M2 EEE Machine Learning

Final Project

Andrew Boomer and Jacob Pichelmann

April 18, 2021

Table of Contents

1 Introduction

This report discusses and presents a replication of a selection of findings from Carrasco and Rossi as well as an empirical application of the methods covered. Carrasco and Rossi discusses in-sample prediction and out-of-sample forecasting in regressions with many exogenous predictors based on four dimension-reduction devices: principal components (PCA), ridge, Landweber Fridman (LF), and partial least squares (PLS). Each involves a regularization or tuning parameter that is selected through generalized cross validation (GCV) or Mallows Cp. Following Carrasco and Rossi we evaluate these estimators in a monte carlo simulation framework with 6 different data generating processes (DGP).

2 Factor Models in Economics

Factor models attempt to explain panels of data in terms of a smaller number of common factors that apply to each of the variables in the dataset. In the case of high dimensional data, factor models are a useful tool to reduce the dimensionality of the dataset, making estimation possible where the dataset would have been rank deficient before. A factor model on panel data can be represented as

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

where X denotes the matrix of observations, F the underlying factors and Λ the corresponding factor loadings. ξ is an idiosyncratic shock.

Additionally, through dimensionality reduction, factor models can find the most important variables that effect the outcome variables. For factor models in general, a crucial part of the estimation procedure is determining the number of factors to use. This is the context that Carrasco and Rossi is set in. The parameter used to select the number of factors in a factor model is also known as the regularization parameter. Carrasco and Rossi run simulations to analyze each of the different dimension reduction devices.

3 Data Generating Process

To study how accurate each estimation method is, Carrasco and Rossi simulate six different data generating processes, both in the large and small sample cases. In the large sample case, the size of the data set is $N = 200$ and $T = 500$. In the small sample case, the size is $N = 100$ and $T = 50$

- ▷ 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):

$\theta = 1$, $r = 1$, $r_{max} = r + 10$, $\Lambda' = \frac{1}{\sqrt{N}} 1_{r \times N}$

4 Estimation Methods

4.1 Notation

In matrix notation the model is

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix}, X = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_T' \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{pmatrix}$$

where y is a $(T \times 1)$ vector, X is a $(T \times N)$ matrix of predictors and ε is a $(T \times 1)$ vector. We then write

$$S_{xx} = \frac{X^T X}{T} \quad \text{and} \quad S_{xy} = \frac{X^T y}{T}$$

$$\widehat{y} = M_T^\alpha y = X \widehat{\delta}^\alpha$$

Additionally, the degrees of freedom (DOF) for each estimator is calculated as the trace of M_T^α. For PC and PLS, the trace of M_T^α is equal to the number of factors k. Moreover we denote as α the choice of penalty parameter which is obtained from one of the selection methods discussed in section 5.

For some of the estimators, we need to calculate the matrix of eigenvectors of X. Given that X is a non-square matrix in both the large and small sample cases, we decided to decompose X using the singular value decomposition (SVD). The SVD of X is represented as:

$$\underbrace{X}_{(T\times N)} = \underbrace{U}_{T\times T}\underbrace{\Sigma}_{T\times N}\underbrace{V}_{N\times N}$$

From this decomposition, we have that U is the $T \times T$ matrix of orthonormalized eigenvectors of $\frac{XX^T}{T}$, sorted in descending order of each vectors' eigenvalues. V^T is the matrix of orthonormalized eigenvectors of $\frac{X^TX}{T}$. Lastly, the diagonal of Σ is the vector of the square root of the eigenvalues of $\frac{X^TX}{T}$.

In the notation of the Carrasco and Rossi paper, $U = \hat{\psi}$ and $diag(\Sigma)^2 = \lambda^2$.

4.2 Estimators

Carrasco and Rossi specify multiple different expressions for each of their four estimation methods. We implemented each formulation of the method to check for accuracy and computational efficiency. We unfortunately found that for some of the estimation methods, the results were not consistent between the different formulations. Additionally, where possible, we vectorized the matrix sums to improve computation time. Lastly, no matter the formulation, for some of the estimators, we were not able to get the regularization parameters to agree exactly with the tables in the paper by Carrasco and Rossi.

For each estimator we choose the expression that is the most computationally efficient and whose optimized parameters are the closest to the tables presented by Carrasco and Rossi. The four estimation methods in the paper we replicated are:

4.2.1 Principal Components/Spectral Cutoff

For this estimator, the implementation we use is:

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

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

As explained in the Notation section, $\hat{\Psi}$ is estimated from the singular value decomposition of X. The number of vectors k included in $\hat{\Psi}$ is the regularization parameter that is optimized for the PC method. We decided to use this implementation as it was the most straightforward, agreed with the results from Carrasco and Rossi, and was fully vectorized.

