Online Empirical Supplement to "A One Covariate at a Time, Multiple Testing Approach to Variable Selection in High-Dimensional Linear Regression Models"

A. Chudik
Federal Reserve Bank of Dallas

G. Kapetanios
King's College, London

M. Hashem Pesaran
University of Southern California and Trinity College, Cambridge

24 June 2018

1 Introduction

This supplement to Chudik, Kapetanios, and Pesaran (2018, hereafter CKP) provides a description of the individual methods employed in the empirical illustration, and additional empirical results. The empirical illustration is set out in Section 6 of CKP. Section 2 below describes the forecasting exercise, and Section 3 reports additional empirical results.

2 Description of the forecasting exercise

We forecast the U.S. GDP growth and CPI inflation using a set of macroeconomic variables. We use the smaller dataset considered in Stock and Watson (2012), which contains 109 series. The series are transformed by taking logarithms and/or differencing following Stock and Watson (2012).[1] After transformations, the available sample is 1960Q3:2008Q4, or $T = 194$. Let $\boldsymbol{\xi}_t = \left(\xi_{1t}, \xi_{2t}, ..., \xi_{n-1,t}\right)'$ be a vector of the 109 transformed variables. Define the $n \times 1$ vector $\mathbf{x}_t = (\boldsymbol{\xi}_t, y_t, y_{t-1}, y_{t-2}, y_{t-3})'$ considered below, where y_t is either the first-differenced log of real gross domestic product, or the second differenced log of consumer price index.

We are interested in forecasting y_{t+1} with the predictors in $\mathbf{x}_t$ and common factors $\mathbf{f}_t$ extracted from variables in $\mathbf{z}_t^s$, where $\mathbf{z}_t^s$ is the standardized $\mathbf{z}_t = \left(y_t, \boldsymbol{\xi}_t'\right)'$ (by subtracting its sample mean and dividing each series by its sample standard deviation). We consider:

(a) the AR(h) model,

$$y_t = \sum_{\ell=1}^{h} \rho_\ell y_{t-\ell} + v_t,$$

which we use as a benchmark. The lag order h is selected using the SBC criterion with the maximum number of lags set equal to $h_{\max} = 4$.

[1] For further details, see the online supplement of Stock and Watson (2012), in particular columns E and T of their Table B.1.

Data-rich forecasting methods are:

(b) The factor-augmented AR,

$$y_t = \sum_{\ell=1}^{h} \rho_\ell y_{t-\ell} + \boldsymbol{\gamma}' \mathbf{f}_{t-1} + v_t,$$

where $\mathbf{f}_t$ is $m \times 1$ vector of unobserved common factors extracted from variables in $\mathbf{z}_t^s$. We use Bai and Ng's PC_{p1} criterion to select the number of factors (m) with the maximum number of factors set to 5. The vector of unobserved factors, $\mathbf{f}_t$, is estimated using the method of principal components. Same as in the AR case, the lag order h is selected using the SBC criterion with the maximum number of lags set equal to $h_{\max} = 4$.

(c) Lasso method, implemented in the same way as described in Section 2 of the online Monte Carlo supplement of CKP using $\left(\mathbf{x}'_{t-1}, \mathbf{f}'_{t-1}\right)$ as the vector of predictors for y_t.

(d) Adaptive Lasso method, implemented in the same way as described in Section 2 of the online Monbte Carlo supplement of CKP using $\left(\mathbf{x}'_{t-1}, \mathbf{f}'_{t-1}\right)$ as the vector of predictors for y_t.

(e-g) OCMT method. We use OCMT described in CKP to select the relevant variables from the vector $\mathbf{x}_{t-1}$ to forecasts the target variable y_t. We set $p = 0.01$ (e), 0.05 (f) and 0.1 (g), and $(\delta, \delta^*) = (1, 2)$, and we always include c (intercept) , and $\mathbf{f}_{t-1}$ (lagged factors) in the testing regressions. Next, we use the selected variables together with c, and $\mathbf{f}_{t-1}$ in an ordinary least squares regression for y_t.

