Dimension Reduction for High Dimensional Vector Autoregressive Models

Gianluca Cubadda   Alain Hecq

Università di Roma "Tor Vergata"

Maastricht University

February 24, 2022

Introduction

- The VAR is a standard tool for investigating multivariate time series in economics and finance (forecasting, IRF, Granger causality).
- However, it is subject to the curse of dimensionality.
- Several solutions were proposed, following either
  - the dimension reduction approach (e.g. reduced-rank VAR, FAVAR)
  - the regularization paradigm (e.g. BVAR, Lasso).
- This paper is on the first type of solution.

Continued...

- Lam *et al.* (2011) and Lam and Yao (2012) showed how to decompose a large multivariate time series Y_t into two parts:
    - a linear function of a small scale dynamic component x_t and
    - a static component ε_t that is unpredictable from the past.
- No loss of information.
- We go a step further...

Continued...

- We assume that Y_t is generated by a large VAR and provide the conditions under which the dynamic component x_t is generated by a small scale VAR.
- We show that it is required that the large VAR model of series Y_t is endowed with both
  - the serial correlation common feature (Engle and Kozicki, 1993)
  - and an index structure (Reinsel, 1983).

Continued...

- Based on the eigenanalysis (Lam *et al.* (2011), Lam and Yao (2012) as well as IC) , we provide statistical tools to verify whether there exists in series Y_t a dynamic component x_t that is generated by such a small scale VAR model and to estimate the associated parameters.
- Prety cool:
  - This allows to obtain forecasts of the large time series Y_t using those of the dynamic component x_t only,
  - to recover the structural shocks from the reduced form shocks of the dynamic component only.
- In this paper: Monte Carlo analysis and by an empirical applications on US macroeconomic series.

Model representation

- The n-vector time series Y_t is generated by the stationary VAR(p)

$$Y_t = \sum_{j=1}^{p} \Phi_j Y_{t-j} + u_t,$$

  where $t = 1, ..., T$, and u_t is an n-vector of innovations.
- **Condition 1** (Serial Correlation Common Feature, SCCF): For $\Phi = [\Phi_1, ..., \Phi_p]'$ it holds that $\Phi' = \bar{A}\bar{\Omega}'$, where $\bar{A}$ is a full rank $n \times r$ ($r < n$) matrix and $\bar{\Omega} = [\bar{\omega}_1', ..., \bar{\omega}_p']'$ is a full rank $np \times r$ matrix.
- SCCF: Common left null space (reduced rank) in all Φ_j matrices

$$Y_t = \bar{A}\bar{\omega}_1' Y_{t-1} + ... + \bar{A}\bar{\omega}_p' Y_{t-p} + u_t$$

Continued...

- We can always use the equivalent factorization $\Phi' = A\Omega'$, where $A = \bar{A}(\bar{A}'\bar{A})^{-1/2}$ and $\Omega = \bar{\Omega}(\bar{A}'\bar{A})^{1/2}$, we assume without loss of generality that A is an orthogonal matrix with $A'A = I_r$
- Under Condition 1 we can decompose series Y_t using $I_n = AA' + A_\perp A'_\perp$ as

$$Y_t = Ax_t + \varepsilon_t,$$

  where $x_t = A'Y_t$, $\varepsilon_t = A_\perp A'_\perp u_t$,
- such that $\mathrm{E}(Y_{t+k}|\mathcal{F}_t) = A\mathrm{E}(x_{t+k}|\mathcal{F}_t)$ for $k > 0$, and $\mathcal{F}_t$ is the natural filtration of process Y_t.

Continued...

- Notice that under Condition 1 (SCCF) the dynamic component x_t is generated as

$$x_t = \sum_{j=1}^{p} \omega_j' Y_{t-j} + \underbrace{A' u_t}_{\xi_t},$$

- No loss of information (vs. e.g. PCA).
- We could have stop here. This is what inter alia my coauthors and I do for many years (common features, reduced rank) but that is not a small-scale model since the number of parameters grows with n (although not with n^2 as in the unrestricted VAR).

