

Econometric Analysis of High Dimensional VARs Featuring a Dominant Unit

Alexander Chudik (1) Hashem Pesaran (2)
(1) CIMF and ECB
(2) Cambridge University, CIMF and USC

October 2009

Outline

- Related literature
- Weak vs strong cross section dependence
- IVAR Model
- Estimation and inference
- Monte Carlo experiments
- An empirical application

Motivation

- VAR models play a very important role in empirical macroeconomics and finance
- VAR theory is well developed when the number of variables N is small (6-7) and the number of time periods, T , is large
- For many empirical application such as global macroeconomic modelling, modelling of regions, firms, households, financial markets,... this theoretical framework is not appropriate
- Panel data models focus on small T and resolve the curse of dimensionality by dealing with homogeneous slope models with no cross section error dependence. See, panel VAR in Binder Hsiao and Pesaran (2005, ET)
- Our aim is to develop a theory for the analysis of VARs when both N and T are large

Related Literature: dynamics and cross section dependence

- Dealing with the **curse of dimensionality**:
 - (i) Data shrinkage (e.g. along the lines of index models) and
 - (ii) Shrinkage of the parameter space (e.g. in form of Bayesian priors or spatial weights matrices).
- Dealing with **cross section dependence (CD)**:
 - (i) Spatial processes
 - (ii) Factor structures

- Spatial processes were pioneered by Whittle (1954), and later by Cliff and Ord (1973). Anselin(1988), Kelejian and Robinson (1995), Kelejian and Prucha (1999, 2004, 2007), and Lee (2002, 2004, 2007) have made important contributions in econometrics.
- Factor models were introduced by Hotelling (1933), applied in economics by Stone (1947), and recently extensively in finance and macroeconomics (Chamberlain and Rothschild 1983; Connor and Korajczyk, 1993; Kapetanios and Pesaran, 2007; Forni and Reichlin, 1998; Stock and Watson, 2002).
- Chudik, Pesaran and Tosetti (2009) distinguish between weak and strong, which will be useful in the analysis below.
- Chudik and Pesaran (2009) consider an IVAR model without a dominant unit. We extend this paper and allow for a dominant unit.

Weak and Strong Cross Section Dependence

- Understanding concepts of weak and strong CS dependence is important for the analysis of IVAR models.
- Let $\mathbf{z}_t = (z_{1t}, \dots, z_{Nt})'$, with $E(\mathbf{z}_t | \mathcal{I}_{t-1}) = \mathbf{0}$, $\text{Var}(\mathbf{z}_t | \mathcal{I}_{t-1}) = \boldsymbol{\Sigma}_t$, where \mathcal{I}_{t-1} is the information set at time $t - 1$, and for each t where $\boldsymbol{\Sigma}_t$ has diagonal elements $0 < \sigma_{ii,t} \leq K$, for $i = 1, 2, \dots, N$.
- Let $\mathbf{w}_t = (w_{1t}, \dots, w_{Nt})'$ be a vector of weights satisfying the *granularity conditions*

$$\|\mathbf{w}_t\|_2 = O\left(N^{-\frac{1}{2}}\right), \quad \frac{w_{jt}}{\|\mathbf{w}_t\|_2} = O\left(N^{-\frac{1}{2}}\right) \quad \text{for any } j \leq N \quad (1)$$

An obvious example is equal weights, $w_i = N^{-1}$.

- The process $\{z_{it}\}$ is *weakly cross sectionally dependent* (CWD) at a point in time t , if for *all* \mathbf{w}_t

$$\lim_{N \rightarrow \infty} \text{Var}(\mathbf{w}'_t \mathbf{z}_t | \mathcal{I}_{t-1}) = 0$$

The process $\{z_{it}\}$ is *cross sectionally strongly dependent* (CSD) at a point in time t , if there exists \mathbf{w}_t such that

$$\text{Var}(\mathbf{w}'_t \mathbf{z}_t | \mathcal{I}_{t-1}) \geq K > 0$$

where K is a constant independent of N .