We use a rolling window of $T = 120$ time periods, which leaves us with the last $H = 74$ out-of-sample evaluation periods, 1990Q3-2008Q4. We also consider pre-crisis evaluation subsample, 1990Q3-2007Q2 with $H = 68$ periods, to evaluate the sensitivity of results to exclusion of the global financial crisis from the sample.

3 Results

Table 1 reports the root mean squared forecasting error (RMSFE) findings for all forecasting methods. Diebold-Mariano (DM) test statistics for testing $H_0 : E\left(\hat{v}_{ij,t}\right) = 0$, where $\hat{v}_{ij,t} = \hat{e}_{i,t}^2 - \hat{e}_{j,t}^2$ is the difference between the squared forecasting errors of methods i and j, are presented in Table 2. The DM statistics is computed assuming serially uncorrelated one-step-ahead forecasting errors. Specifically

$$DM_{ij} = \sqrt{H} \frac{\overline{\hat{v}}_{H,ij}}{\hat{\sigma}_{H,ij}}, \tag{1}$$

where $H = 68$ or 74 (depending on the evaluation period) is the length of the evaluation period, $\overline{\hat{v}}_{H,ij} = H^{-1} \sum_{t=T+1}^{T+H} \hat{v}_{ij,t}$ is the sample mean of $\hat{v}_{ij,t}$, and

$$\hat{\sigma}_{H,ij} = \sqrt{\frac{1}{H} \sum_{t=T+1}^{T+H} \hat{v}_{ij,t}^2}.$$

Table 1: RMSFE performance of the AR, factor-augmented AR, Lasso, adaptive Lasso, and OCMT methods

Evaluation sample:	Full 1990Q3-2008Q4		Pre-crisis 1990Q3-2007Q2	
	RMSFE ($\times$100)	Relative RMSFE	RMSFE ($\times$100)	Relative RMSFE
	Real output growth			
(a) *AR* benchmark	0.561	1.000	0.505	1.000
(b) Factor-augmented *AR*	0.484	0.862	0.470	0.930
(c) Lasso	0.510	0.910	0.465	0.922
(d) Adaptive Lasso	0.561	1.000	0.503	0.996
(e) OCMT, $p = 0.01$	0.495	0.881	0.479	0.948
(f) OCMT, $p = 0.05$	0.477	0.850	0.461	0.912
(g) OCMT, $p = 0.1$	0.490	0.874	0.464	0.918
	Inflation			
(a) $AR(1)$ benchmark	0.601	1.000	0.435	1.000
(b) Factor-augmented $AR(1)$	0.557	0.927	0.415	0.954
(c) Lasso	0.599	0.997	0.462	1.063
(d) Adaptive Lasso	0.715	1.190	0.524	1.205
(e) OCMT, $p = 0.01$	0.596	0.992	0.472	1.086
(f) OCMT, $p = 0.05$	0.590	0.982	0.464	1.068
(g) OCMT, $p = 0.1$	0.595	0.990	0.471	1.084

Notes: RMSFE is computed using a rolling forecasting scheme with a rolling window of 120 observations. We use the smaller dataset considered in Stock and Watson (2012) which contains 109 series. The series are transformed by taking logarithms and/or differencing following Stock and Watson (2012). The transformed series span 1960Q3 to 2008Q4 and are collected in the vector $\boldsymbol{\xi}_t$. Set of regressors in Lasso and adaptive-Lasso contains $h_{\max} = 4$ lags of y_t (lagged target variables), $\boldsymbol{\xi}_{t-1}$, and a lagged set of principal components obtained from the large dataset given by $(y_t, \boldsymbol{\xi}_t')'$. OCMT procedure is applied to regressions of y_t conditional on lagged principal components, with elements of $\boldsymbol{\xi}_{t-1}$ and $h_{\max} = 4$ lags of y_t considered one at a time. OCMT is reported for $\delta = 1$ in the first stage, and $\delta^* = 2$ in the subsequent stages of the OCMT procedure, and three choices of p, similarly to the MC section of CKP. The number of principal components in the factor-augmented *AR*, Lasso, adaptive Lasso, and OCMT methods is determined in a rolling scheme by using criterion PC_{p_1} of Bai and Ng (2002) (with the maximum number of PCs set to 5). See Section 2 for further details.