Continued...

- It wouldn't matter much for small n but it is a challenge for high dimensional systems.
- We add a right null space condition.
- **Condition 2**: For any $j = 1, ..., p$ it holds that $\omega_j = A\alpha_j'$, where α_j is a $r \times r$ matrix. In words, the lags of the same linear combinations of Y_t that are unpredictable from the past are also irrelevant predictors of the dynamic component x_t.

Continued...

- Under Condition 2, the dynamic component x_t is then generated as

$$x_t = \sum_{j=1}^{p} \omega_j' Y_{t-j} + \xi_t = \sum_{j=1}^{p} \alpha_j x_{t-j} + \xi_t,$$

- Under Condition 1 and 2, the VAR model of series Y_t can be rewritten as

$$Y_t = \sum_{j=1}^{p} A\alpha_j \underbrace{A' Y_{t-j}}_{x_{t-j}} + u_t. \qquad (1)$$

- Model (1) is interesting since it combines the features of the reduced-rank VAR model with those of the multivariate autoregressive index (MAI) model proposed by Reinsel (1983)

Continued...

- **Remark 1**: The reduced-rank VAR and the MAI have been considered separately in the literature, whereas Conditions 1 and 2 reveal that the combination of the two models allows for an important dimension reduction in large VARs.
- In what follows, we call Model (1) as the dimension-reducible VAR model (DRVAR).

- **Remark 2**: If we invert the polynomial coefficient matrix of the small VAR we can write the Wold representation of series Y_t as

$$Y_t = A\gamma(L)\xi_t + \varepsilon_t, \tag{2}$$

where $\gamma(L)^{-1} = I_n - \sum_{j=1}^{p} \alpha_j L^j$. Moreover, by linearly projecting ε_t on ξ_t, we can decompose the static component as $\varepsilon_t = \rho\xi_t + \nu_t$, where $\rho = A_\perp A'_\perp \Sigma_u A(A'\Sigma_u A)^{-1}$, and then rewrite Equation (2) as

$$Y_t = C(L)\xi_t + \nu_t \equiv \chi_t + \nu_t, \tag{3}$$

where $C_0 = A + \rho$ and $C_j = A\gamma_j$ for $j > 0$.

- Representation (3) highlights that the system dynamics are entirely generated by errors ξ_t. Hence, we label χ_t as the common component of Y_t and ν_t as the ignorable errors, as we assume the latter are not endowed with a structural interpretation. Since the errors ξ_t and ν_t are uncorrelated at any lead and lag, it is then possible to recover the structural shocks solely from the reduced form errors ξ_t of the common component χ_t.

Continued. . .

- Any of the procedures that are commonly employed in structural VAR analysis may be used. For instance, one may obtain the structural shocks as $u_t = C^{-1}D\xi_t$ and the impulse response functions from $\Psi(L) = C(L)D^{-1}C$, where D is the matrix formed by the first r rows of C_0 and $C = \mathrm{Chol}(DA'\Sigma_u AD')$. Notice that the first r rows of $\Psi(0)$, being equal to C, form a lower triangular matrix, thus allowing for the usual interpretation of the structural shocks u_t as long as the s $(s \leq r)$ variables of interest are placed and properly ordered in the first s elements of Y_t.
- It is always possible to identify the structural shocks from the reduced form errors of the large VAR. However, the advantage of the structural DRVAR is that it requires to identify r shocks only instead of n of them. Hence, the number of structural shocks is (much) smaller than the number of variables, as it is typical in both structural factor models (see e.g. Forni *et al*, 2009) and dynamic stochastic general equilibrium models (see e.g. Fernández-Villaverde *et al.*, 2016).