- The process $\{z_{it}\}$ is CSD at time $t \in \mathcal{T}$ if and only if

$$\lim_{N \rightarrow \infty} \frac{1}{N} \lambda_1(\boldsymbol{\Sigma}_t) = K > 0,$$

i.e. $\lambda_1(\boldsymbol{\Sigma}_t)$ increases to infinity at the rate N .

- If $\lambda_1(\boldsymbol{\Sigma}_t) = O(N^{1-\epsilon})$ for any $\epsilon > 0$, then

$$\lim_{N \rightarrow \infty} (\mathbf{w}'_t \mathbf{w}_t) \lambda_1(\boldsymbol{\Sigma}_t) = 0,$$

and the underlying process will be CWD. Hence, the bounded eigenvalue condition is *sufficient but not necessary* for CWD.

- CWD and CSD can be defined equally with respect to any information set, such as \mathcal{I}_{-M} , for any fixed M , or as M tends to infinity (if the underlying process is stationary).

Stationary Infinite Dimensional VAR Model

Suppose that for each $N \in \mathbb{N}$, vector of N endogenous variables $\mathbf{x}_{(N),t} = (x_{(N),1t}, \dots, x_{(N),Nt})'$ is given by the following VAR model,

$$\mathbf{x}_{(N),t} = \mathbf{\Phi}_{(N)} \mathbf{x}_{(N),t-1} + \mathbf{u}_{(N),t}, \quad (2)$$

where $\mathbf{\Phi}_{(N)}$ is $N \times N$ matrix of coefficients, $\mathbf{u}_{(N),t}$ is $N \times 1$ vector of error terms given by

$$\mathbf{u}_{(N),t} = \delta_{(N)} \mathbf{e}_t + \mathbf{R} \boldsymbol{\varepsilon}_{(N),t}, \quad (3)$$

where $\boldsymbol{\varepsilon}_{(N),t} = (\varepsilon_{1t}, \varepsilon_{2t}, \dots, \varepsilon_{Nt})'$, and the individual elements of the double index array $\{\varepsilon_{jt}, j \in \mathbb{N}, t \in \mathbb{Z}\}$ are *IID*.

- The results that follow hold irrespective of whether the parameters $\phi_{(N),ij}$, $\delta_{(N),i}$, and $r_{(N),ij}$ are assumed to be varying with N or not.
- Note that even if individual coefficients do not change with N , covariance between individual units, $\text{cov}\left(x_{(N),it}, x_{(N),jt}\right)$ in general must be changing with N , unless matrices $\Phi_{(N)}$ and $\mathbf{R}_{(N)}$ are lower triangular.
- Dependence on N is suppressed in the remainder of this presentation to simplify the notations, but it is understood that the parameters and the dimension of the random variables \mathbf{x}_t and \mathbf{u}_t vary with N , unless otherwise stated.

Spatial or Network Dependence

- The spatial dependence, pioneered by Whittle (1954), models cross section correlations by means of spatial processes that relate each cross section unit to its neighbour(s). (Cliff and Ord, 1973, Anselin, 1988, and Kelejian and Robinson, 1995).
- It is not necessary that proximity is measured in terms of physical space.
- In the case of the IVAR model contemporaneous dependence can be modelled through an $N \times N$ network topology matrix \mathbf{R} so that

$$\mathbf{u}_t = \mathbf{R}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim IID(\mathbf{0}, \mathbf{I}_N), \quad (6)$$

- For example, \mathbf{R} could characterize a first order SMA, or SAR, or \mathbf{R} could correspond to the 'star' network.

$$\mathbf{R}_{SMA} = \mathbf{I}_N + \rho_s \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 1/2 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 1/2 & 0 & 1/2 \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 & 0 \end{pmatrix},$$

$$\mathbf{R}_{Star} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ r_{21} & 1 & \dots & 0 & 0 \\ r_{31} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & 1 & 0 \\ r_{N1} & 0 & \dots & 0 & 1 \end{pmatrix}, \text{ with } \sum_{j=2}^N r_{j1} = O(N).$$

