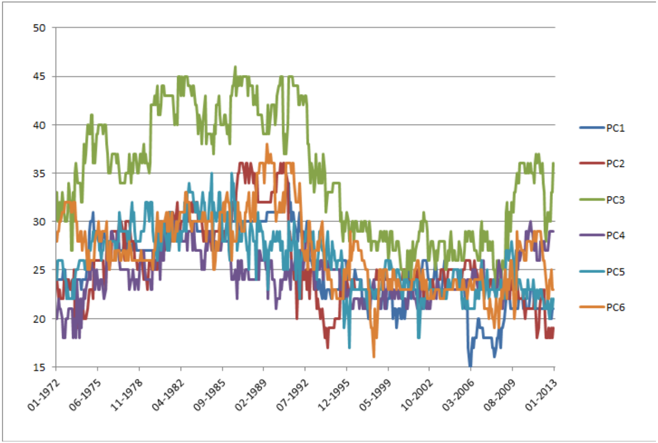
- Principal components are linear combinations of variables that are ordered by covariance contributions, and selection is of a small number of components which maximize the variance that is explained.
- However, factor loading coefficients are all typically nonzero, making interpretation of estimated components difficult.
- SPCA aids in the interpretation of principal components by placing (zero) restrictions on various factor loading coefficients.

Sparse Principal Component Analysis

Introduction

- [Zou et al., 2006] develop a regression optimization framework.
- Namely, they consider X as a dependent variables, F as explanatory variables, and the loadings as coefficients.
- They then use of the lasso (and elastic net) to derive a sparse loading matrix.
- Suppose we derive principle components (PCs), F via ordinary PCA.
- Then, let the estimated j-th principal component , F_j be the dependent variable and X be the independent variables.

Sparse Principal Component Analysis



Number of Nonzero Loadings in First 6 Sparse Components

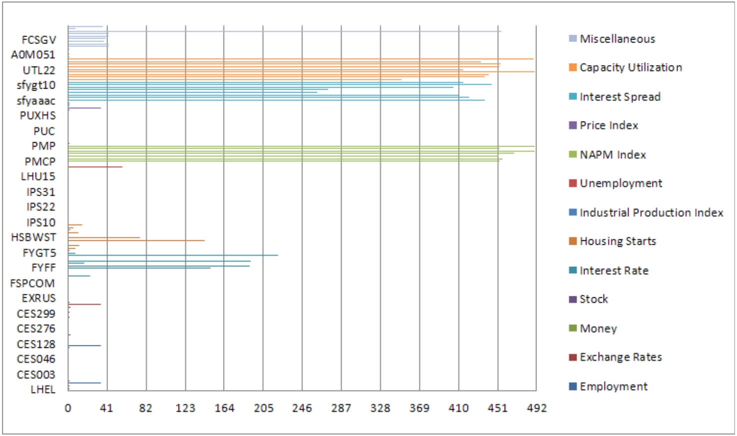
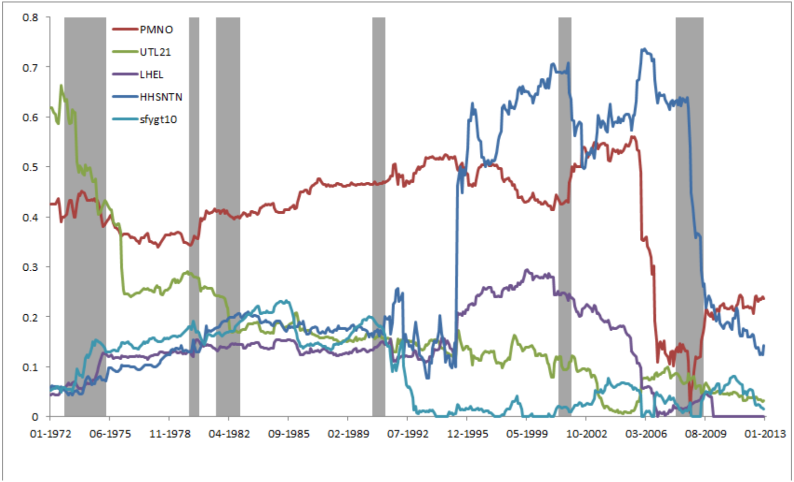


Figure: Number of Selection of Variables in First Sparse Principal Component

Sparse Principal Component Analysis



Time Varying Loadings of SPC

- Our generic forecasting equation is:

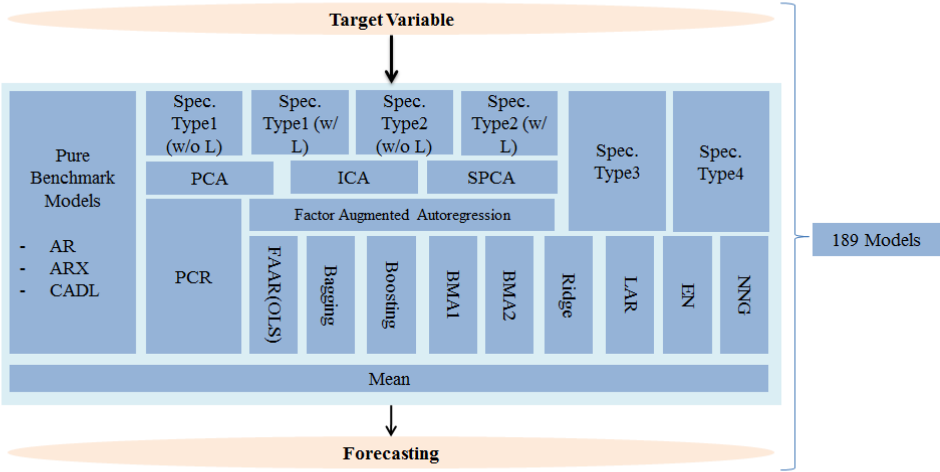
$$Y_{t+h} = W_t \beta_W + F_t \beta_F + \varepsilon_{t+h}, \tag{12}$$

where h is the forecast horizon,

- W_t is a $1 \times s$ vector (possibly including lags of Y), and
- F_t is a $1 \times r$ vector of factors, extracted from F.
- The parameters, β_W and β_F are defined conformably, and
- ε_{t+h} is a disturbance term.

- Forecasts of Y_{t+h} involve a two-step process
  1. The data X_t are used to estimate the factors, $\hat{F}_t$,
  2. Obtain the estimators $\hat{\beta}_F$ and $\hat{\beta}_W$ by regressing Y_{t+h} on $\hat{F}_t$ and W_t.
- As a first step, estimate $\hat{F}_t$ using
  - PCA, ICA and SPCA
- As a second step, estimate $\hat{\beta}_F$ with various robust estimation techniques including
  - Bagging, Boosting, Ridge Regression, Least Angle Regression, Elastic Net and Non-Negative Garotte
- Other than that, various benchmark models are considered including
  - Autoregressive model, CADL and BMA

Experimental Setup



Experimental Setup

- **Specification Type 1**: Various type of factor components are first constructed using large set of data; and then prediction models are formed using the shrinkage methods to select functions of and weights for the factors
- **Specification Type 2**: Various type of factor components are first constructed using subsets of variables from the large-scale dataset that are pre-selected via application of the robust shrinkage methods discussed.
- **Specification Type 3**: Prediction models are constructed using only the shrinkage methods, without use of factor analysis at any stage.
- **Specification Type 4**: Prediction models are constructed using only shrinkage methods, and only with variables which have nonzero coefficients, as specified via pre-selection using SPCA.

Table 4: Summary of MSFE-"Best" Models*

Panel A: Recursive Window Estimation

Forecast Horizon	Specification Method		UR	PI	TB10Y	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
h = 1	SP1	PCA	**FAAR**	PCR	Ridge	PCR	PCR	FAAR	ARX	PCR	**Mean**	Mean	**ARX**
		ICA	ARX	FAAR	FAAR	**FAAR**	**FAAR**	Ridge	ARX	FAAR	Mean	Boost	**ARX**
		SPCA	FAAR	**PCR**	**PCR**	BMA1	BMA2	**Mean**	**FAAR**	**FAAR**	Mean	**Boost**	**ARX**
	SP1L	PCA	**FAAR**	**PCR**	**Mean**	**PCR**	**Mean**	**Mean**	**ARX**	**BMA1**	**Mean**	**Boost**	**ARX**
		ICA	ARX	Mean	Mean	ARX	Mean	Mean	**ARX**	Mean	Mean	AR	**ARX**
		SPCA	ARX	Mean	CADL	ARX	Mean	Boost	**ARX**	Mean	Mean	Mean	**ARX**
	SP2	PCA	**Boost**	**Mean**	**Mean**	**Boost**	Mean	**Mean**	**ARX**	**BMA1**	BMA2	**Mean**	Boost
		ICA	ARX	Mean	Mean	ARX	**Mean**	Mean	**ARX**	ARX	EN	Mean	**Boost**
		SPCA	ARX	Mean	Mean	ARX	Mean	Mean	**ARX**	BMA1	**Boost**	Mean	Boost
	SP2L	PCA	**Boost**	**Mean**	**Mean**	Boost	Mean	**Mean**	**ARX**	**BMA1**	**BMA2**	**Mean**	Boost
		ICA	Boost	Mean	Mean	**Boost**	Mean	Mean	**ARX**	Boost	EN	Mean	**Boost**
		SPCA	Boost	Mean	Mean	ARX	**Mean**	Mean	**ARX**	ARX	Boost	Mean	Boost
	SP3		ARX	Mean	CADL	Mean	Mean	Mean	ARX	ARX	Mean	Boost	Mean
	SP4		ARX	Mean	Mean	ARX	Mean	Mean	ARX	BMA1	Mean	Mean	ARX