Statistical inference (reduced rank in large systems)

- Let us indicate with $\hat{V}_q$ the matrix formed by the eigenvectors associated with the q $(\leq n)$ largest eigenvalues of the matrix

$$\hat{M} = \sum_{j=1}^{p_0} \hat{\Sigma}_y(j)\hat{\Sigma}_y(j)',$$

  where $\hat{\Sigma}_y(j)$ denotes the sample autocovariance matrix of Y_t at lag j.
- Under regularity conditions that are compatible with our Conditions, $\hat{V}_r$ estimates A (up to an orthonormal transformation) with the standard $\sqrt{T}$ rate when r is fixed, $n, T \to \infty$, and the factors are strong, i.e. $\bar{a}_i'\bar{a}_i = O(n)$ for $i = 1, ..., r$, where $\bar{A} = [\bar{a}_1, ..., \bar{a}_r]$ (see Theorem 1 of Lam *et al.*, 2011).

Continued...

- Moreover, Lam and Yao (2012) provided the following consistent estimator of r

$$\hat{r} = \arg \min_{i=1,..R} \left\{ \hat{\lambda}_{i+1} / \hat{\lambda}_i \right\},$$

where R is a constant such that $r < R < n$ and $\hat{\lambda}_i$ is the $i-$th largest eigenvalue of matrix $\hat{M}$.

- However, this procedure takes into account Condition 1 only and not Condition 2.

Continued...

- In order to estimate the parameters of the DRVAR assuming that r is known and having fixed A equal to $\hat{V}_r$, let's rewrite the model in its matrix form

$$Y = Z\alpha A' + u,$$

where $Y = \left[y_{p+1}, ..., y_T\right]'$, $u = \left[u_{p+1}, ..., u_t\right]'$, $z_t = \left[x_t', ..., x_{t-p+1}'\right]'$, and $Z = \left[z_p, ..., z_{T-1}\right]'$. Then the OLS estimator of $\mathrm{Vec}(\alpha)$ is:

$$\mathrm{Vec}(\hat{\alpha}) = [A' \otimes (Z'Z)^{-1} Z']\mathrm{Vec}(Y).$$

which is equivalent to OLS on the small-scale VAR of x_t.

Continued...

- The GLS estimator of $\mathrm{Vec}(\alpha)$ takes instead the following form:

$$\mathrm{Vec}(\tilde{\alpha}) = \left[\left(A'\Sigma_u^{-1} A \right)^{-1} A'\Sigma_u^{-1} \otimes \left(Z'Z \right)^{-1} Z' \right] \mathrm{Vec}(Y). \quad (4)$$

  The relation, in terms of efficiency, between these estimators is provided in the following theorem.
- **Theorem**: Assuming that A and Σ_u are known, the GLS Estimator of $\mathrm{Vec}(\alpha)$ has a MSE matrix, conditionally on Z, that is not larger that the one of the OLS estimator. The two estimators have the same MSE matrix when $A'u_t$ and $A'_{\perp}u_t$ are not correlated.
- We offer a switching algorithm to compute a FGLS estimator of $\mathrm{Vec}(\alpha)$. A practical issue that arises when n is large is that the estimate of matrix Σ_u is (nearly) singular. We solve this problem by ignoring the error cross-correlations, i.e. we use a diagonal matrix Δ_u with the same diagonal as Σ_u in place of Σ_u itself.

Continued...

- In order to identify the number of factors r, for $q = 1, ..., R(\geq r)$ we estimate either by OLS or FGLS the models

$$Y_t = \sum_{j=1}^{p} \hat{V}_q \alpha_{j,q} \hat{V}_q' Y_{t-j} + u_t(q),$$

  where $\alpha_{j,q}$ is a $q \times q$ matrix for $j = 1, ..., p$, and estimate r as the index $\widehat{r}$ that minimizes an information criterion, where the measure of fit is $\operatorname{trace}(\ln(\Delta_u))$, and the overall number of parameters is $k = nq + (p-1)q^2$.
- **Proposition**: Under conditions such that OLS and FGLS estimate the DRVAR parameters (up to an orthonormal transformation) with the standard $\sqrt{T}$ rate as $n, T \to \infty$, and assuming that $\gamma = \lim_{n\to\infty} \left(\prod_{i=1}^{n} \sigma_i^2 \right)^{1/n}$ exists, where $\operatorname{diag}[\sigma_1^2, ..., \sigma_n^2] = \Delta_u$, the BIC and HQIC provide weakly consistent estimators for the number of dynamic components r but not for the overall number of the DRVAR parameters k.