4.2.2 Ridge Estimator

For the ridge estimator, the implementation we use is:

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

where I is the $(N \times N)$ identity matrix.

For this estimator, we were not able to get the estimated α parameter to agree with the simulation results from Carrasco and Rossi. We also tested the implementation of the ridge which involves the eigenvectors of $\frac{XX^T}{T}$:

$$M_T^\alpha y = \sum_{j=1}^{\min(N,T)} \frac{\hat{\lambda}_j^2}{\hat{\lambda}_j^2 + \alpha} \left\langle y, \hat{\psi}_j \right\rangle_T \hat{\psi}_j$$

However, this implementation yielded estimation results that were further away from the Carrasco and Rossi values, and took longer to implement, so we used the specification involving the regularized inverse of $S_x x$.

4.2.3 Landweber Fridman (LF) Estimator

For the LF estimator, we implemented:

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

Here d denotes XXXX. We follow Carrasco and Rossi and choose $d = 0.018/max(\lambda^2)$. XXXXX WHAT IS NOW LAMBDA SQ AGAIN XXXX.

With the LF estimator, the α values reported by Carrasco and Rossi were mostly indistiguishable from 0, so it was difficult to verify that we were implementing it correctly. However our values for the DOF of the LF estimator were not exactly as reported, but there was no other implementation to try.

4.2.4 Partial Least Squares (PLS) Estimator

For the PLS estimator, we initially implemented the specification:

$$\widehat{y} = M_T^\alpha y = XV_k \left(V_k' X' X V_k\right)^{-1} V_k' X' y$$
$$\text{where } V_k = \left(X'y, \quad \left(X'X\right) X'y, \ldots, \left(X'X\right)^{k-1} X'y\right)$$

With this implementation, we were not able to get our parameter estimates for k to agree with the results of Carrasco and Rossi. We looked at how PLS regressions were implemented in various programming softwares, and came across an implementation known as the SIMPLS algorithm by De Jong. This algorithm for solving PLS is as follows:

$S = X^T y$

for $i \in 1 : k$

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

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

$T_k[:, i-1] = XR_k[:, i-1]$

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

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

Utilising this algorithm both brought our estimated values of k closer to those of Carrasco and Rossi and was computationally faster, so this is the method we used.

5 Selection Methods

As outlined above the choice of regularization parameter is crucial. We hence implement selection on three criteria.

- ▷ Generalized Cross Validation (GCV):

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

- ▷ 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. In practice this translates to the variance of ϵ being taken from the errors of the largest model, or from the model with all regressors in the case of PCA.

- ▷ Leave-one-out Cross Validation (LOO-CV):

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

Note that for PC and PLS $\operatorname{tr}\left(M_T^{\alpha}\right) = k$, i.e. the number of factors. For the Ridge and LF estimators, the trace of M_t is equivalent to the degrees of freedom. Moreover, we only use LOO-CV for PLS.

6 Simulation Results

We run simulations of the six DGPs outlined in section 3 for a small sample ($N = 100, T = 50$) and a large sample ($N = 200, T = 500$). Due to computational limitations we only ran 25 simulations for each of the DGPs in the large sample.[1]

XXXX DISCUSSION OF FINDINGS XXXX

(GCV, $N = 200, T = 500$)

		PC	PLS	Ridge		LF	
	r	k	k	α	DOF	α	DOF
DGP 1	4.00	4.87	7.43	0.41	10.98	0.00	0.02
(s.e.)	–	(1.87)	(3.81)	(0.16)	(0.81)	(0.00)	(0.01)
DGP 2	50.00	50.93	109.38	1.97	93.09	0.00	0.09
(s.e.)	–	(2.48)	(18.26)	(0.06)	(0.94)	(0.00)	(0.00)
DGP 3	5.00	1.86	1.16	0.61	11.15	0.00	0.01
(s.e.)	–	(2.01)	(1.04)	(0.24)	(0.93)	(0.00)	(0.00)
DGP 4	5.00	0.88	1.26	18.12	3.87	0.04	0.02
(s.e.)	–	(2.09)	(1.58)	(5.11)	(1.54)	(0.04)	(0.02)
DGP 5	200.00	0.99	8.30	26.70	12.08	0.03	0.07
(s.e.)	–	(2.42)	(9.40)	(7.17)	(6.81)	(0.04)	(0.07)
DGP 6	1.00	6.54	1.01	2.36	5.28	0.00	0.40
(s.e.)	–	(3.28)	(0.13)	(4.44)	(2.04)	(0.00)	(0.00)

[1]For the small sample our personal machines provided sufficient computing power to execute 1000 simulations per DGP.

