



Contents lists available at ScienceDirect

## International Journal of Forecasting

journal homepage: [www.elsevier.com/locate/ijforecast](http://www.elsevier.com/locate/ijforecast)

# Macroeconomic forecasting and structural analysis through regularized reduced-rank regression

Emmanuela Bernardini<sup>a</sup>, Gianluca Cubadda<sup>b,\*</sup><sup>a</sup> Banca d'Italia, Roma, Italy<sup>b</sup> Dipartimento di Economia e Finanza, Università di Roma "Tor Vergata", Via Columbia 2, 00133 Roma, Italy

## ARTICLE INFO

## Keywords:

Canonical correlation analysis  
Vector autoregressive models  
Shrinkage estimation  
Macroeconomic prediction

## ABSTRACT

This paper proposes a strategy for detecting and imposing reduced-rank restrictions in medium vector autoregressive models. It is known that Canonical Correlation Analysis (CCA) does not perform well in this framework, because inversions of large covariance matrices are required. We propose a method that combines the richness of reduced-rank regression with the simplicity of naïve univariate forecasting methods. In particular, we suggest the usage of a proper shrinkage estimator of the autocovariance matrices that are involved in the computation of CCA, in order to obtain a method that is asymptotically equivalent to CCA, but numerically more stable in finite samples. Simulations and empirical applications document the merits of the proposed approach for both forecasting and structural analysis.

© 2013 International Institute of Forecasters. Published by Elsevier B.V. All rights reserved.

## 1. Introduction

In principle, the existence of co-movements among economic time series implies that multivariate forecasting methods should outperform univariate techniques. Indeed, this is the main theoretical motivation for the development of models for large multivariate data sets, such as factor models (Forni, Hallin, Lippi, & Reichlin, 2000, 2005; Stock & Watson, 2002a,b) and Bayesian vector autoregressions (Banbura, Giannone, & Reichlin, 2010). In practice, however, univariate forecasting models are hard to beat, particularly for short horizons. In a recent paper, Carriero, Kapetanios, and Marcellino (2011) explore the merits of the best available technology for forecasting large datasets, and conclude that, for one- and two-step-ahead forecasts, there are no multivariate models that are able to beat the univariate autoregressive benchmark. One possible

explanation of this finding is that the existence of common components among time series leads to simple univariate models (Cubadda, Hecq, & Palm, 2009) which can be identified and estimated easily, whereas the efficient multivariate modelling of many time series is a hard task.

In a vector autoregressive (VAR) framework, the presence of common components is equivalent to imposing proper reduced rank structures on the model coefficient matrices. At the statistical level, these restrictions can be tested for and imposed on the estimation by means of Reduced-Rank Regression (RRR) techniques, see *inter alia* Cubadda (2007) and the references therein.

We propose a method that combines the richness of RRR with the simplest univariate forecasting method, i.e., each series is forecast by its unconditional mean. To this end, we resort to a proper shrinkage estimator of the autocorrelation matrices for computing RRR, in place of the natural estimator. The resulting estimator is asymptotically equivalent to the Maximum Likelihood (ML) solution, but it is numerically more stable in finite samples.

In order to check the practical value of the proposed method, we consider its performances when applied to medium VAR models. Indeed, Banbura et al. (2010) and

\* Corresponding author.

E-mail addresses: [emmanuela.bernardini@bancaditalia.it](mailto:emmanuela.bernardini@bancaditalia.it) (E. Bernardini), [gianluca.cubadda@uniroma2.it](mailto:gianluca.cubadda@uniroma2.it) (G. Cubadda).

URL: <http://www.economia.uniroma2.it/nuovo/facolta/docenti/curriculum/GianlucaCubadda.html> (G. Cubadda).

Koop (2013) provide convincing evidence that no substantial predictive gains are obtained by increasing the dimension of the VAR beyond 20. Hence, we focus on reduced-rank VAR models, where the number of predictors is considerably larger than in typical small-scale forecasting models, but not large enough to resort to statistical inference that is based on methods involving double asymptotics (see for example Cubadda & Guardabascio, 2012). We show, by both simulations and empirical applications, that the proposed approach performs well with respect to traditional medium-size macroeconomic methods, and demonstrate that our new method improves both the forecasting and estimation of structural parameters.

The paper is organized as follows. Section 2 discusses the theoretical aspects. Section 3 compares various forecasting procedures in an empirical application to US economic variables. Section 4 uses simulations to evaluate the merits of the various methods in terms of model specifications, forecasting performances, and precision in estimating structural parameters. Section 5 draws some conclusions, and the proofs of the theorems are reported in the Appendix.

## 2. Theory

We start this section by briefly reviewing the reduced-rank VAR model for stationary time series (see Reinsel & Velu, 1998, and the references therein) and the associated estimation issues.

Consider an  $n$ -vector time series  $y_t$  generated by the following stable VAR( $p$ ) model:

$$y_t = \sum_{i=1}^p \Phi_i y_{t-i} + \varepsilon_t = \beta' x_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, T,$$

where  $\beta$  is a  $pn \times n$  coefficient matrix,  $x_t = [y_t', \dots, y_{t-p+1}']'$ , and  $\varepsilon_t$  are i.i.d. innovations with  $E(\varepsilon_t) = 0$ ,  $E(\varepsilon_t \varepsilon_t') = \Sigma_{\varepsilon\varepsilon}$  (positive definite), and finite fourth moments. For the sake of simplicity, no deterministic terms are included.

If we assume that the series  $y_t$  exhibits the serial correlation common feature (Engle & Kozicki, 1993; Vahid & Engle, 1993), we can rewrite the VAR as an RRR model

$$y_t = A\psi'x_{t-1} + \varepsilon_t = AF_{t-1} + \varepsilon_t,$$

where  $A$  and  $\psi$  are, respectively, full-rank  $n \times q$  and  $np \times q$  matrices, and  $F_t = \psi'x_t$  are  $q$  common factors.

One well-known method of obtaining the factor weights  $\psi$  is Canonical Correlation Analysis (CCA). In particular, the matrix  $\psi$  lies in the space generated by  $[v_1, \dots, v_q]$ , where  $v_i$  ( $i = 1, 2, \dots, q$ ) is the eigenvector associated with the  $i$ th largest eigenvalue of the matrix

$$\Sigma_{xx}^{-1} \Sigma_{xy} \Sigma_{yy}^{-1} \Sigma_{yx},$$

where  $\Sigma_{xx} = E(x_{t-1}x_{t-1}')$ ,  $\Sigma_{yy} = E(y_t y_t')$ , and  $\Sigma_{xy} = E(x_{t-1}y_t')$ . Finally, the matrix  $A$  is obtained (up to an identification matrix) by regressing  $y_t$  on the  $q$  canonical factors  $[v_1, \dots, v_q]'x_{t-1}$ , see, *inter alia*, Anderson (1984) for further details.

As a statistical method, CCA has numerous merits, of which the most important are that it provides the ML

solution under the Gaussianity assumption and is invariant to non-singular linear transformations of both  $y_t$  and  $x_{t-1}$ . Moreover, Carriero et al. (2011) have recently extended consistency results for the case that the number of predictors  $N = np$  diverges more slowly than  $T$ .