Monte Carlo analysis

- We consider the following n-dimensional stationary VAR(2) process

$$Y_t = \underbrace{\bar{A}\mathrm{diag}(\delta_1)\bar{A}^+}_{\Phi_1} Y_{t-1} + \underbrace{\bar{A}\mathrm{diag}(\delta_2)\bar{A}^+}_{\Phi_2} Y_{t-2} + u_t, \quad (5)$$

  where $\bar{A}$ is a $n \times r$ matrix such that its columns are generated by r i.i.d. $\mathrm{N}_n(0, I_n)$, $\bar{A}^+ = (\bar{A}'\bar{A})^{-1}\bar{A}'$ is the Moore–Penrose pseudo inverse of the matrix $\bar{A}$, $\delta_1 = 2\mathrm{diag}(m)\cos(\omega)$, m is a $r-$vector drawn from a $\mathrm{U}_r[0.3, 0.9]$, ω is a $r-$vector drawn from a $\mathrm{U}_r[0, \pi]$, $\delta_2 = -m^2$, and u_t are i.i.d. $\mathrm{N}_n(0, \Sigma_u)$.
- The covariance matrix of the VAR errors Σ_u is not diagonal. This allows us to evaluate the performances of both the FGLS estimator and the information criteria when $\ln(\det(\Sigma_u))$ is not equal to $\mathrm{trace}(\ln(\Delta_u))$.

Continued...

- We generate 1000 VAR(2) systems of $n = 150, 300, 600, 1200$ variables with $r = 3, 9$ dynamic components, and $T = \frac{1}{2}n, n, 1.5n$. We look at two statistics:(i) the percentage with which the true number of dynamic components r is correctly identified using both the test by Lam and Yao (2012) (LY) and by the usual information criteria (IC); (ii) the Frobenius distance between the estimates of $\Phi = [\Phi_1', \Phi_2']'$ and the true ones relative to the Frobenius norm of Φ.
- The main results are: (i) all the methods perform better as both n and T get larger; (ii) OLS and FGLS perform quite similarly; (iii) HQIC identifies the correct model better than the competitors but in 3 cases where BIC performs best; (iv) LY [AIC] systematically underestimates [overestimates] the true r; (iv) models identified by the BIC [AIC] provide estimates of Φ that are more [less] accurate than those obtained by the other criteria over all the settings.

Continued...

MC results, $r = 3$, $N = 150, 300$, OLS estimator

T/n		$n = 150$			$n = 300$		
		$\%\widehat{r} = 3$	$\overline{\widehat{r}}$	RFD	$\%\widehat{r} = 3$	$\overline{\widehat{r}}$	RFD
$T = \frac{n}{2}$	LY	40.8	2.100	-	53.7	2.389	-
	BIC	67.2	2.605	39.46	78.6	2.748	27.55
	HQ	72.6	3.094	74.88	86.6	2.993	42.44
	AIC	29.1	5.983	274.27	42.5	4.801	192.21
$T = n$	LY	52.8	2.311	-	62.3	2.493	-
	BIC	80.3	2.784	26.94	90.0	2.898	18.56
	HQ	86.2	3.002	38.11	92.7	2.992	23.72
	AIC	45.1	4.633	144.22	51.1	4.444	122.71
$T = 1.5n$	LY	55.8	2.328	-	64.3	2.555	-
	BIC	83.6	2.824	22.15	91.8	2.917	14.916
	HQ	91.6	3.005	27.93	95.0	3.003	19.140
	AIC	53.5	4.245	105.08	52.8	4.217	99.841

Continued...