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

		PC	PLS	Ridge		LF	
	r	k	k	α	DOF	α	DOF
DGP 1	4.00	4.78	4.23	1.25	7.83	0.00	0.15
(s.e.)	–	(2.08)	(3.43)	(0.56)	(1.23)	(0.00)	(0.02)
DGP 2	50.00	19.50	3.31	0.99	23.37	0.00	0.51
(s.e.)	–	(6.06)	(5.23)	(0.02)	(0.14)	(0.00)	(0.07)
DGP 3	5.00	1.99	1.16	1.02	9.33	0.00	0.10
(s.e.)	–	(2.33)	(1.07)	(0.39)	(1.02)	(0.00)	(0.01)
DGP 4	5.00	1.18	1.34	8.70	5.00	0.02	0.21
(s.e.)	–	(2.61)	(1.70)	(2.95)	(1.73)	(0.02)	(0.21)
DGP 5	100.00	1.34	3.73	12.47	3.99	0.01	0.41
(s.e.)	–	(3.28)	(4.14)	(4.70)	(2.96)	(0.02)	(0.38)
DGP 6	1.00	3.32	1.17	5.85	2.59	0.10	0.13
(s.e.)	–	(3.15)	(0.91)	(3.99)	(1.95)	(0.08)	(0.25)

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

		PC	PLS	Ridge		LF	
	r	k	k	α	DOF	α	DOF
DGP 1	4.00	14.00	14.00	0.00	14.00	0.00	0.78
(s.e.)	–	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.03)
DGP 2	50.00	25.00	25.00	0.00	25.00	0.00	0.66
(s.e.)	–	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
DGP 3	5.00	15.00	15.00	0.00	15.00	0.00	0.51
(s.e.)	–	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)
DGP 4	5.00	15.00	15.00	0.00	15.00	0.00	0.51
(s.e.)	–	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)
DGP 5	100.00	25.00	25.00	0.00	25.00	0.00	0.87
(s.e.)	–	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.04)
DGP 6	1.00	11.00	11.00	0.00	11.00	0.00	1.00
(s.e.)	–	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

7 Empirical Application

7.1 Introduction and Data

Building on the long history of machine learning in forecasting macroeconomic variables[2] we use the Federal Reserve Bank's monthly database (FRED-MD) to apply the estimators discussed above on real data.[3] This database was established for empirical analysis that requires 'big data' and hence constitutes an ideal environment to employ the methods discussed above. We took inspiration from the work of Coulombe et al. (2020) but limit ourselves to PC, Ridge, PLS and LF. 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 housing starts (HOUST).[4] For each of these variables of interest Y_t we follow Coulombe et al (2020) in defining the forecast objective as

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

where h denotes the number of periods ahead. This allows us to assess the performance of our predictive methods for further periods ahead. Given the nature of the data we expect the underlying factor structure to be similar to DGP XXXXXXXXX

7.2 Evaluation

We evaluate the performance of our methods on the out of sample MSE. To be able to compute this metric we split our data into a training and a test set where the former spans all observations from ... to ... amounting to 80% of the data. Denoting N the number of observations in the test set we calculate the MSE as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \widehat{Y}_i)^2$$

where $\widehat{Y}_i = \widehat{\Psi}\widehat{\delta}_{pc}$ for PCA and $\widehat{Y}_i = X_{test}\widehat{\delta}_m, m \in \{R, LF, PLS\}$ for all other models.

We conduct forecasts for $h = \{1, 3, 9\}$ periods ahead. Given the simulation results we expect Ridge to deliver the best results, here defined as yielding the smallest out of sample MSE. Subsequently we report results similar to the simulation framework; we provide tables showing for each combination of estimator and parameter selection method the chosen penalty parameter/the number of factors as well as the degrees of freedom and the resulting out of sample MSE for $h = 1$. Moreover, we visualize the out of sample MSE to ease comparison across methods and settings.

[2]See e.g. ...

[3]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.

[4]Fortunately McCracken and Ng (2016), the accompanying paper of the dataset, outlines a transformation method for each variable to achieve stationarity. We apply those transformations in our data preparation.

7.3 Results and Discussion

From tables 1 to 3 we can immediately see that the estimated factor structure as well as the chosen penalty parameters are remarkably stable across both selection criteria and variables of interest. We cautiously take this as indication for the underlying macro data to indeed exhibit a stable factor structure, i.e. that economic variables are driven by a set of common underlying factors. The number of estimated factors ranges from 9 to 15 depending on the setting and choice of method, which is in line with the literature. In terms of forecasting power, we can see that LF yields the smallest out of sample MSE for $h = 1$ across all variables of interest. The difference to Ridge is, however, close to negligible.