However, CCA suffers from some limitations when the system dimension is large. First, CCA is unfeasible when the number of observations  $T$  is less than the number of predictors  $N$ . Second, even when  $T$  is large, the CCA solution is numerically unstable and statistical inference is unreliable in a medium  $N$  framework, see Cubadda and Hecq (2011).

We propose to solve these problems by shrinking the sample autocorrelation matrix of series  $y_t$ , which CCA is based on at the sample level. In particular, CCA is usually performed by solving the eigenvector equation  $S_{xx}^{-1}S_{xy}S_{yy}^{-1}S_{yx}\hat{v} = \hat{\lambda}\hat{v}$ , where

$$S = \begin{bmatrix} S_{yy} & S_{yx} \\ S_{xy} & S_{xx} \end{bmatrix}$$

is the sample covariance matrix of the series  $w_t = [y_t', x_{t-1}']'$ . We suggest that a proper shrinkage estimator of the covariance matrix  $\Sigma = E(w_t w_t')$  be used in place of  $S$ .

In general, a shrinkage estimator (Stein, 1956) is an optimally weighted average of two existing estimators, an asymptotically unbiased estimator that suffers from a large estimation error, and a target one that is biased, but with a lower estimation error. We propose a regularized version of CCA that requires the solution of the eigenvector equation  $Z_{xx}^{-1}Z_{xy}Z_{yy}^{-1}Z_{yx}\tilde{v} = \tilde{\lambda}\tilde{v}$ , where

$$Z = \alpha D + (1 - \alpha)S,$$

$D$  is a diagonal matrix having the same diagonal as  $S$ , and  $\alpha \in [0, 1]$ .

Note that when  $\alpha = 1$ , the full-rank regression case coincides with  $n$  univariate white noises, whereas when  $\alpha = 0$ , we go back to the usual CCA solution. Hence, in a way, we are shrinking RRR towards the simplest forecasting univariate model, i.e., the white noise. When variables are made stationary by differencing, our target is equivalent to the so-called Minnesota prior in Bayesian VAR modelling, where the forecasts of the time series levels are shrunk towards random walks (see, *inter alia*, Litterman, 1986).

In the choice of the optimal  $\alpha$ , we follow the data-based approach of Ledoit and Wolf (2003), which has the advantage of providing a closed form solution to the optimization problem. In particular, Ledoit and Wolf (2003) propose the minimization of a risk function based on the Frobenius norm of the difference between the shrinkage estimator  $Z$  and the covariance matrix  $\Sigma$ , i.e.,

$$\begin{aligned} R(\alpha) &= E(\|Z - \Sigma\|^2) \\ &= E \left[ \sum_i^{N+n} \sum_j^{N+n} (\alpha d_{ij} + (1 - \alpha)s_{ij} - \sigma_{ij})^2 \right], \end{aligned}$$

and the solution to this optimization problem is

$$\alpha^* = \frac{\sum_{i \neq j} \text{Var}(s_{ij})}{\sum_{i \neq j} [\text{Var}(s_{ij}) + \sigma_{ij}^2]},$$

which has a neat interpretation: the weight on the target estimator, which is a diagonal matrix having the same diagonal as  $S$ , increases as either the squared covariances between the series  $w_t$  decrease or the variances of the entries of  $S$  increase.

Ledoit and Wolf (2003) propose a feasible estimator of  $\alpha^*$ , but their solution is only valid for i.i.d. data. Hence, we need to extend their approach to the case that data are generated by vector stationary processes. To this end, we provide the following theorems.

**Theorem 1.** *If  $y_t$  is generated by a linear process with absolutely summable coefficient matrices and errors  $\varepsilon_t$  with finite fourth cumulants, then*

$$\alpha^* = O(1/T).$$

**Proof.** See Appendix.

**Theorem 2.** *A consistent estimator of the optimal shrinkage constant  $\lim_T T\alpha^*$  is obtained using  $T\hat{\alpha}^*$ , where*

$$\hat{\alpha}^* = \frac{\sum_{i \neq j} \widehat{LRV}(v_t^{ij})}{\sum_{i \neq j} [\widehat{LRV}(v_t^{ij}) + Ts_{ij}^2]},$$

$v_t^{ij} = w_{it}w_{jt}$ , and  $\widehat{LRV}(v_t^{ij})$  is a consistent kernel estimator of  $LRV(v_t^{ij})$ , the long-run variance of  $v_t^{ij}$ .

**Proof.** See the Appendix.

**Remark 3.** Sancetta (2008) also provides an extension of the Ledoit and Wolf (2003) approach to the case of dependent data. However, our results differ from those of Sancetta (2008) in two ways. First, he assumes that both  $N$  and  $T$  diverge, whereas we actually keep  $N$  fixed, as it is appropriated for the medium  $N$  framework that we are considering here. Second, our theorems refer to the whole autocorrelation matrix  $S$ , rather than to the contemporaneous variance matrix  $S_{yy}$  only.

Note that if we compute CCA using  $Z^*$  (i.e.,  $Z$  with  $\alpha = \hat{\alpha}^*$ ) in place of  $S$ , we obtain estimators of the canonical coefficients and weights that are more stable numerically than those obtained by CCA. In what follows, we call this procedure Regularized Canonical Correlation Analysis (RCCA).

**Remark 4.** In light of Theorems 1 and 2, we see that RCCA is asymptotically equivalent to CCA. Hence, standard inferential results for CCA apply to RCCA as well.

### 3. Empirical analysis

In order to perform our empirical out-of-sample forecasting exercise, we use a data-set similar to those of Banbura et al. (2010) and Cubadda and Guardabascio (2012). It consists of 20 US monthly time series divided into three groups: (i) real variables such as industrial production and employment; (ii) asset prices such as stock prices and exchange rates; and (iii) nominal variables such as consumer and producer price indices, wages, and money aggregates.

The time period is from 1960.01 to 2011.12, leading to an overall sample size of  $T^* = 624$ . All of the variables are transformed in order to achieve stationarity, then demeaned and standardized to a unit variance. The series under analysis and the transformations used are reported in Table 1.<sup>1</sup>

The forecasting exercise is performed using a rolling window of 10 years, and hence,  $T = 120$ . For five forecasting methods, direct forecasts of  $y_{\tau+h}$  are computed using the predictors  $x_\tau = [y'_\tau, \dots, y'_{\tau-p+1}]'$  for  $\tau = T + p, \dots, T^* - h$  and  $h = 1, 3, 6, 12$ . In particular, the competing methods are:

1. The  $h$ -step-ahead Ordinary Least Squares (OLS) forecasts, obtained as  $\hat{\beta}^{h'}x_\tau$ , where

$$\hat{\beta}^h = (X'X)^{-1}X'Y.$$

$$X = [x_{\tau-T+1}, \dots, x_{\tau-h}]', \text{ and } Y = [y_{\tau-T+1+h}, \dots, y_\tau]'$$

2. The Regularized Least Squares (RLS) forecast, obtained as  $\hat{\beta}_{\text{RLS}}^{h'}x_\tau$ , where

$$\hat{\beta}_{\text{RLS}}^h = Z^{*-1}Z_{xy}^*,$$

and  $Z^* = \begin{bmatrix} Z_{yy}^* & Z_{yx}^* \\ Z_{xy}^* & Z_{xx}^* \end{bmatrix}$  is our shrinkage estimator of the covariance matrix of  $[y'_\tau, x'_{\tau-h}]'$ .