Strong Dependence in IVAR models

- Strong dependence in the IVAR model could be due to CSD errors $\{u_{it}\}$, or dominant columns/rows of Φ , or both.
- Suppose the column $i = 1$ in matrices Φ , and \mathbf{R} is dominant. The following assumption postulates that for any i , coefficient vector ϕ_i can be decomposed into a sparse vector $\phi_{ai} = (\phi_{i1}, 0, \dots, 0, \phi_{ii}, 0, \dots, 0)'$ and a vector $\phi_{bi} = \phi_{-1,-i}$ where $\phi_{-1,-i} = (0, \phi_{i2}, \dots, \phi_{i,i-1}, 0, \phi_{i,i+1}, \dots, \phi_{iN})'$.

Assumption 5: Let $\Phi = \sum_{i=1}^N \phi_i \mathbf{e}_i' = \phi_1 \mathbf{e}_1' + \Phi_{-1}$ where $\phi_i = (\phi_{1i}, \dots, \phi_{Ni})'$ is the i^{th} column of matrix Φ , \mathbf{e}_i is an $N \times 1$ selection vector for unit i , with the i^{th} element of \mathbf{e}_i being one and the remaining zero. Suppose as $N \rightarrow \infty$,

$$\|\phi_1\|_r = O(1), \quad \|\phi_{-1,-i}\|_r = O(N^{-1}) \quad \text{for any } i \in \mathcal{S}.$$

- The equation for unit i in system (2) is

$$x_{it} = \sum_{j=1}^N \phi_{ij} x_{j,t-1} + u_{it}.$$

- We shall assume that the following absolute summability condition holds,

$$\sum_{j=1}^N |\phi_{ij}| < K \text{ for any } i \text{ and any } N \in \mathbb{N}. \quad (4)$$

- This is not a restrictive condition, and is similar to the concept of absolutely summability in the time series literature, but the summing is over lagged values of cross section units as opposed to AR terms of unit i , or MA terms.

- Similar constraint is used in 'data mining' literature, in particular Lasso (Tibshirani, 1996) and Ridge regression shrinkage method. Using Lasso around 5-10% of the regression coefficients end up being exactly equal to zero.
- The Lasso minimizes residual sum of squares subject to $\sum_{j=1}^N |\phi_{ij}| \leq K$.
- Ridge regression minimizes residual sum of squares subject to $\sum_{j=1}^N \phi_{ij}^2 \leq K$.
- We do not choose the value for K , as it is necessary in shrinkage methods, but rather assume only its existence instead.

- For each i we divide the units into a finite number of 'neighbors', denoted by \mathcal{N}_i , that have fixed coefficients $|\phi_{ij}| < K$ for $j \in \mathcal{N}_i$, and nonneighbors $j \in \mathcal{N}_i^c \equiv \{1, \dots, N\} \setminus \mathcal{N}_i$ that have coefficients that are of order $1/N$:

$$\begin{aligned}
 x_{it} &= \underbrace{\sum_{j \in \mathcal{N}_i} \phi_{ij} x_{j,t-1}}_{\text{Neighbors}} + \underbrace{\sum_{j \in \mathcal{N}_i^c} \phi_{ij} x_{j,t-1}}_{\text{Nonneighbors}} + u_{it} \\
 &= \overset{\circ}{\phi}'_i \mathbf{x}_{t-1} + \tilde{\phi}'_i \mathbf{x}_{t-1} + u_{it},
 \end{aligned}$$

$\overset{\circ}{\phi}_i$ and $\tilde{\phi}_i$ are obtained from ϕ_i by replacing nonneighbors and neighbors coefficients with zeros.