Table 2: DM statistics for the forecasting performance of the AR, factor-augmented AR, Lasso, adaptive Lasso, and OCMT methods

	DM_{ij} test statistics													
Full evaluation sample: 1990Q3-2008Q4														
	Real output growth							**Inflation**						
Method pair i (below), j (on right)	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(a)	(b)	(c)	(d)	(e)	(f)	(g)
(a) AR(1)	.	1.50	1.95	0.00	1.49	1.73	1.44	.	1.12	0.06	-2.55	0.12	0.28	0.14
(b) Factor-augmented $AR(1)$	-1.50	.	-0.67	-1.39	-0.59	0.43	-0.38	-1.12	.	-1.89	-2.06	-2.39	-2.07	-2.09
(c) Lasso	-1.95	0.67	.	-1.76	0.45	0.92	0.57	-0.06	1.89	.	-1.82	0.14	0.45	0.20
(d) Adaptive Lasso	0.00	1.39	1.76	.	1.29	1.56	1.31	2.55	2.06	1.82	.	1.61	1.69	1.62
(e) OCMT, $p = 0.01$	-1.49	0.59	-0.45	-1.29	.	1.32	0.24	-0.12	2.39	-0.14	-1.61	.	0.49	0.08
(f) OCMT, $p = 0.05$	-1.73	-0.43	-0.92	-1.56	-1.32	.	-1.21	-0.28	2.07	-0.45	-1.69	-0.49	.	-0.71
(g) OCMT, $p = 0.05$	-1.44	0.38	-0.57	-1.31	-0.24	1.21	.	-0.14	2.09	-0.20	-1.62	-0.08	0.71	.
Pre-Crisis evaluation sample: 1990Q3-2007Q2														
(a) AR(1)	.	0.95	1.60	0.13	0.84	1.19	1.11	.	0.98	-1.13	-2.28	-1.54	-1.01	-1.18
(b) Factor-augmented $AR(1)$	-0.95	.	0.14	-0.88	-0.48	0.52	0.34	-0.98	.	-1.66	-2.31	-2.46	-2.21	-2.21
(c) Lasso	-1.60	-0.14	.	-1.39	-0.48	0.16	0.06	1.13	1.66	.	-1.78	-0.47	-0.07	-0.37
(d) Adaptive Lasso	-0.13	0.88	1.39	.	0.66	1.07	1.00	2.28	2.31	1.78	.	1.22	1.31	1.15
(e) OCMT, $p = 0.01$	-0.84	0.48	0.48	-0.66	.	1.22	0.82	1.54	2.46	0.47	-1.22	.	0.46	0.05
(f) OCMT, $p = 0.05$	-1.19	-0.52	-0.16	-1.07	-1.22	.	-0.33	1.01	2.21	0.07	-1.31	-0.46	.	-0.71
(g) OCMT, $p = 0.05$	-1.11	-0.34	-0.06	-1.00	-0.82	0.33	.	1.18	2.21	0.37	-1.15	-0.05	0.71	.

Notes: This table reports results for DM_{ij} statistics defined in (1). See also notes to Table 1.

References

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*, 191–221.

Chudik, A., G. Kapetanios, and M. H. Pesaran (2018). A one-covariate at a time, multiple testing approach to variable selection in high-dimensional linear regression models. Econometrica, forthcoming.

Stock, J. H. and M. W. Watson (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics 30*, 481–493.