3. The  $h$ -step-ahead CCA forecasts, obtained as

$$\hat{\beta}_{\text{CCA}}^{h'}x_\tau = Y'\hat{F}_{\text{CCA}}(\hat{F}'_{\text{CCA}}\hat{F}_{\text{CCA}})^{-1}\hat{F}_{\text{CCA},\tau},$$

where  $\hat{F} = (\hat{F}_{\text{CCA},\tau-T+1}, \dots, \hat{F}_{\text{CCA},\tau-h})'$  and  $\hat{F}_{\text{CCA},t} = (\hat{f}_{\text{CCA},1,t}, \dots, \hat{f}_{\text{CCA},q,t})'$  are the first  $q$  canonical factors estimated by CCA.

4. The  $h$ -step-ahead RCCA forecasts, obtained as

$$\hat{\beta}_{\text{RCCA}}^{h'}x_\tau = Y'\hat{F}_{\text{RCCA}}(\hat{F}'_{\text{RCCA}}\hat{F}_{\text{RCCA}})^{-1}\hat{F}_{\text{RCCA},\tau},$$

where  $\hat{F} = (\hat{F}_{\text{RCCA},\tau-T+1}, \dots, \hat{F}_{\text{RCCA},\tau-h})'$  and  $\hat{F}_{\text{RCCA},t} = (\hat{f}_{\text{RCCA},1,t}, \dots, \hat{f}_{\text{RCCA},q,t})'$  are the first  $q$  canonical factors estimated by RCCA.

5. The  $h$ -step-ahead Principal Component Regression (PCR) forecasts, obtained as  $\hat{\beta}_{\text{PCR}}^{h'}x_\tau$ , where

$$\hat{\beta}_{\text{PCR}}^h = \hat{\tau}_q \hat{\Lambda}_q^{-1} \hat{\gamma}_q' x'_\tau Y.$$

$\hat{\Lambda}_q$  is the diagonal matrix of the  $q$  largest eigenvalues of  $\sum_{t=\tau-T+1}^\tau x'_t x_t$ , and  $\hat{\tau}_q$  are the associated eigenvectors.

In order to fix the lag order  $p$ , we look at the information criteria by Akaike (AIC), Schwarz (SIC) and Hannan–Quinn (HQIC), having estimated the VAR models by both OLS and RLS. With the former method, AIC suggests  $p = 2$ , whereas both SIC and HQIC indicate  $p = 1$ ; on the other hand, when RLS is used in estimation, AIC suggests  $p = 3$ , whereas SIC and HQIC indicate  $p = 1$  and  $p = 2$ , respectively. Hence, we compare the results for the various methods, letting  $p$  vary from 1 to 3, and find no significant gains from using a lag order longer than 2. We therefore estimate a VAR(2) model for all of the multivariate forecasting methods.

<sup>1</sup> Since the JoC-ECRI Industrial Price Index was not available before 1985, its values for the period 1960.01–1984.12 are reconstructed using the index of sensitive materials prices.

**Table 1**  
Data description.

Variable	Transformation	Acronym
Personal Income	$(1 - L) \ln$	PI
Real Consumption	$(1 - L) \ln$	RC
Industrial Production Index	$(1 - L) \ln$	IP
Total Capacity Utilization	$(1 - L)$	TCU
Unemployment Rate	$(1 - L)$	UR
Employees on Nonfarm Payrolls Private	$(1 - L) \ln$	EMP
Housing Starts	$\ln$	HS
Money Stock: M1	$(1 - L^{12})(1 - L) \ln$	M1
Money Stock: M2	$(1 - L^{12})(1 - L) \ln$	M2
Institutional Total Adjusted Reserves	$(1 - L^{12})(1 - L) \ln$	TAR
Nonborrowed Reserves	$(1 - L^{12})(1 - L) \ln$	NBR
Standard and Poor's Common Stock Price Index	$(1 - L) \ln$	SP500
Effective Federal Funds Interest Rate	$(1 - L)$	FYFF
US Treasury Interest Rate	$(1 - L)$	TIR
US Effective Exchange Rate	$(1 - L) \ln$	EER
Producer Price Index: Finished Goods	$(1 - L^{12})(1 - L) \ln$	PPI
Consumer Price Index	$(1 - L^{12})(1 - L) \ln$	CPI
Personal Consumption Expenditure Deflator	$(1 - L^{12})(1 - L) \ln$	CED
JoC-ECRI Industrial Price Index	$(1 - L^{12})(1 - L) \ln$	JEIPI
Average Hourly Earnings	$(1 - L^{12})(1 - L) \ln$	AHE

The application of CCA, RCCA and PCR requires the number of factors  $q$  to be fixed. For each of these methods,  $q$  is chosen recursively according to AIC, SIC and HQIC.

The application of RCCA and RLS requires consistent estimates of the long-run variances of the cross-products of elements of  $w_t$ . To this end, we opt for the estimator proposed by Newey and West (1987), with a bandwidth equal to  $T^{1/3}$ .<sup>2</sup>

Finally, the benchmark univariate forecasts are obtained through the  $n$  univariate AR( $p$ ) models that are selected recursively by the AIC with  $0 \leq p \leq 6$ .

In order to take the effects of the “Great Moderation” into account, we consider three forecast evaluation samples: 1960.01–2011.12 (full sample), 1960.01–1984.12 (pre-Great Moderation) and 1985.01–2007.12 (Great Moderation).

Tables 2–4 report the sums of the mean square forecast errors relative to the AR univariate forecasts (TRMSFE) of all of the series and the three samples.

The empirical findings indicate that RCCA delivers the most accurate forecasts in 11 cases out 12; the only exception is in the pre-Great Moderation sample for  $h = 6$ , where CCA with  $q$  selected by SIC is the best forecasting method. In particular, the RCCA method performs best when  $q$  is chosen by HQIC. Focusing on the shortest forecasting horizon, the performance of RCCA is comparable to those of the AR benchmark in the full and pre-Great Moderation samples, whereas in the Great Moderation sample, the univariate forecasts outperform all of the multivariate ones for  $h = 1$ .

In addition to considering the overall forecasting performances of the models, we also focus on four key target series: the Industrial Production Index (IP), Employees on non-farm payrolls (EMP), the Effective Federal Funds Interest Rate (FYFF), and the Consumer Price Index (CPI).

