M2 EEE Machine Learning

Final Project

Andrew Boomer and Jacob Pichelmann

April 16, 2021

Table of Contents

1 Introduction **2**

2 Data Generating Process **2**

3 Estimation Methods **3**

3.1 Via SIMPLS Algorithm 3

4 Evaluation Methods **4**

5 Empirical Application **5**

1 Introduction

Four dimension reduction devices: (1) Principal Component Analysis, (2) Ridge Regression, (3) Landweber Fridman (LF) regularization, (4) Partial least squares. Each involves a regularization or tuning parameter that is selected through genearlized cross validation (GCV).

Take two different data generating processes, (1) The eigenvalues of $\frac{X'X}{T}$ are bounded and decline to zero gradually. (2) Popular factor model with a finite number, r, of factors. Here, the r largest eigenvalues grow with N, while the remaining are bounded. In both cases, $\frac{X'X}{T}$ is ill-conditioned, which means the ratio of the largest to smallest eigenvalue diverges, and a regularization terms is needed to invert the matrix.

2 Data Generating Process

Large Sample Case: $N = 200$ and $T = 500$ Small Sample Case: $N = 100$ and $T = 50$

$$\underbrace{x_t}_{(N\times 1)} = \underbrace{\Lambda}_{(N\times r)} \underbrace{F_t}_{(r\times 1)} + \underbrace{\xi_t}_{(N\times 1)}$$

$$\underbrace{y_t}_{(1\times 1)} = \underbrace{\theta'}_{(1\times r)} \underbrace{F_t}_{(r\times 1)} + \underbrace{\nu_t}_{(1\times 1)}$$

$$\underbrace{y}_{(T\times 1)} = \underbrace{F}_{(T\times r)} \underbrace{\theta}_{(r\times 1)} + \underbrace{\nu}_{(T\times 1)}$$

$$\underbrace{X}_{(T\times N)} = \underbrace{F}_{(T\times r)} \underbrace{\Lambda'}_{(r\times N)} + \underbrace{\xi}_{(T\times N)}$$

DGP 1 (Few Factors Structure): θ is the $(r\times 1)$ vector of ones, $r = 4$ and $r_{max} = r+10$

DGP 2 (Many Factors Structure): θ is the $(r \times 1)$ vector of ones, $r = 50$ and $r_{max} = min(N, \frac{T}{2})$

DGP 3 (Five Factors but only One Relevant): $\theta = (1, 0_{1\times 4})$, $r = 5$ and $r_{max} = min(r + 10, min(N, \frac{T}{2}))$

$$F = [F_1, F_2']' \text{ and } F \times F' = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \end{bmatrix}$$

$y = \hat{F}\theta + \nu$ where $\hat{F}$ is generated from X equation in DGP 3, and $\sigma_\nu = 0.1$

DGP 4 (x_t Has a Factor Structure but Unrelated to y_t):

θ is a vector of zeros with dimension $(r \times 1)$. $r = 5$, $r_{max} = r + 10$. $F \times F'$ is defined as in DGP 3.

DGP 5 (Eigenvalues Declining Slowly):

θ is an $(N \times 1)$ vector of ones. $r = N$, $r_{max} = min(N, \frac{T}{2}$.

$\Lambda = M \odot \xi$, with $\xi \sim (N \times N)$ matrix of $iidN(0, 1)$

$$M \sim (N \times N) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \frac{1}{2} & \frac{1}{2} & \cdots & \frac{1}{2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} \end{bmatrix}$$

DGP 6 (Near Factor Model):
$\theta = 1$, $r = 1$, $r_{max} = r + 10$, $\Lambda' = \frac{1}{\sqrt{N}} 1_{r \times N}$
Estimation:
Set 1: Bai-Ng, PCA, PLS, Ridge, LF, LASSO Set 2: GCV, Mallows, AIC, BIC Set 3: Small Sample, Large Sample

Parameter Iteration (For Later) Simulation (Andy) Model Estimation (Jacob) Evaluation (Jacob) Output (Andy)

3 Estimation Methods

Ridge Estimator:

$$\widehat{y} = M_T^\alpha y = X(S_{xx} + \alpha I)^{-1} S_{xy}$$

LF Estimator:

$$\widehat{y} = M_T^\alpha y = X \sum_{j=1}^{\min(N,T)} \frac{\left(1 - \left(1 - d\widehat{\lambda}_j^2\right)^{1/\alpha}\right)}{\widehat{\lambda}_j^2} \left\langle y, \hat{\psi}_j \right\rangle_T \frac{X'\hat{\psi}_j}{T}$$