Table 1: $Y_t = INDPRO$

Method	OOS MSE, $h = 1$	$alpha/k$	DOF
LF: GCV	0.000126	0.0001	14.532053
LF: Mallow	0.000126	0.0001	14.532053
PC: GCV	0.000162	13.0000	13.000000
PC: Mallow	0.000162	15.0000	15.000000
PLS: GCV	0.002139	15.0000	15.000000
PLS: Mallow	0.002139	15.0000	15.000000
Ridge: GCV	0.000130	0.1170	20.074885
Ridge: Mallow	0.000130	0.1170	20.074885

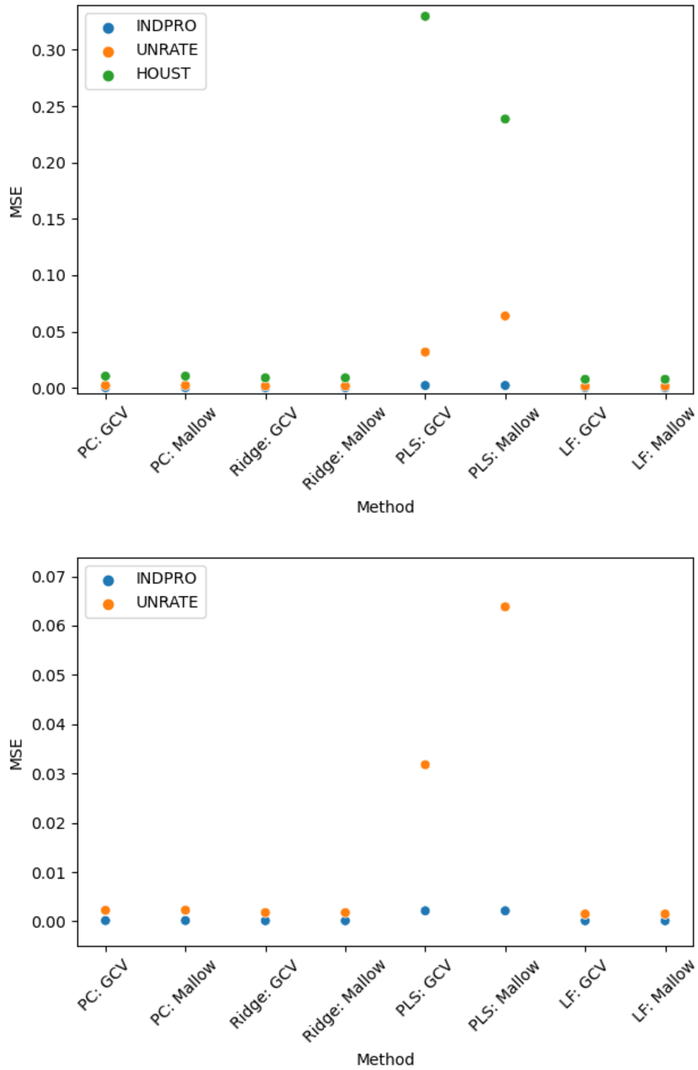
Table 2: $Y_t = UNRATE$

Method	OOS MSE, $h = 1$	$alpha/k$	DOF
LF: GCV	0.001488	0.0001	14.418940
LF: Mallow	0.001488	0.0001	14.418940
PC: GCV	0.002247	9.0000	9.000000
PC: Mallow	0.002276	15.0000	15.000000
PLS: GCV	0.031782	13.0000	13.000000
PLS: Mallow	0.063822	15.0000	15.000000
Ridge: GCV	0.001779	0.1170	20.008091
Ridge: Mallow	0.001779	0.1170	20.008091