**Table 2**  
TRMSFE, sample 1960–2011.

Method/criterion	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RCCA/AIC	108.30	101.49	112.02	114.80
RCCA/SIC	118.73	99.35	<b>93.05</b>	<b>87.28</b>
RCCA/HQIC	<b>101.30</b>	<b>96.04</b>	93.91	92.95
CCA/AIC	125.73	107.22	119.69	122.54
CCA/SIC	106.86	98.48	98.04	104.22
CCA/HQIC	109.19	101.77	105.83	114.93
PCR/AIC	104.26	108.82	115.52	106.15
PCR/SIC	119.49	107.33	103.17	90.71
PCR/HQIC	104.59	105.83	111.45	102.63
OLS	139.58	166.24	167.86	186.82
RLS	107.59	119.50	124.74	122.01

Note: TRMSFE is the sum of the mean square forecast errors relative to the AR univariate forecast. (R) CCA is (regularized) canonical correlation analysis. PCR is principal component regression. OLS (RLS) is ordinary (regularized) least squares. AIC (SIC) (HQIC) is the Akaike (Schwarz) (Hannan–Quinn) information criterion. The best result for each forecasting horizon  $h$  is in bold.

**Table 3**  
TRMSFE, sample 1960–1984.

Method/criterion	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RCCA/AIC	107.49	93.06	92.61	90.60
RCCA/SIC	112.02	91.16	86.57	<b>82.49</b>
RCCA/HQIC	<b>100.31</b>	<b>89.44</b>	87.36	84.97
CCA/AIC	117.49	96.57	95.97	93.11
CCA/SIC	110.67	90.60	<b>85.61</b>	82.54
CCA/HQIC	110.76	94.85	89.42	86.98
PCR/AIC	104.50	101.54	104.82	98.27
PCR/SIC	114.66	98.01	92.69	84.99
PCR/HQIC	105.06	99.52	100.58	93.96
OLS	132.60	143.45	145.71	149.52
RLS	108.03	114.61	118.98	120.80

See the notes to Table 2.

In order to check whether the differences between the MSFEs of the two best forecasting methods are statistically significant, we performed the version of the Diebold and Mariano (1995) test introduced by Harvey, Leybourne, and Newbold (1997). In particular, the null hypothesis of equal

<sup>2</sup> Alternative choices of the bandwidth do not alter the results significantly.

**Table 4**  
TRMSFE, sample 1985–2007.

Method/criterion	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RCCA/AIC	120.51	104.02	106.83	106.66
RCCA/SIC	122.03	105.41	94.50	<b>88.52</b>
RCCA/HQIC	<b>109.87</b>	<b>96.68</b>	<b>90.15</b>	89.98
CCA/AIC	134.49	109.81	117.14	113.46
CCA/SIC	122.08	104.31	93.66	89.27
CCA/HQIC	125.93	102.47	96.59	96.34
PCR/AIC	116.51	112.05	110.89	104.62
PCR/SIC	118.97	104.36	98.95	89.16
PCR/HQIC	112.68	104.81	100.89	96.36
OLS	155.69	181.89	189.62	187.10
RLS	123.46	136.60	144.55	134.74

See the notes to Table 2.

**Table 5**  
IP, RMSFE: sample 1960–2011.

Method/criterion	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RCCA/AIC	106.59	100.63	110.02	138.82
RCCA/SIC	95.12	<b>89.43</b>	<b>86.23</b> *	<b>83.01</b> *
RCCA/HQIC	95.46	91.72	90.10	91.95
CCA/AIC	124.58	110.83	117.61	142.62
CCA/SIC	106.81	94.09	95.78	106.75
CCA/HQIC	107.70	99.85	101.99	124.44
PCR/AIC	100.46	100.13	121.42	136.73
PCR/SIC	<b>87.89</b> **	92.32	92.48	94.64
PCR/HQIC	99.23	94.75	111.10	131.32
OLS	143.08	175.59	177.11	263.82
RLS	105.58	120.49	132.42	151.66

Note: RMSFE is the mean square forecast error relative to that of the AR univariate forecast. See the notes to Table 2 for the meanings of the other acronyms.

\* Indicates significance at the 10% level for the test of equal RMSFEs of the two best methods.

\*\* Indicates significance at the 5% level for the test of equal RMSFEs of the two best methods.

\*\*\* Indicates significance at the 1% level for the test of equal RMSFEs of the two best methods.

**Table 6**  
EMP, RMSFE: sample 1960–2011.

Method/criterion	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RCCA/AIC	115.05	97.17	101.21	124.16
RCCA/SIC	122.16	102.21	<b>86.95</b> **	<b>77.17</b>
RCCA/HQIC	107.18	98.85	88.50	92.10
CCA/AIC	140.12	109.32	107.20	134.44
CCA/SIC	132.47	114.42	98.96	113.12
CCA/HQIC	125.42	109.15	98.70	127.02
PCR/AIC	101.88	100.04	114.99	114.10
PCR/SIC	<b>94.85</b> ***	<b>94.03</b>	96.51	83.69
PCR/HQIC	95.68	100.90	109.28	110.70
OLS	146.52	131.14	165.79	196.44
RLS	111.57	104.13	125.50	130.56

See the notes to Table 5.

MSFEs is tested against the alternative that the second best forecasting model has a larger MSFE.

Tables 5–8 report the MSFEs relative to AR univariate forecasts (RMSFE) for the four series, as well as the test results of equal MSFEs of the two best forecasting methods. In order to save space, the results refer to the full sample only.

The empirical findings indicate that RCCA delivers the most accurate forecasts in about 62.5% of cases. It performs best for IP and EMP for the longer forecasting horizons,

**Table 7**  
IR, Relative MSFE: sample 1960–2011.

Method/criterion	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RCCA/AIC	92.93	94.78	93.79	88.29
RCCA/SIC	91.95	<b>79.77</b>	84.59	<b>80.09</b> **
RCCA/HQIC	94.92	85.69	87.96	84.73
CCA/AIC	107.46	98.84	97.93	88.31
CCA/SIC	94.83	80.56	<b>83.98</b>	83.82
CCA/HQIC	96.51	94.36	88.96	87.74
PCR/AIC	85.19	90.40	97.89	104.07
PCR/SIC	85.70	82.48	86.18	83.31
PCR/HQIC	<b>84.61</b>	87.99	97.67	102.24
OLS	102.41	139.59	167.29	173.45
RLS	88.17	112.65	123.73	137.52

See the notes to Table 5.

**Table 8**  
CPI, RMSFE: sample 1960–2011.

Method/criterion	$h = 1$	$h = 3$	$h = 6$	$h = 12$
RCCA/AIC	100.89	95.92	102.91	109.29
RCCA/SIC	103.55	<b>86.12</b>	<b>83.30</b>	83.59
RCCA/HQIC	<b>92.38</b>	91.06	85.86	101.27
CCA/AIC	121.95	101.82	113.34	106.93
CCA/SIC	101.10	90.54	88.87	103.24
CCA/HQIC	111.77	99.53	99.78	113.09
PCR/AIC	92.66	106.83	110.90	77.65
PCR/SIC	93.93	89.13	98.78	78.50
PCR/HQIC	92.48	99.41	107.49	<b>75.33</b>
OLS	139.15	183.84	161.99	105.06
RLS	98.49	124.22	110.62	81.80

See the notes to Table 5.

and for CPI for short term forecasts. Interestingly, RCCA outperforms the AR benchmark even for short forecasting horizons. The second-best performer is PCR, which produces results superior to those of CCA. For all of the methods, the most accurate forecasts are often obtained when the number of factors  $q$  is chosen according to the SIC.