- If it is possible to divide units in this way, then we no longer have dimensionality problem.
- Notation: We use $\|\cdot\|_\infty$ and $\|\cdot\|_1$ to denote maximum absolute row and column sum matrix norms, respectively.
- Division of units into neighbors and nonneighbors implies:
$$\left\| \overset{\circ}{\phi}_i \right\|_\infty = \max_{j \in \mathcal{N}_i} |\phi_{ij}| < K \text{ and}$$
$$\left\| \tilde{\phi}_i \right\|_\infty = \max_{j \in \mathcal{N}_i^c} |\phi_{ij}| < \frac{K}{N}.$$
- Remark: This assumption could be relaxed without the loss of generality by assuming more general restrictions in form of spatial weights matrices as in Chudik and Pesaran (2009)

- Consider aggregate spatiotemporal impact of nonneighbors given by $\tilde{\phi}'_i \mathbf{x}_{t-1}$.
- The absolute sum $\sum_{j \in \mathcal{N}_i^c} |\phi_{ij}| = \|\tilde{\phi}_i\|_1 < K$ and nonneighbors could have large aggregate impact on unit i .
- However, the euclidean norm $\|\tilde{\phi}_i\| \leq \sqrt{\|\tilde{\phi}_i\|_\infty \|\tilde{\phi}_i\|_1} = \sqrt{O\left(\frac{1}{N}\right) O(1)} = O\left(N^{-\frac{1}{2}}\right)$ and it follows that $\tilde{\phi}'_i \mathbf{x}_{t-1} \xrightarrow{q.m.} 0$ if and only if $\{x_{it}\}$ is weakly CS dependent.
- If $\{x_{it}\}$ is strongly CS dependent process then $\lim_{N \rightarrow \infty} \text{Var}\left(\tilde{\phi}'_i \mathbf{x}_{t-1}\right)$ is not necessarily zero.
- This paper allows for dominant unit, which is a source of strong CS dependence.

Assumptions

- To simplify the exposition, we consider only the case where $\mathcal{N}_i = \{1, i\}$, that is only lagged dominant unit and own lagged coefficient are $O(1)$, and the remaining elements are nonneighbors:

Assumption 1: (*Influence of unit 1 on the rest of the system is unrestricted.*) There exists a constant $K < \infty$ (independent of N) such that $|\phi_{ii}| < K$, $|\phi_{i1}| < K$, and

$$\|\boldsymbol{\phi}_{-1,-i}\|_{\infty} < \frac{K}{N}, \text{ for any } i, N \in \mathbb{N},$$

where $\boldsymbol{\phi}_{-1,-i} = (0, \phi_{i2}, \dots, \phi_{i,i-1}, 0, \phi_{i,i+1}, \dots, \phi_{iN})'$.

Assumption 2 (Stationarity) There exists a constant $0 < \rho < 1$ (independent of N) such that for any $N \in \mathbb{N}$:

$$\|\Phi\|_{\infty} < \rho < 1, \quad \|\Phi_{(-1)}\|_{\infty} < \frac{\rho}{2}, \quad \text{and} \quad \|\Phi_{(-1)}\|_{\infty} \|\Phi_{(-1)}\|_1 < \rho$$

where $\Phi_{(-1)}$ is $N \times N$ matrix constructed from matrix Φ by replacing its first column with a zero vector.

- Remark: Condition $\|\Phi\|_{\infty} < \rho < 1$ implies $\rho(\Phi) \leq \|\Phi\|_{\infty} < 1$ and therefore $\max_{1 \leq j \leq N} |\lambda_j(\Phi)| < 1$, which is the well known sufficient and necessary condition for stationarity when N is fixed. We need stronger conditions than eigenvalues inside unit circle. (We need $\text{Var}(x_{it}) < K$ for any $N \in \mathbb{N}$).
- Assumption 2 can be relaxed.

Assumption 3 (Starting values) Available observations are $\mathbf{x}_0, \mathbf{x}_1, \dots, \mathbf{x}_T$ with the starting values $\mathbf{x}_0 = \sum_{\ell=0}^{\infty} \Phi^{\ell} \mathbf{u}(-\ell)$.