Table 3: $Y_t = HOUST$

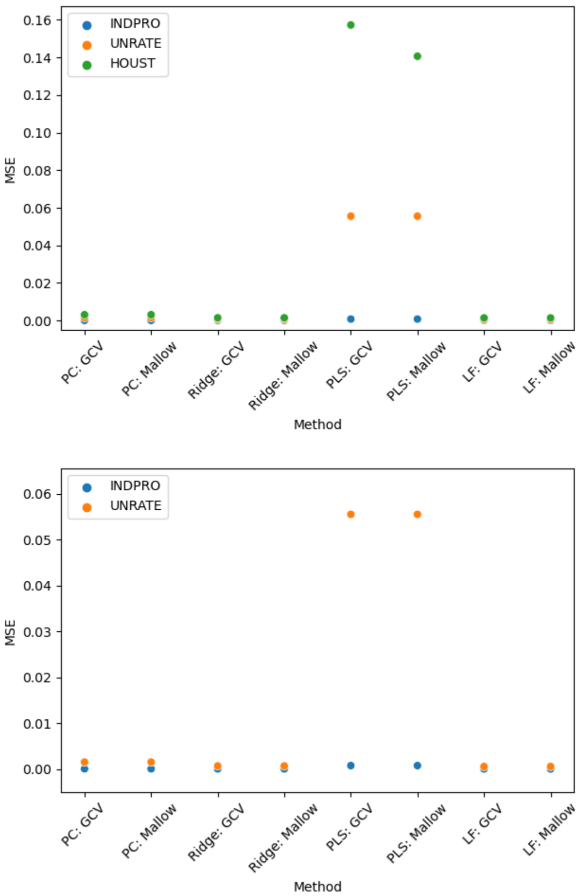
Method	OOS MSE, $h = 1$	$alpha/k$	DOF
LF: GCV	0.007627	0.0001	14.441119
LF: Mallow	0.007627	0.0001	14.441119
PC: GCV	0.010409	15.0000	15.000000
PC: Mallow	0.010409	15.0000	15.000000
PLS: GCV	0.329750	11.0000	11.000000
PLS: Mallow	0.238590	15.0000	15.000000
Ridge: GCV	0.008906	0.1170	20.052968
Ridge: Mallow	0.008906	0.1170	20.052968

Comparing the performance across variables we make the interesting finding that the PLS performs very poorly when predicting $HOUST$. We therefore report for each

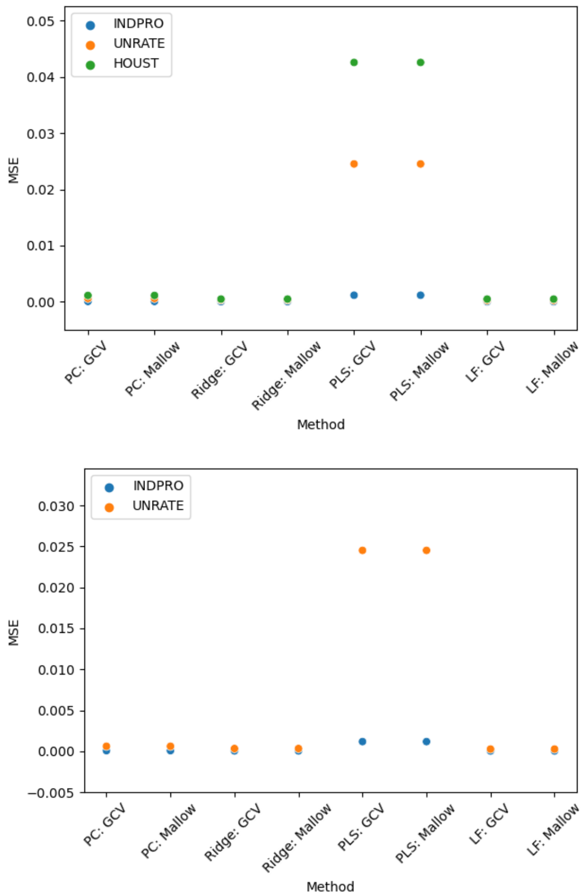


Out of sample MSE, $h = 1$

time horizon also the subset of only $INDPRO$ and $UNRATE$ to increase readability. Generally PLS yields the highest out of sample MSE while Ridge and LF perform best. This is somewhat surprising as the simulation results did not indicate a similar pattern XXX CHECK THIS XXX. In line with the findings of Coulombe et al (2020) we observe that the performance increases for further periods ahead.



Out of sample MSE, $h = 3$



Out of sample MSE, $h = 9$

References

[1] Jushan Bai and Serena Ng. "Determining the number of factors in approximate factor models". In: *Econometrica* 70.1 (2002), pp. 191–221.

[2] Marine Carrasco and Barbara Rossi. "In-sample inference and forecasting in misspecified factor models". In: *Journal of Business & Economic Statistics* 34.3 (2016), pp. 313–338.

[3] Sijmen De Jong. "SIMPLS: an alternative approach to partial least squares regression". In: *Chemometrics and intelligent laboratory systems* 18.3 (1993), pp. 251–263.