#### 4. Monte Carlo analysis

In this section, we perform a Monte Carlo study in order to evaluate the finite sample performances of CCA, RCCA, OLS and RLS in a VAR framework. We do this by means of various criteria: first, the sum of the MSFEs relative to univariate AR( $p$ ) forecasts for the forecasting horizon  $h = 1$ ; second, the percentage of cases in which the various methods/criteria detect the true number  $q$  of relevant factors; and third, the Frobenius distance between the estimated structural parameters and the true ones, relative to the Frobenius distance of OLS (RFD). The structural parameters are recovered through the Cholesky factorization of the error covariance matrix  $\Sigma_{EE}$ .

For each criterion, we check whether the differences between the two best performers are statistically significant. In particular, we compute significance tests for the mean of the differences between the two smallest TRMSFEs and RFDs. In order to evaluate the significance of the differences between the proportions in which the estimated number of factors  $\hat{q}$  is equal to the true one  $q$ , we use the McNemar test for the null hypothesis that the two best methods identify the correct number of factors with the same probability.



**Table 9**  
Simulation results.

<i>N</i> = 20, <i>T</i> = 120				
Method/criterion	Statistics	Full sample <i>q</i> = 5	Pre-great moderation <i>q</i> = 2	Great moderation <i>q</i> = 2
RCCA/AIC	% in which $\hat{q} = q$	<b>29.32<sup>***</sup></b>	10.96	15.56
	% in which $\hat{q} > q$	67.96	89.04	84.44
	RFD	72.72	70.43	<b>55.72<sup>***</sup></b>
	TRMSFE	<b>105.85<sup>***</sup></b>	100.87	106.01
RCCA/SIC	% in which $\hat{q} = q$	0.00	1.56	2.64
	% in which $\hat{q} > q$	0.00	0.00	0.00
	RFD	<b>71.61<sup>***</sup></b>	71.54	57.12
	TRMSFE	133.84	106.79	116.88
RCCA/HQIC	% in which $\hat{q} = q$	0.80	89.68	80.36
	% in which $\hat{q} > q$	0.00	0.10	0.06
	RFD	77.48	71.55	56.10
	TRMSFE	106.51	<b>94.51<sup>***</sup></b>	<b>100.97<sup>***</sup></b>
CCA/AIC	% in which $\hat{q} = q$	13.86	0.84	0.90
	% in which $\hat{q} > q$	85.88	99.16	99.10
	RFD	97.01	101.38	87.24
	TRMSFE	110.86	105.02	111.07
CCA/SIC	% in which $\hat{q} = q$	0.00	4.70	3.14
	% in which $\hat{q} > q$	0.00	0.00	0.00
	RFD	90.44	108.79	79.11
	TRMSFE	128.84	105.08	113.55
CCA/HQIC	% in which $\hat{q} = q$	1.38	<b>95.30<sup>***</sup></b>	<b>86.68<sup>***</sup></b>
	% in which $\hat{q} > q$	0.00	0.54	0.80
	RFD	101.64	106.23	86.54
	TRMSFE	109.40	95.41	101.85
RLS	RFD	73.35	<b>68.79<sup>***</sup></b>	63.12
	TRMSFE	116.66	118.79	125.85
OLS	TRMSFE	131.75	135.42	143.52

Note: RFD is the Frobenius distance between the estimated structural parameters and the true ones, relative to the Frobenius distance of OLS. The best result for each criterion and sample design is in bold. See the notes to Table 2 for the meanings of the other acronyms.

<sup>\*\*\*</sup> Indicates significance at the 1% level of the tests of equal TRMSFEs and RFDs, and of the McNemar test on the differences between the proportions of cases in which  $\hat{q} = q$  for the two best methods.

The Monte Carlo design is as follows. Three different processes are simulated according to the parameter estimates in the three samples of the empirical analysis. In particular, we simulate 20 time series accordingly to the following stationary VAR(2) process:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \epsilon_t.$$

In order to challenge the RCCA estimator, the coefficient matrices  $\Phi_1$  and  $\Phi_2$  are set equal to their estimates, obtained in the empirical analysis through CCA, using the SIC;  $\epsilon_t$  are i.i.d.  $N_n(0, S_{\epsilon\epsilon})$ , and  $S_{\epsilon\epsilon}$  is the covariance matrix of the RRR errors. Hence, the Monte Carlo analysis is designed explicitly to be unfair with respect to the RCCA approach.<sup>3</sup>

We generate systems of 20 variables (i.e.,  $N = 40$  predictors) with a number of observations of vector series  $y_t$  equal to  $T + 150$ , for  $T = 120, 240$ . The first 50 points are used as a burn-in period, the last 100 observations to compute the  $h$ -step-ahead forecast errors for  $h = 1$ , and the intermediate  $T$  observations to estimate the competing models.

The averages of RFD and TRMSFE and the percentages in which  $\hat{q} = q$  over 5000 replications are reported in Tables 9–10 for  $T = 120, 240$ .

<sup>3</sup> In an alternative Monte Carlo design, the parameters of the simulated VAR are those obtained in the empirical application through RCCA using the HQIC. The results, available upon request, are clearly more favourable for the RCCA method.

The results indicate that, in terms of RFD, the RCCA estimator is the best one, with the exception of the Pre-Great Moderation design when  $T = 120$ , where RLS delivers the best results. The correct number of factors is obtained more often by using CCA in combination with the HQIC. However, there is no strong connection between the exact choice of factors and predictive ability. Indeed, RCCA's forecasts outperform those of its multivariate competitors, although no method can beat the univariate forecasts with the full-sample and Great Moderation designs and when  $T = 120$ . As  $T$  increases, the results in terms of TRMSFE improve for all of the multivariate methods, but the winner is still RCCA, especially if it is associated with the HQIC for the choice of  $q$ .

## 5. Conclusions

Based on the work of Ledoit and Wolf (2003), we have proposed a method that allows for shrinking RRR towards the simplest univariate models, i.e., the white noise. The suggested method, named RCCA, can be seen as a small-sample correction for CCA, since the two methods are asymptotically equivalent.