Spectral Cutoff/Principal Components Estimator:

$$\widehat{y} = M_T^\alpha y = \widehat{\Psi} \left(\widehat{\Psi}'\widehat{\Psi}\right)^{-1} \widehat{\Psi}' y$$

$$\text{Where } \widehat{\Psi} = \left[\widehat{\psi}_1 \left| \widehat{\psi}_2 \right| \ldots \mid \widehat{\psi}_k\right]$$

Partial Least Squares Estimator:

$$\widehat{y} = M_T^\alpha y = X V_k \left(V_k' X' X V_k\right)^{-1} V_k' X' y$$

$$\text{Where } V_k = \Big(X'y, \quad (X'X)\, X'y, \ldots, (X'X)^{k-1} X'y\Big)$$

3.1 Via SIMPLS Algorithm

$S = X^T y$

for $i \in 1 : k$

if $i = 1, [u, s, v] = svd(S)$

if $i > 1, [u, s, v] = svd(S - (P_k[:, i-1](P_k[:, i-1]^T P_k[:, i-1])^{-1} P_k[:, i-1]^T S))$

$T_k[:, i-1] = X R_k[:, i-1]$

$P_k[:, i-1] = \dfrac{X^T T_k[:, i-1]}{T_k[:, i-1]^T T_k[:, i-1]}$

$\widehat{y} = M_T^\alpha y = X R_k (T_k^T T_k)^{-1} T_k^T y$

4 Evaluation Methods

Generalized Cross Validation:

$$\hat{\alpha} = \arg \min_{\alpha \in A_T} \frac{T^{-1} \left\| y - M_T^{\alpha} y \right\|^2}{\left(1 - T^{-1} \operatorname{tr}\left(M_T^{\alpha}\right)\right)^2}$$

Mallows Criterion:

$$\hat{\alpha} = \arg \min_{\alpha \in A_T} T^{-1} \left\| y - M_T^{\alpha} y \right\|^2 + 2\widehat{\sigma}_{\varepsilon}^2 T^{-1} \operatorname{tr}\left(M_T^{\alpha}\right)$$

Where $\widehat{\sigma}_{\epsilon}^2$ is a consistent estimator of the variance of ϵ

So variance of ϵ is taken from the errors of the largest model, or from the model with all regressors for PC.

Leave-one-out Cross Validation:

$$\hat{\alpha} = \arg \min_{\alpha \in A_T} \frac{1}{T} \sum_{t=1}^{T} \left(\frac{y_i - \hat{y}_{i,\alpha}}{1 - M_T^{\alpha}[ii]} \right)^2$$

(GCV, $N = 100, T = 50$)

		PC	PLS	Ridge		LF	
	r	k	k	α	DOF	α	DOF
DGP 1	4.00	4.56	4.16	6.56	16.21	0.00	7.48
(s.e.)	–	(2.17)	(3.25)	(2.96)	(9.53)	(0.00)	(1.31)
DGP 2	50.00	20.24	4.76	0.56	47.49	0.00	31.85
(s.e.)	–	(5.55)	(6.32)	(0.46)	(2.00)	(0.00)	(2.20)
DGP 3	5.00	1.76	1.12	5.08	16.99	0.00	7.91
(s.e.)	–	(1.58)	(0.59)	(1.83)	(4.20)	(0.00)	(5.11)
DGP 4	5.00	1.76	1.84	8.48	14.52	0.03	3.52
(s.e.)	–	(3.18)	(2.88)	(3.01)	(7.83)	(0.01)	(6.64)
DGP 5	100.00	1.44	3.56	12.83	11.14	0.03	2.14
(s.e.)	–	(2.35)	(3.13)	(4.46)	(8.50)	(0.01)	(2.79)
DGP 6	1.00	2.44	1.16	4.85	16.69	0.28	2.58
(s.e.)	–	(2.55)	(0.78)	(3.27)	(7.89)	(0.45)	(1.00)

5 Empirical Application

We build on the long history of machine learning in forecasting macroeconomic variables[1] and use the Federal Reserve Bank's monthly database, which was established for empirical analysis that requires 'big data', hence constituting an ideal environment for the estimators discussed above. We took inspiration from the work of Coulombe et al. (2020) but limit ourselves to PC, Ridge, PLS and LF.[2] The dataset contains 134 monthly US macroeconomic and financial indicators observed from January 1959 to January 2021. An overview of all variables is given in the appendix. Following Coulombe et al. (2020) we predict three indicators which are of key economic interest, namely Industrial Production (INDPRO), Unemployment Rate (UNRATE), and the Consumer Price Index (INF). For each of these variables of interest Y_t we define the forecast objective as

$$y_{t+h} = (1/h)ln(Y_{t+h}/Y_t) \tag{1}$$

[1]See e.g. ...

[2]Unfortunately we were unable to use (1) Gu et al. (2020) as TSE does not have access to WDRS returns, (2) as only data on the resulting factors is available or (3) as they do not offer any replication data.