Appendix

Data Dictionary

Group 1: Output and income

	id	tcode	fred	description	gsi	gsi:description
1	1	5	RPI	Real Personal Income	M_14386177	PI
2	2	5	W875RX1	Real personal income ex transfer receipts	M_145256755	PI less transfers
3	6	5	INDPRO	IP Index	M_116460980	IP: total
4	7	5	IPFPNSS	IP: Final Products and Nonindustrial Supplies	M_116460981	IP: products
5	8	5	IPFINAL	IP: Final Products (Market Group)	M_116461268	IP: final prod
6	9	5	IPCONGD	IP: Consumer Goods	M_116460982	IP: cons gds
7	10	5	IPDCONGD	IP: Durable Consumer Goods	M_116460983	IP: cons dble
8	11	5	IPNCONGD	IP: Nondurable Consumer Goods	M_116460988	IP: cons nondble
9	12	5	IPBUSEQ	IP: Business Equipment	M_116460995	IP: bus eqpt
10	13	5	IPMAT	IP: Materials	M_116461002	IP: matls
11	14	5	IPDMAT	IP: Durable Materials	M_116461004	IP: dble matls
12	15	5	IPNMAT	IP: Nondurable Materials	M_116461008	IP: nondble matls
13	16	5	IPMANSICS	IP: Manufacturing (SIC)	M_116461013	IP: mfg
14	17	5	IPB51222s	IP: Residential Utilities	M_116461276	IP: res util
15	18	5	IPFUELS	IP: Fuels	M_116461275	IP: fuels
16	19	1	NAPMPI	ISM Manufacturing: Production Index	M_110157212	NAPM prodn
17	20	2	CUMFNS	Capacity Utilization: Manufacturing	M_116461602	Cap util

Group 2: Labour market

	id	tcode	fred	description	gsi	gsi:description
1	21*	2	HWI	Help-Wanted Index for United States		Help wanted indx
2	22*	2	HWIURATIO	Ratio of Help Wanted/No. Unemployed	M_110156531	Help wanted/unemp
3	23	5	CLF16OV	Civilian Labor Force	M_110156467	Emp CPS total
4	24	5	CE16OV	Civilian Employment	M_110156498	Emp CPS nonag
5	25	2	UNRATE	Civilian Unemployment Rate	M_110156541	U: all
6	26	2	UEMPMEAN	Average Duration of Unemployment (Weeks)	M_110156528	U: mean duration
7	27	5	UEMPLT5	Civilians Unemployed - Less Than 5 Weeks	M_110156527	U < 5 wks
8	28	5	UEMP5TO14	Civilians Unemployed for 5-14 Weeks	M_110156523	U 5-14 wks
9	29	5	UEMP15OV	Civilians Unemployed - 15 Weeks & Over	M_110156524	U 15+ wks
10	30	5	UEMP15T26	Civilians Unemployed for 15-26 Weeks	M_110156525	U 15-26 wks
11	31	5	UEMP27OV	Civilians Unemployed for 27 Weeks and Over	M_110156526	U 27+ wks
12	32*	5	CLAIMSx	Initial Claims	M_15186204	UI claims
13	33	5	PAYEMS	All Employees: Total nonfarm	M_123109146	Emp: total
14	34	5	USGOOD	All Employees: Goods-Producing Industries	M_123109172	Emp: gds prod
15	35	5	CES1021000001	All Employees: Mining and Logging: Mining	M_123109244	Emp: mining
16	36	5	USCONS	All Employees: Construction	M_123109331	Emp: const
17	37	5	MANEMP	All Employees: Manufacturing	M_123109542	Emp: mfg
18	38	5	DMANEMP	All Employees: Durable goods	M_123109573	Emp: dble gds
19	39	5	NDMANEMP	All Employees: Nondurable goods	M_123110741	Emp: nondbles
20	40	5	SRVPRD	All Employees: Service-Providing Industries	M_123109193	Emp: services
21	41	5	USTPU	All Employees: Trade, Transportation & Utilities	M_123111543	Emp: TTU
22	42	5	USWTRADE	All Employees: Wholesale Trade	M_123111563	Emp: wholesale
23	43	5	USTRADE	All Employees: Retail Trade	M_123111867	Emp: retail
24	44	5	USFIRE	All Employees: Financial Activities	M_123112777	Emp: FIRE
25	45	5	USGOVT	All Employees: Government	M_123114411	Emp: Govt
26	46	1	CES0600000007	Avg Weekly Hours : Goods-Producing	M_140687274	Avg hrs
27	47	2	AWOTMAN	Avg Weekly Overtime Hours : Manufacturing	M_123109554	Overtime: mfg
28	48	1	AWHMAN	Avg Weekly Hours : Manufacturing	M_14386098	Avg hrs: mfg
29	49	1	NAPMEI	ISM Manufacturing: Employment Index	M_110157206	NAPM empl
30	127	6	CES0600000008	Avg Hourly Earnings : Goods-Producing	M_123109182	AHE: goods
31	128	6	CES2000000008	Avg Hourly Earnings : Construction	M_123109341	AHE: const
32	129	6	CES3000000008	Avg Hourly Earnings : Manufacturing	M_123109552	AHE: mfg