Using both simulations and empirical examples, we have shown that RCCA outperforms well-known methods for macroeconomic forecasting when applied to medium VAR models. Moreover, our Monte Carlo study reveals that

**Table 10**  
Simulation results.

$N = 20, T = 240$				
Method/criterion	Statistics	Full sample $q = 5$	Pre-great moderation $q = 2$	Great moderation $q = 2$
RCCA/AIC	% in which $\hat{q} = q$	<b>66.12***</b>	55.22	59.16
	% in which $\hat{q} > q$	33.88	44.78	40.84
	RFD	81.34	78.98	65.19
	TRMSFE	<b>91.24***</b>	92.04	95.77
RCCA/SIC	% in which $\hat{q} = q$	0.00	87.90	79.84
	% in which $\hat{q} > q$	0.00	0.00	0.00
	RFD	89.21	82.08	67.01
	TRMSFE	101.08	91.78	96.82
RCCA/HQIC	% in which $\hat{q} = q$	55.86	<b>100.00</b>	<b>100.00</b>
	% in which $\hat{q} > q$	0.00	0.00	0.00
	RFD	<b>81.06***</b>	<b>78.51***</b>	<b>64.31***</b>
	TRMSFE	91.71	<b>90.91***</b>	<b>94.67***</b>
CCA/AIC	% in which $\hat{q} = q$	59.26	42.58	42.68
	% in which $\hat{q} > q$	40.74	57.42	57.32
	RFD	92.62	95.64	86.47
	TRMSFE	91.86	92.43	96.37
CCA/SIC	% in which $\hat{q} = q$	0.00	90.26	80.06
	% in which $\hat{q} > q$	0.00	0.00	0.00
	RFD	95.51	98.33	86.30
	TRMSFE	101.63	91.70	96.86
CCA/HQIC	% in which $\hat{q} = q$	54.60	<b>100.00</b>	<b>100.00</b>
	% in which $\hat{q} > q$	0.00	0.00	0.00
	RFD	92.09	95.22	85.20
	TRMSFE	92.46	91.00	94.84
RLS	RFD	85.83	80.97	74.81
	TRMSFE	98.69	103.28	107.76
OLS	TRMSFE	101.85	107.08	111.71

See the notes to Table 9.

RCCA provides more accurate estimates of the structural parameters than its competitors.

The above findings leave room for several developments of potential interest, such as the application of RCCA to the forecasting and analysis of medium dynamic systems with more elaborate forms of common autocorrelation features (see for example Cubadda, 2007) and cointegration (see for example Ahn, 1997). The latter extension is theoretically challenging, since it would require a systematic analysis of the properties of the shrinking approach when some of the variables are non-stationary. These issues are on our research agenda.

## Acknowledgments

Previous drafts of this paper were presented at the 32nd Annual International Symposium on Forecasting in Boston, the 6th CSDA International Conference on Computational and Financial Econometrics in Oviedo, the International Statistical Conference SIS 2013 in Brescia, and the 67th European Meeting of the Econometric Society in Gothenburg. We wish to thank the participants, as well as Alain Hecq, Marco Lippi, Marcella Corduas and two anonymous referees, for helpful comments and suggestions. The usual disclaimers apply.

## Appendix

In this appendix we provide proofs of Theorems 1 and 2 from Section 2. Unlike Sancetta (2008), we derive

our asymptotic results by keeping  $N$  fixed and letting  $T$  diverge.

### A.1. Asymptotic theory for the regularization term $\alpha^*$

**Proof of Theorem 1.** Let us first derive the asymptotics for

$$T\alpha^* = \frac{\sum_{i \neq j}^{N+n} \text{Var}(\sqrt{T}s_{ij})}{\sum_{i \neq j}^{N+n} [\text{Var}(s_{ij}) + \sigma_{ij}^2]}.$$

We assume, without loss of generality, that  $E(w_t) = 0$ . We can write

$$s_{ij} = \frac{1}{T} \sum_{t=1}^T \dot{w}_{it} \dot{w}_{jt},$$

where  $\dot{w}_{it} = w_{it} - \bar{w}_i$ , and  $\bar{w}_i = \frac{1}{T} \sum_t w_{it}$ ,  $\forall i$ .

Let us define  $\tilde{\sigma}_{ij}$

$$\tilde{\sigma}_{ij} = \frac{1}{T} \sum_{t=1}^T w_{it} w_{jt},$$

so that

$$\begin{aligned} \tilde{\sigma}_{ij} - s_{ij} &= \frac{1}{T} \sum_{t=1}^T w_{it} w_{jt} - \frac{1}{T} \sum_{t=1}^T \dot{w}_{it} \dot{w}_{jt} \\ &= \frac{1}{T} \sum_{t=1}^T w_{it} w_{jt} - \left( \frac{1}{T} \sum_{t=1}^T w_{it} w_{jt} - \bar{w}_i \bar{w}_j \right) \end{aligned}$$

and

$$\sqrt{T}(\tilde{\sigma}_{ij} - s_{ij}) = \sqrt{T}\tilde{w}_i\tilde{w}_j.$$

By applying the Law of Large Numbers and the Central Limit Theorem for stationary time series, we get

$$\begin{aligned}\lim_{T \rightarrow \infty} \sqrt{T}(\tilde{\sigma}_{ij} - s_{ij}) &= \lim_{T \rightarrow \infty} \frac{1}{\sqrt{T}} \sum_t w_{it} \frac{1}{T} \sum_t w_{jt} \\ &= \mathcal{N}(0, V)E(w_{jt}) = o_p(1).\end{aligned}$$

Given that the asymptotic expression for the covariances of  $\tilde{\sigma}_{ij}$  also applies to  $s_{ij}$ , we can study the limit behavior of  $\sqrt{T}(\tilde{\sigma}_{ij} - \sigma_{ij})$ .

Now, let us consider as a natural estimator for the matrix  $\Sigma(\tau) = E(w_t w'_{t+\tau})$  the quantities

$$\tilde{\Sigma}(\tau) = \frac{1}{T} \sum_{t=1}^T w_t w'_{t+\tau}.$$

When we have  $T$  observations, we are able only to form

$$\hat{\Sigma}(\tau) = \frac{1}{T-\tau} \sum_{t=1}^{T-\tau} w_t w'_{t+\tau}, \quad \tau > 0.$$

Since the fourth order cumulants of  $\varepsilon_t$  are finite, the fourth order cumulants of  $y_t$  and  $w_t$  are too; then,  $T^{-1}\{w_{T-j} w_{T+n-j}\}$  converges in mean square to zero, so that all convergence properties of  $\tilde{\Sigma}(\tau)$  apply to  $\hat{\Sigma}(\tau)$  as well.

We find the following expression for the variance of  $\tilde{\sigma}_{ij}$ :

$$\begin{aligned}\text{Var}[\tilde{\sigma}_{ij}] &= \frac{1}{T^2} \text{Var} \left( \sum_t w_{it} w_{jt} \right) \\ &= \frac{1}{T} \gamma_{ij}(0) + \frac{1}{T} \sum_{\tau \neq 0, \tau = -T+1}^{T-1} \left( 1 - \frac{|\tau|}{T} \right) \gamma_{ij}(\tau) \\ &= \frac{1}{T} \sum_{\tau = -T+1}^{T-1} \left( 1 - \frac{|\tau|}{T} \right) \gamma_{ij}(\tau),\end{aligned}$$

where  $\gamma_{ij}(\tau) = \text{Cov}(w_{it} w_{jt}, w_{i,t-\tau} w_{j,t-\tau})$ ,  $\tau = 0, \dots, T-1$ .