MC results, $r = 3$, $n = 600, 1200$, OLS estimator

T/n		$n = 150$			$n = 300$		
		$\%\widehat{r} = 9$	$\overline{\widehat{r}}$	RFD	$\%\widehat{r} = 9$	$\overline{\widehat{r}}$	RFD
$T = \frac{n}{2}$	LY	59.9	2.560	-	65.7	2.734	-
	BIC	89.6	2.882	18.39	94.0	2.936	12.89
	HQ	92.5	3.017	29.07	95.5	3.003	17.84
	AIC	51.5	4.286	147.14	52.8	4.171	137.20
$T = n$	LY	67.7	2.683	-	73.2	2.815	-
	BIC	94.0	2.944	13.06	97.6	2.976	8.85
	HQ	96.8	3.009	16.31	97.4	3.010	12.36
	AIC	52.2	4.167	110.85	52.0	4.176	112.70
$T = 1.5n$	LY	73.5	2.755	-	77.5	2.819	-
	BIC	96.5	2.964	10.54	99.1	2.991	7.27
	HQ	96.4	3.021	15.08	97.7	3.025	11.72
	AIC	50.8	4.162	98.16	52.3	4.127	94.88

Continued...

MC results, $r = 9$, $N = 150, 300$, OLS estimator

$T\,/\,n$		$n = 150$			$n = 300$		
		$\%\widehat{r} = 9$	$\widehat{r}$	RFD	$\%\widehat{r} = 9$	$\widehat{r}$	RFD
$T = \frac{n}{2}$	LY	22.7	3.639	-	39.2	4.753	-
	BIC	44.7	7.781	59.54	50.0	8.135	38.57
	HQIC	63.1	9.262	109.20	85.7	8.946	48.29
	AIC	18.7	10.405	202.09	35.2	10.059	153.15
$T = n$	LY	34.5	4.443	-	47.1	5.389	-
	BIC	47.9	8.108	39.41	61.4	8.459	27.38
	HQ	83.0	8.928	46.82	91.4	8.958	30.55
	AIC	41.3	9.911	115.60	45.5	9.836	100.99
$T = 1.5n$	LY	42.7	5.075	-	56.8	6.019	-
	BIC	59.8	8.386	32.17	74.3	8.681	22.04
	HQIC	90.2	8.940	34.47	94.8	8.974	23.45
	AIC	48.9	9.765	88.17	51.4	9.729	78.47

Continued...

MC results, $r = 9$, $n = 600, 1200$, OLS estimator

$T\,/\,n$		$n = 600$			$n = 1200$		
	IC	$\%\widehat{r} = 9$	$\widehat{r}$	RFD	$\%\widehat{r} = 9$	$\widehat{r}$	RFD
$T = \frac{n}{2}$	LY	51.1	5.534	-	59.5	6.162	-
	BIC	61.1	8.442	27.03	73.3	8.676	18.96
	HQIC	91.7	8.953	30.69	96.1	8.970	19.98
	AIC	41.6	9.893	131.91	46.9	9.800	119.15
$T = n$	LY	64.4	6.540	-	74.2	7.164	-
	BIC	76.2	8.719	18.83	88.9	8.873	13.16
	HQIC	96.8	8.984	19.85	99.5	8.998	13.47
	AIC	48.8	9.760	90.40	45.6	9.818	93.59
$T = 1.5n$	LY	64.9	6.559	-	75.7	7.285	-
	BIC	83.1	8.809	15.25	93.1	8.927	10.58
	HQIC	97.9	8.986	15.54	99.5	9.001	10.97
	AIC	51.1	9.746	74.52	48.0	9.800	76.15