Panel C: Summary of Panel A and B

h=1

	Recursive Window Estimation							Rolling Window Estimation						
	SP1	SP1L	SP2	SP2L	SP3	SP4	Total	SP1	SP1L	SP2	SP2L	SP3	SP4	Total
AR	0	1	0	0	0	0	1	3	6	5	5	3	2	24
ARX	6	10	8	5	3	4	36	7	7	6	6	2	2	30
CADL	0	1	0	0	1	0	2	0	1	0	0	1	0	2
FAAR	10	1	0	0	0	0	11	4	0	0	0	0	0	4
PCR	6	2	0	0	0	0	8	2	1	0	0	0	0	3
Bagg	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Boost	2	2	6	10	1	0	21	0	0	3	2	3	2	10
BMA1	1	1	2	1	0	1	6	0	0	0	0	0	0	0
BMA2	1	0	1	1	0	0	3	0	0	0	1	0	1	2
Ridge	2	0	0	0	0	0	2	0	0	0	0	0	0	0
LAR	0	0	0	0	0	0	0	2	1	4	4	1	0	12
EN	0	0	1	1	0	0	2	0	0	3	2	1	0	6
NNG	0	0	0	0	0	0	0	1	0	0	0	0	0	1
Mean	5	15	15	15	6	6	62	14	17	12	13	0	4	60

Empirical Results

- Various our benchmarks do not dominate more complicated nonlinear methods, and that using a combination of factor and other shrinkage methods often yields superior predictions.

UR	PI	TB	CPI	PPI	NPE	HS	IPX	M2	SNP	GDP
SP1	SP1	SP1	SP4	SP1	SP1	SP1	SP1	SP1L	SP1	SP2
REC	REC	REC	ROL	REC	REC	REC	REC	ROL	REC	REC
PCA	SPC	SPC	N/A	ICA	SPC	SPC	SPC	SPC	SPC	ICA
FAAR	PCR	PCR	BMA2	FAAR	Mean	FAAR	FAAR	Mean	Boost	Boost

MSFE-best Specification Type/Window/PC/Model Combo for $h = 1$

Empirical Results

- Our benchmark econometric models are never found to be MSFE-best, regardless of the target variable being forecast, and the forecast horizon.

- Additionally, pure shrinkage type prediction models and standard (linear) regression models, do not MSFE-dominate models based on the use of factors constructed using either principal component analysis, independent component analysis or sparse component analysis.
  - This result provides strong new evidence of the usefulness of factor based forecasting.

- Recursive estimation window strategies only dominate rolling strategies at the 1-step ahead forecast horizon
- Including lags in factor model approaches does not generally yield improved predictions.

Concluding Remarks

- In this paper, I find
  - the simplest principal components type models "win" around 40% of the time.
  - Interestiingly, ICA and SPCA type models also "win" around 40% of the time.
  - hybrid methods including factor approaches coupled with shrinkage "win" around 1/3 of the time,
  - simple linear autoregressive type models never "win" in our experiments.
- I take these results as evidence of the usefulness of new methods in factor modelling and shrinkage, when the objective is prediction of macroeconomic time series variables

Hyvärinen, A. and Oja, E. (2000).
Independent component analysis: algorithms and applications.
Neural Networks, 13(4-5):411–430.

Zou, H., Hastie, T., and Tibshirani, R. (2006).
Sparse principal component analysis.
Journal of Computational and Graphical Statistics, 15(2):262–286.