Let us denote by  $\kappa_{ijkl}(t_1, t_2, t_3, t_4)$  the joint cumulants of  $w_{it1}, w_{jt2}, w_{kt3}, w_{lt4}$ . Then, we have that

$$\begin{aligned}\text{Cov}[w_{it1} w_{jt2} w_{kt3} w_{lt4}] &= E[w_{it1} w_{jt2} w_{kt3} w_{lt4}] \\ &\quad - \sigma_{ij}(t_2 - t_1) \sigma_{kl}(t_4 - t_3) \\ &= \sigma_{ik}(t_3 - t_1) \sigma_{jl}(t_4 - t_2) \\ &\quad + \sigma_{il}(t_4 - t_1) \sigma_{jk}(t_3 - t_2) \\ &\quad + \kappa_{ijkl}(t_1, t_2, t_3, t_4),\end{aligned}$$

where  $\sigma_{ij}(t_2 - t_1) = E[w_{it1} w_{jt2}]$ . In this way, we get the following expression for the covariance between  $\tilde{\sigma}_{ij}(\tau)$  and  $\tilde{\sigma}_{kl}(\tau + k)$  (Hannan, 1970, p. 209):

$$\begin{aligned}\text{Cov}[\tilde{\sigma}_{ij}(\tau) \tilde{\sigma}_{kl}(\tau + k)] \\ = \frac{1}{T} \sum_{u=-T+1}^{T-1} \left( 1 - \frac{|u|}{T} \right) \left\{ \sigma_{ik}(u) \sigma_{jl}(u + k) \right.\end{aligned}$$

$$\begin{aligned}&+ \sigma_{il}(u + \tau + k) \sigma_{jk}(u - \tau) + \left( \frac{1}{T - |u|} \right) \\ &\times \sum_v' \kappa_{ijkl}(v, v + \tau, v + u, v + u + \tau + k) \Big\}, \quad (1)\end{aligned}$$

where the index  $v$  for the summation  $\sum'$  is such that  $1 \leq v + u \leq T$ .

As  $T$  diverges, the covariance in Eq. (1) converges to zero. Indeed, since the fourth order cumulants of  $w_t$  are finite, and, given that  $w_t$  is a linear process with absolutely summable autocovariances  $\sigma_{ij}(\tau)$ , we have that

$$\sum_q \sum_r \sum_s |\kappa_{ijkl}(0, q, r, s)| < \infty,$$

and the result follows.

Let us now focus on the limit behavior of  $\sqrt{T}\tilde{\sigma}_{ij}$ . From Hannan (1970, Theorem 14, p. 228), we know that if  $w_t$  is generated by a linear process, with  $\varepsilon_t$  having the finite fourth-cumulant function  $\kappa_{ijkl}$ , then the quantities

$$\begin{aligned}\sqrt{T}(\tilde{\sigma}_{ij}(\tau_u) - \sigma_{ij}(\tau_u)); \\ i, j = 1, \dots, n, \quad u = 1, \dots, q, \quad q \geq 0,\end{aligned}$$

where  $\sigma_{ij}(\tau_u) = E(w_{it} w_{jt+\tau_u})$ , have an asymptotic Gaussian distribution with a zero mean and variance equal to the asymptotic variance of  $\tilde{\sigma}_{ij}$

$$\pi_{ij} = LRV(v_t^{ij}) = \lim_{T \rightarrow \infty} T \times \text{Var}(\tilde{\sigma}_{ij}),$$

given that  $\sqrt{T}(s_{ij} - \tilde{\sigma}_{ij}) = o(1)$ ,  $\sum_{i \neq j} \text{Var}(\sqrt{T}s_{ij}) \rightarrow \pi < \infty$ .

Since we have previously shown that  $\text{Var}(s_{ij})$  is  $O(\frac{1}{T})$ , then  $\sum_{i \neq j} \text{Var}(s_{ij})$  is  $O(\frac{1}{T})$  as well, and finally, we get

$$\alpha^* = \frac{1}{T} \left( \frac{\pi}{\gamma} \right) + O(T^{-2}),$$

where  $\sum_{i \neq j} \sigma_{ij}^2 \equiv \gamma$ . ■

## A.2. A consistent estimator of $\lim_T T\alpha^*$

**Proof of Theorem 2.** In order to get a consistent estimator of the optimal shrinkage constant, we need a consistent estimator of  $\pi_{ij}$  and  $\sigma_{ij}^2$  for  $\forall i, j$ .

We start by showing the ergodicity in the second moments of  $w_t$ . Given Proposition 10.2(b), (d) of Hamilton (1994), the sample variance in a vector autoregressive model, in a sample of dimension  $T$ , is a consistent estimator of the population variance. Given that  $s_{ij} = \frac{1}{T} \sum_t \dot{w}_{it} \dot{w}_{jt} \xrightarrow{P} E(\dot{w}_{it} \dot{w}_{jt}) = \sigma_{ij}$ , and given that the square is a continuous transformation, we can safely say that  $s_{ij}^2$  is a consistent estimator of  $\sigma_{ij}^2$ .

Now, we need to estimate consistently the sum of the asymptotic variances of  $s_{ij}$ , for  $i \neq j = 1, \dots, N + n$ . To this end, let us write down the long run variance of  $v_t^{ij}$  as



follows:

$$\begin{aligned} \text{LRV}(v_t^{ij}) &= \lim_{T \rightarrow \infty} \sqrt{T} \text{Var}(s_{ij}) = \lim_{T \rightarrow \infty} \text{Var} \left( \frac{1}{\sqrt{T}} \sum_t \dot{w}_{it} \dot{w}_{jt} \right) \\ &= \lim_{T \rightarrow \infty} \frac{1}{T} T \gamma_{ij}(0) + \sum_{\tau=-T+1, \tau \neq 0}^{T-1} \left( 1 - \frac{|\tau|}{T} \right) \gamma_{ij}(\tau) \\ &= \sum_{\tau=-\infty}^{+\infty} \gamma_{ij}(\tau). \end{aligned}$$

As a natural estimator for  $\gamma_{ij}(\tau)$  ( $\tau = 0, 1, 2, \dots$ ), we can use the sample autocovariance of  $w_{it} w_{jt}$ :

$$\hat{\gamma}_{ij}(\tau) = \frac{1}{T} \sum_t [\dot{w}_{it} \dot{w}_{jt} - s_{ij}] [\dot{w}_{it-\tau} \dot{w}_{jt-\tau} - s_{ij}].$$

In order for this estimator to be consistent, we need  $\frac{1}{T} \sum_t w_{it} w_{jt} w_{it-\tau} w_{jt-\tau} \xrightarrow{p} E(w_{it} w_{jt} w_{it-\tau} w_{jt-\tau})$ .