Co-movements in quarterly US time series

- We use $n = 211$ variables with $T = 242$ quarterly observations (1959Q3-2019Q4) from the FRED-QD database, to which we added the total factor productivity time series corrected for utilization produced by Fernald (2012).
- Series are first stationarized then corrected for outliers and finally demeaned and standardized.
- We fix $p_0 = 5$, a rather typical lag length of a VAR model for quarterly data, and $R = 14$ as the upper bound of the dimension of the dynamic component.
- In order to determine the largest VAR order, we use the traditional information criteria to estimate the lag length in a VAR model for series $\hat{V}_R' Y_t$. We get $p = 1$ according to the BIC, $p = 2$ according to the HQIC, and $p = 4$ according to the AIC. Consequently, we consider successively $p = 1, ..., 4$ lags when estimating r through the information criteria.

Continued...

- We estimate a VAR(2) with $r = 8$, which is the indication coming from the BIC [HQIC] when it is used to determine r [p].
- We compute two statistics in order to evaluate the fit of the model to the data. First, we consider the coefficients of determination of each element of Y_t as obtained by the reduced form DRVAR. Second, we compute the squared correlation coefficients between each element of Y_t and its counterpart in the common component χ_t. We denote the former statistic as $R^2_{Y,Z}$ and the latter as $R^2_{Y,\Xi}$. It is easy to see that $R^2_{Y,\Xi} \geq R^2_{Y,Z}$.
- Whereas $R^2_{Y,Z}$ has the usual interpretation in terms of predictability, $R^2_{Y,\Xi}$ indicates the fraction of the variability of each element of Y_t that is explained by a linear projection on the present and past values of the dynamic errors ξ_t. Hence, $R^2_{Y,\Xi}$ measures the importance of the common component χ_t in the variability of each series.

Continued...

- Based on the FGLS estimates of the coefficients of the DRVAR model, we report the averages as well as the quartiles of the empirical distributions of both $R^2_{Y,Z}$ and $R^2_{Y,\Xi}$. in the followimg table

Table: Average and quartiles of the measures of fit

	Mean	Q_1	Q_2	Q_3
$R^2_{Y,Z}$	0.30	0.11	0.25	0.47
$R^2_{Y,\Xi}$	0.53	0.34	0.56	0.73

- As expected, the estimates of $R^2_{Y,\Xi}$ are considerably larger than those of $R^2_{Y,Z}$.

Continued...

- Moreover, we report the estimates of both $R^2_{Y,Z}$ and $R^2_{Y,\Xi}$ for nine key macroeconomic variables: Gross Domestic Product (GDP), Consumption (Con), Investment (Inv), Unemployment Rate, (UR), Worked Hours (Hours), Inflation Rate (Inf), Interest Rate (IR), Labor Productivity, (LP), and Total Factor Productivity (TFP):

Table: Measures of fit for 9 key aggregate variables

	GDP	Con	Inv	UR	Hours	Inf	IR	LP	TFP
$R^2_{Y,Z}$	0.40	0.38	0.40	0.64	0.61	0.22	0.28	0.20	0.10
$R^2_{Y,\Xi}$	0.83	0.71	0.73	0.89	0.85	0.90	0.68	0.66	0.37

- Remarkably, the role of the common component of TFP is smaller than the one of the other key variables. This finding may reflect the partial exogenous nature of TFP, as well as the presence of large estimation errors in a variable that it is not directly observable.

Identification of the shock driving the business cycle

- Based on Angeletos *et al.* (2020), we identify the unique shock that is responsible for most of the *common volatility* at the business cycle frequency band. In this way, we also filter out the effect of the ignorable errors, which cannot generate cyclical fluctuations.
- We compute the matrix that measures the (co-)volatility of the common component χ_t at the frequency band $[\varpi_0, \varpi_1]$,

$$\Theta(\varpi_0, \varpi_1) = \int_{\varpi_0}^{\varpi_1} \mathrm{Re}\, F_{\varkappa}(\varpi) \mathrm{d}\varpi$$

  where $F_{\varkappa}(\varpi)$ is spectral density matrix of the common component χ_t
- Let Q be the matrix formed by the eigenvectors that are associated with the first r non-increasing eigenvalues of the matrix $\Theta(\varpi_0, \varpi_1)$, then the linear combinations $Q'\chi_t$ represent the (static) principal components of χ_t at the frequency band $[\varpi_0, \varpi_1]$.