Group 3: Housing

	id	tcode	fred	description	gsi	gsi:description
1	50	4	HOUST	Housing Starts: Total New Privately Owned	M_110155536	Starts: nonfarm
2	51	4	HOUSTNE	Housing Starts, Northeast	M_110155538	Starts: NE
3	52	4	HOUSTMW	Housing Starts, Midwest	M_110155537	Starts: MW
4	53	4	HOUSTS	Housing Starts, South	M_110155543	Starts: South
5	54	4	HOUSTW	Housing Starts, West	M_110155544	Starts: West
6	55	4	PERMIT	New Private Housing Permits (SAAR)	M_110155532	BP: total
7	56	4	PERMITNE	New Private Housing Permits, Northeast (SAAR)	M_110155531	BP: NE
8	57	4	PERMITMW	New Private Housing Permits, Midwest (SAAR)	M_110155530	BP: MW
9	58	4	PERMITS	New Private Housing Permits, South (SAAR)	M_110155533	BP: South
10	59	4	PERMITW	New Private Housing Permits, West (SAAR)	M_110155534	BP: West

Group 4: Consumption, orders and inventories

	id	tcode	fred	description	gsi	gsi:description
1	3	5	DPCERA3M086SBEA	Real personal consumption expenditures	M_123008274	Real Consumption
2	4*	5	CMRMTSPLx	Real Manu. and Trade Industries Sales	M_110156998	M&T sales
3	5*	5	RETAILx	Retail and Food Services Sales	M_130439509	Retail sales
4	60	1	NAPM	ISM : PMI Composite Index	M_110157208	PMI
5	61	1	NAPMNOI	ISM : New Orders Index	M_110157210	NAPM new ordrs
6	62	1	NAPMSDI	ISM : Supplier Deliveries Index	M_110157205	NAPM vendor del
7	63	1	NAPMII	ISM : Inventories Index	M_110157211	NAPM Invent
8	64	5	ACOGNO	New Orders for Consumer Goods	M_14385863	Orders: cons gds
9	65*	5	AMDMNOx	New Orders for Durable Goods	M_14386110	Orders: dble gds
10	66*	5	ANDENOx	New Orders for Nondefense Capital Goods	M_178554409	Orders: cap gds
11	67*	5	AMDMUOx	Unfilled Orders for Durable Goods	M_14385946	Unf orders: dble
12	68*	5	BUSINVx	Total Business Inventories	M_15192014	M&T invent
13	69*	2	ISRATIOx	Total Business: Inventories to Sales Ratio	M_15191529	M&T invent/sales
14	130*	2	UMCSENTx	Consumer Sentiment Index	hhsntn	Consumer expect

Group 5: Money and credit

	id	tcode	fred	description	gsi	gsi:description
1	70	6	M1SL	M1 Money Stock	M_110154984	M1
2	71	6	M2SL	M2 Money Stock	M_110154985	M2
3	72	5	M2REAL	Real M2 Money Stock	M_110154985	M2 (real)
4	73	6	BOGMBASE	Monetary Base	M_110154995	MB
5	74	6	TOTRESNS	Total Reserves of Depository Institutions	M_110155011	Reserves tot
6	75	7	NONBORRES	Reserves Of Depository Institutions	M_110155009	Reserves nonbor
7	76	6	BUSLOANS	Commercial and Industrial Loans	BUSLOANS	C&I loan plus
8	77	6	REALLN	Real Estate Loans at All Commercial Banks	BUSLOANS	DC&I loans
9	78	6	NONREVSL	Total Nonrevolving Credit	M_110154564	Cons credit
10	79*	2	CONSPI	Nonrevolving consumer credit to Personal Income	M_110154569	Inst cred/PI
11	131	6	MZMSL	MZM Money Stock	N.A.	N.A.
12	132	6	DTCOLNVHFNM	Consumer Motor Vehicle Loans Outstanding	N.A.	N.A.
13	133	6	DTCTHFNM	Total Consumer Loans and Leases Outstanding	N.A.	N.A.
14	134	6	INVEST	Securities in Bank Credit at All Commercial Banks	N.A.	N.A.