Assumption 4 (Errors) Vector of errors \mathbf{u}_t is given by (3), $\delta_1 = 1$, $\|\delta\|_{\infty} = O(1)$, e_t and the individual elements of the double index array $\{\varepsilon_{jt}, j \in \mathbb{N}, t \in \mathbb{Z}\}$ are identically and independently distributed with mean 0, unit variances and finite fourth moments uniformly bounded in $j \in \mathbb{N}$. Furthermore, matrix \mathbf{R} has bounded row and column matrix norms and $|r_{1j}| \leq \frac{K}{N}$ for any $j \in \mathbb{N}$ and any $N \in \mathbb{N}$.

Convergence Results

- Under Assumptions 1-4 we have:

$$\mathbf{x}_t = \sum_{\ell=0}^{\infty} \mathbf{\Phi}_{(-1)}^{\ell} \boldsymbol{\phi}_1 x_{1,t-1-\ell} + \sum_{\ell=0}^{\infty} \mathbf{\Phi}_{(-1)}^{\ell} \delta \mathbf{e}_{t-\ell} + \mathbf{v}_t,$$

where $\mathbf{v}_t = \sum_{\ell=0}^{\infty} \mathbf{\Phi}_{(-1)}^{\ell} \mathbf{R} \boldsymbol{\varepsilon}_{t-\ell}$ is CWD, in particular $\mathbf{w}' \mathbf{v}_t = O_p \left(N^{-\frac{1}{2}} \right)$.

- As a consequence, the equation for unit $i > 1$ is

$$x_{it} = \phi_{ii} x_{i,t-1} + \beta_i(L) x_{1t} + \eta_{it} + \zeta_{it},$$

where $\eta_{it} = \mathbf{r}'_i \boldsymbol{\varepsilon}_t$, ζ_{it} is $O_p \left(N^{-\frac{1}{2}} \right)$ and in general the polynomial $\beta_i(L)$ is a function of all elements of $\mathbf{\Phi}$ and δ .

Objectives of the exercise

- Focus is on the consistent estimation of the unit-specific unknown coefficients ϕ_{ij} for $i > 1$ and also on the estimation of the impact of the dominant unit on the remaining units in the system, captured by $\beta_i(L)$.
- We allow for cross section dependence of innovations $\eta_t = \mathbf{R}\varepsilon_t$, as characterized by matrix \mathbf{R} , but we ignore this matrix for estimation and therefore our estimators are not necessarily efficient in the presence of cross section dependence induced by the matrix \mathbf{R} .
- This paper also does not deal with the pooled estimation of the mean coefficients over the cross section units.

- ϕ_{ii} and $\beta_i(L)$ can be consistently estimated.
- Based on the asymptotic representation of unit i , we consider the following auxiliary regression for $i > 1$:

$$\begin{aligned}x_{it} &= \phi_{ii}x_{i,t-1} + \sum_{\ell=0}^k \beta_{i\ell}x_{i,t-\ell} + \epsilon_{it} \\ &= \mathbf{g}'_{it}\boldsymbol{\pi}_i + \epsilon_{it},\end{aligned}\tag{5}$$

where $\boldsymbol{\pi}_i = (\phi_{ii}, \beta_{i0}, \dots, \beta_{ik})'$.

- Identification requires $\mathbf{C}_i = E(\mathbf{g}_{it}\mathbf{g}'_{it})$ to be positive definite.

Theorems

Theorem 1 (Consistency) Under Assumptions 1-4, and invertibility of $\frac{1}{T-k} \sum_{t=k+1}^T \mathbf{g}_{it} \mathbf{g}'_{it}$, we have

$$\|\hat{\boldsymbol{\pi}}_i - \boldsymbol{\pi}_i\|_{\infty} \xrightarrow{P} 0, \text{ for any } i > 1,$$

as $N, T \xrightarrow{j} \infty$ at any order, and $k^2/T \rightarrow 0$ such that there exists constants $r_1, r_2 > 0$ satisfying $k > r_1 T^{r_2}$.