Continued...

- We label the (standardized) shock of the first principal components of χ_t at frequencies $[2\pi/32, 2\pi/6]$ as the Main Business Cycle Common Shock (MBCCS). The associated impulse response function (IRF) for series Y_t is

$$\Psi_{\bullet 1}(L) = C(L)D^{-1}C_{\bullet 1},$$

where $D = Q'C_0$, $C = \mathrm{Chol}(DA'\Sigma_u AD')$, and $C_{\bullet 1}$ is the first column of C.

- Given the frequency domain nature of our identification scheme, we evaluate the effects of the MBCCS by its contribution to the variability of the $i-$th series at frequencies $[\pi/16, \pi/3]$ and 0.

Paper (updated version) at http://arxiv.org/abs/2009.03361

Continued...

Table: Contributions of the MBCCS to the variability of the 9 key series at frequencies $[\pi/16, \pi/3]$ and 0

Period	GDP	Con	Inv	UR	Hours	Inf	IR	LP	TFP
$6-32$	40.2 (15.0)	15.3 (14.0)	47.9 (14.7)	44.2 (12.2)	43.7 (12.1)	8.7 (15.6)	28.9 (18.1)	21.0 (13.5)	5.4 (15.7)
∞	33.3 (12.1)	16.8 (12.9)	35.8 (12.0)	39.4 (11.7)	37.7 (12.4)	7.2 (12.3)	19.4 (15.3)	14.4 (12.2)	2.4 (12.5)

Note: bootstrap standard errors in brackets

Continued...

- The MBCCS triggers a procyclical effect on GDP and Inv peaking with one quarter delay, on Con with no delay, as well as on UR [Hours and IR] with a peak at two [three] quarters. Moreover, it explains a large fraction of the cyclical variability of Inv, Hours, UR and GDP. These results corroborate the claim that the considered shock is the main driver of the business cycle fluctuations.
- However, it has a limited positive impact on Inf, which peaks at one quarter and quickly dies out, and it marginally affects both LP and, especially TFP. Moreover, it explains a small portion of the cyclical movements of both Inf and TFP.

Continued...

- Regarding the long-run scenario, the MBCCS explains around 35% of the zero-frequency variability of UR, Hours, Inv, and GDP, whereas it has negligible explanatory power for the permanent variation in Inf and TFP. Surprisingly, the MBCCS is responsible for almost the same portion of variability of Con (around 14%) both in the long and short run.
- All in all, the findings above seem to preclude the interpretation of the MBCCS as either a productivity or a news shock on the one hand, and a traditional demand shock on the other hand. However, a meticulous interpretation of the empirical results of this application, in particular by means of a rigorous comparison with DSGE models, is beyond the scope of the present paper.

Continued...

Table: Information criteria of the DRVAR and FAVAR partial models of the key series

	BIC	HQIC	AIC
DRVAR	-7.52	-8.62	-9.37
FAVAR	-6.62	-8.48	-9.73

Conclusions

- We propose a dimensional reduction approach such that both a common right space and a common left null space are present in the coefficient matrices of a large VAR model. This specification allows to detect a small dimensional VAR that is responsible to generate the whole dynamics of the system.
- This approach has many potential applications such as forecasting big data from a small scale VAR without loosing relevant information, structural VAR analysis, realized covariance matrices modelling, etc.
- We document the practical value of our approach by both a Monte Carlo study and an empirical application to 211 aggregate economic variables.