Group 6: Interest and exchange rates

	id	tcode	fred	description	gsi	gsi:description
1	84	2	FEDFUNDS	Effective Federal Funds Rate	M_110155157	Fed Funds
2	85*	2	CP3Mx	3-Month AA Financial Commercial Paper Rate	CPF3M	Comm paper
3	86	2	TB3MS	3-Month Treasury Bill:	M_110155165	3 mo T-bill
4	87	2	TB6MS	6-Month Treasury Bill:	M_110155166	6 mo T-bill
5	88	2	GS1	1-Year Treasury Rate	M_110155168	1 yr T-bond
6	89	2	GS5	5-Year Treasury Rate	M_110155174	5 yr T-bond
7	90	2	GS10	10-Year Treasury Rate	M_110155169	10 yr T-bond
8	91	2	AAA	Moody's Seasoned Aaa Corporate Bond Yield		Aaa bond
9	92	2	BAA	Moody's Seasoned Baa Corporate Bond Yield		Baa bond
10	93*	1	COMPAPFFx	3-Month Commercial Paper Minus FEDFUNDS		CP-FF spread
11	94	1	TB3SMFFM	3-Month Treasury C Minus FEDFUNDS		3 mo-FF spread
12	95	1	TB6SMFFM	6-Month Treasury C Minus FEDFUNDS		6 mo-FF spread
13	96	1	T1YFFM	1-Year Treasury C Minus FEDFUNDS		1 yr-FF spread
14	97	1	T5YFFM	5-Year Treasury C Minus FEDFUNDS		5 yr-FF spread
15	98	1	T10YFFM	10-Year Treasury C Minus FEDFUNDS		10 yr-FF spread
16	99	1	AAAFFM	Moody's Aaa Corporate Bond Minus FEDFUNDS		Aaa-FF spread
17	100	1	BAAFFM	Moody's Baa Corporate Bond Minus FEDFUNDS		Baa-FF spread
18	101	5	TWEXAFEGSMTHx	Trade Weighted U.S. Dollar Index		Ex rate: avg
19	102*	5	EXSZUSx	Switzerland / U.S. Foreign Exchange Rate	M_110154768	Ex rate: Switz
20	103*	5	EXJPUSx	Japan / U.S. Foreign Exchange Rate	M_110154755	Ex rate: Japan
21	104*	5	EXUSUKx	U.S. / U.K. Foreign Exchange Rate	M_110154772	Ex rate: UK
22	105*	5	EXCAUSx	Canada / U.S. Foreign Exchange Rate	M_110154744	EX rate: Canada

Group 7: Prices

	id	tcode	fred	description	gsi	gsi:description
1	106	6	WPSFD49207	PPI: Finished Goods	M110157517	PPI: fin gds
2	107	6	WPSFD49502	PPI: Finished Consumer Goods	M110157508	PPI: cons gds
3	108	6	WPSID61	PPI: Intermediate Materials	M_110157527	PPI: int matls
4	109	6	WPSID62	PPI: Crude Materials	M_110157500	PPI: crude matls
5	110*	6	OILPRICEx	Crude Oil, spliced WTI and Cushing	M_110157273	Spot market price
6	111	6	PPICMM	PPI: Metals and metal products:	M_110157335	PPI: nonferrous
7	112	1	NAPMPRI	ISM Manufacturing: Prices Index	M_110157204	NAPM com price
8	113	6	CPIAUCSL	CPI : All Items	M_110157323	CPI-U: all
9	114	6	CPIAPPSL	CPI : Apparel	M_110157299	CPI-U: apparel
10	115	6	CPITRNSL	CPI : Transportation	M_110157302	CPI-U: transp
11	116	6	CPIMEDSL	CPI : Medical Care	M_110157304	CPI-U: medical
12	117	6	CUSR0000SAC	CPI : Commodities	M_110157314	CPI-U: comm.
13	118	6	CUUR0000SAD	CPI : Durables	M_110157315	CPI-U: dbles
14	119	6	CUSR0000SAS	CPI : Services	M_110157325	CPI-U: services
15	120	6	CPIULFSL	CPI : All Items Less Food	M_110157328	CPI-U: ex food
16	121	6	CUUR0000SA0L2	CPI : All items less shelter	M_110157329	CPI-U: ex shelter
17	122	6	CUSR0000SA0L5	CPI : All items less medical care	M_110157330	CPI-U: ex med
18	123	6	PCEPI	Personal Cons. Expend.: Chain Index	gmdc	PCE defl
19	124	6	DDURRG3M086SBEA	Personal Cons. Exp: Durable goods	gmdcd	PCE defl: dlbes
20	125	6	DNDGRG3M086SBEA	Personal Cons. Exp: Nondurable goods	gmdcn	PCE defl: nondble
21	126	6	DSERRG3M086SBEA	Personal Cons. Exp: Services	gmdcs	PCE defl: service

Group 8: Stock market

	id	tcode	fred	description	gsi	gsi:description
1	80*	5	S&P 500	S&P's Common Stock Price Index: Composite	M_110155044	S&P 500
2	81*	5	S&P: indust	S&P's Common Stock Price Index: Industrials	M_110155047	S&P: indust
3	82*	2	S&P div yield	S&P's Composite Common Stock: Dividend Yield		S&P div yield
4	83*	5	S&P PE ratio	S&P's Composite Common Stock: Price-Earnings Ratio		S&P PE ratio
5	135*	1	VXOCLSx	VXO		