Recall that  $v_t^{ij} \equiv w_{it} w_{jt}$ . Then, we can resort to Theorem 2 of Hannan (1970, p. 203) to show that

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T v_t^{ij} v_{t-\tau}^{ij} \xrightarrow{a.s.} E(v_t^{ij} v_{t-\tau}^{ij}).$$

Indeed, since  $w_t$  has a finite fourth cumulant function,  $E|w_{it1} w_{jt2} w_{it3} w_{jt4}| < \infty$ , which implies that the conditions of Hannan's theorem are satisfied, and

$$(1/T) \sum_{t=1}^T w_{it} w_{jt} w_{it-\tau} w_{jt-\tau} - E(w_{it} w_{jt} w_{it-\tau} w_{jt-\tau}) \rightarrow 0.$$

Hence,

$$\begin{aligned} \hat{\gamma}_{ij}(\tau) &= \frac{1}{T} \sum_t [(\dot{w}_{it} \dot{w}_{jt} - s_{ij})(\dot{w}_{it-\tau} \dot{w}_{jt-\tau} - s_{ij})] \\ &\rightarrow E[\dot{w}_{it} \dot{w}_{jt} \dot{w}_{it-\tau} \dot{w}_{jt-\tau} - \sigma_{ij}^2], \end{aligned}$$

where

$$\hat{\gamma}_{ij}(\tau) \rightarrow \gamma_{ij}(\tau) \quad \tau = 0, 1, \dots, T-1.$$

In order to estimate the long run variance of  $v_t^{ij} \forall i, j$  consistently, we have to estimate the elements of the whole matrix  $\Gamma = \Gamma_0 + \sum_{\tau=1}^{\infty} (\Gamma_{\tau} + \Gamma'_{\tau})$ , where

$$\Gamma_{\tau} = \begin{bmatrix} \gamma_{11}(\tau) & \cdots & \gamma_{1n}(\tau) \\ \gamma_{21}(\tau) & \cdots & \gamma_{2n}(\tau) \\ \vdots & \ddots & \vdots \\ \gamma_{n1}(\tau) & \cdots & \gamma_{nn}(\tau) \end{bmatrix}.$$

For estimating the matrix  $\Gamma$ , we make the following assumption:

**Assumption 5.** Let  $\hat{\Gamma} = \hat{\Gamma}_0 + \sum_{\tau=1}^{T-1} w(\tau/M)(\hat{\Gamma}_{\tau} + \hat{\Gamma}'_{\tau})$  and let  $\hat{\Gamma}_{\tau}$  be the natural estimator of  $\Gamma_{\tau}$  for  $\tau = 0, 1, \dots, T-1$ . The kernel function is a twice continuously differentiable even function with  $w(0) = 1$ ,  $w'(0) = 0$ ,  $w''(0) = 0$ ,  $w(1+\delta) = 0$  for  $\delta \geq 0$ , and either (a)  $w(x)/(1-|x|)^2$  converges to a constant or (b)  $w(x) = O(x^{-2})$ , as  $|x| \rightarrow 1$ . For the bandwidth, it is assumed that  $M = O(T^m)$ , where  $m \in (0, 1/2)$ .

Under this assumption, as was shown by Phillips (1995),  $\hat{\Gamma}$  is a consistent estimator of  $\Gamma$ , so that

$$T\hat{\alpha}^* = \frac{\sum_{i \neq j} \widehat{\text{LRV}}(v_t^{ij})}{\sum_{i \neq j} [T^{-1} \widehat{\text{LRV}}(v_t^{ij}) + s_{ij}^2]},$$

where  $\widehat{\text{LRV}}(v_t^{ij}) = \hat{\gamma}_{ij}(0) + 2 \sum_{\tau=1}^{T-1} w(\tau/M) \hat{\gamma}_{ij}(\tau)$  is a consistent estimator of  $\lim_T T\alpha^*$ . ■

## References

- Ahn, S. K. (1997). Inference of vector autoregressive models with cointegration and scalar components. *Journal of the American Statistical Association*, 92, 350–356.
- Anderson, T. W. (1984). *An introduction to multivariate statistical analysis* (2nd ed.). New York: John Wiley & Sons.
- Banbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian VARs. *Journal of Applied Econometrics*, 25, 71–92.
- Carriero, A., Kapetanios, G., & Marcellino, M. (2011). Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, 26, 735–761.
- Cubadda, G. (2007). A unifying framework for analyzing common cyclical features in cointegrated time series. *Computational Statistics and Data Analysis*, 52, 896–906.
- Cubadda, G., & Guardabascio, B. (2012). A medium-N approach to macroeconomic forecasting. *Economic Modelling*, 29, 1099–1105.
- Cubadda, G., & Hecq, A. (2011). Testing for common autocorrelation in data rich environments. *Journal of Forecasting*, 30, 325–335.
- Cubadda, G., Hecq, A., & Palm, F. C. (2009). Studying interactions without multivariate modeling. *Journal of Econometrics*, 148, 25–35.
- Diebold, F. X., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics*, 13, 253–263.
- Engle, R. F., & Kozicki, S. (1993). Testing for common features (with comments). *Journal of Business and Economic Statistics*, 11, 369–395.
- Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics*, 82, 540–554.
- Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2005). The generalized dynamic factor model: one-sided estimation and forecasting. *Journal of the American Statistical Association*, 100, 830–840.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton: Princeton University Press.
- Hannan, E. J. (1970). *Multiple time series*. Wiley series in applied statistics, New York.
- Harvey, D., Leybourne, S., & Newbold, P. (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting*, 13, 281–291.
- Koop, G. M. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28, 177–203.
- Ledoit, O., & Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, 10, 603–621.
- Litterman, R. (1986). Forecasting with Bayesian vector autoregressions: five years of experience. *Journal of Business and Economic Statistics*, 4, 25–38.
- Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3), 703–708.
- Phillips, P. C. B. (1995). Fully modified least squares and vector autoregressions. *Econometrica*, 63(5), 1023–1078.
- Reinsel, G., & Velu, R. P. (1998). *Multivariate reduced-rank regression, theory and applications*. Springer-Verlag.
- Sancetta, A. (2008). Sample covariance shrinkage for high dimensional dependent data. *Journal of Multivariate Analysis*, 99, 949–967.
- Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In J. Neyman (Ed.), *Proceedings of the third Berkeley symposium on mathematical and statistical probability*, Vol. 1 (pp. 197–206). Berkeley: University of California.
- Stock, J. H., & Watson, M. W. (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97, 1167–1179.
- Stock, J. H., & Watson, M. W. (2002b). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20, 147–162.

Vahid, F., & Engle, R. F. (1993). Common trends and common cycles. *Journal of Applied Econometrics*, 8, 341–360.

**Emmanuela Bernardini** is a researcher at the Bank of Italy in Rome. Her main research interests are macroeconomic forecasting and time series analysis.

**Gianluca Cubadda** is Professor of Economic Statistics at the University of Rome “Tor Vergata”, Italy. His main research interests are time series analysis and econometrics. His articles have appeared in various journals, including *Computational Statistics & Data Analysis*, *Econometric Reviews*, *Economic Modelling*, *Economics Letters*, *Empirical Economics*, *Journal of Applied Econometrics*, *Journal of Econometrics*, *Journal of Forecasting*, *Journal of Time Series Analysis*, *Macroeconomic Dynamics*, and *Oxford Bulletin of Economics and Statistics*.