Theorem 2. (Inference) Under assumptions of Theorem 1, for any sequence of $(k + 1) \times 1$ dimensional vectors \mathbf{a}_k such that

$\|\mathbf{a}_k\| = 1$ and $\|\mathbf{a}_k\|_1 = O(1)$, and as $N, T \xrightarrow{j} \infty$, $T/N \rightarrow \varkappa < \infty$ ($\varkappa \geq 0$ is not necessarily nonzero), and $k^2/T \rightarrow 0$ such that there exists constants $r_1, r_2 > 0$ satisfying $k > r_1 T^{r_2}$, we have

$$\sqrt{T - k} \frac{1}{\sigma_i} \mathbf{a}_k' \mathbf{C}_i^{\frac{1}{2}} (\hat{\pi}_i - \pi_i) \xrightarrow{d} N(0, 1), \text{ for } i > 1, \quad (6)$$

where $\hat{\pi}_i$ is LS estimator of π_i in regression (5), matrix

$\mathbf{C}_i = E(\mathbf{g}_{it} \mathbf{g}_{it}')$ can be consistently estimated by

$\hat{\mathbf{C}}_i = \frac{1}{T-k} \sum_{t=k+1}^T \mathbf{g}_{it} \mathbf{g}_{it}'$, and $\sigma_i^2 = \text{Var}(\eta_{it})$.

Monte Carlo Simulations

Monte Carlo Design

- We consider factor augmented IVAR model featuring dominant unit,

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \gamma v_t + \mathbf{u}_t, \text{ with } v_t \sim IIDN(0, \sigma_v^2),$$

$$\mathbf{u}_t = \delta e_t + \boldsymbol{\eta}_t, \text{ and } \boldsymbol{\eta}_t = \mathbf{R}\boldsymbol{\varepsilon}_t.$$

- We need to generate coefficient matrix Φ , errors \mathbf{u}_t , and factor loadings γ .
- Alternative would be to consider $(\mathbf{x}_t - \gamma f_t) = \Phi (\mathbf{x}_{t-1} - \gamma f_{t-1}) + \mathbf{u}_t$ as DGP with factor f_t generated as a persistent stationary process (work in progress).

Construction of Coefficient Matrix

- We generate $O_p(N^{-1})$ random variables λ_{ij} such that $\sum_{j=1}^N \lambda_{ij} = 1$ first.
- Matrix Φ is then constructed as follows.
 - (Dominant unit $i = 1$) $\phi_{11} = 0.7$, and $\phi_{1j} = \alpha_1 \lambda_{1j}$ for $j = 2, \dots, N$, with $\alpha_1 = 0.1$.
 - (Unit $i = 2$) $\phi_{21} = 0.1$, $\phi_{22} = 0.5$, and $\phi_{2j} = \alpha_2 \lambda_{2j}$ for $j = 3, \dots, N$, with $\alpha_2 = 0.1$.
 - (Remaining units $i > 2$) $\phi_{ii} \sim IIDU(0.3, 0.5)$,
 $\phi_{i1} \sim IIDU(0, 0.2)$, $\phi_{ij} = \alpha_i \lambda_{ij}$ for $j > 1, j \neq i$ where
 $\alpha_i \sim IIDU(0.05, 0.15)$.

Construction of Reduced Form Errors

- $\mathbf{u}_t = \delta e_t + \boldsymbol{\eta}_t$, where $e_t \sim IIDN(0, 0.15)$, we set $\delta_1 = 1$, $\delta_2 = 0.1$ and generate $\delta_i \sim IIDU(0, 0.2)$ for $i = 3, \dots, N$.
- We set $\eta_{1t} = 0$ and $\{\eta_{2t}, \dots, \eta_{Nt}\}$, are generated from a stationary bilateral Spatial Autoregressive Model (SAR) in order to show that our estimators are invariant to the weak cross section dependence of innovations:

$$\eta_{it} = \frac{a_\eta}{2} (\eta_{i-1,t} + \eta_{i+1,t}) + \vartheta_{eit},$$

where $\vartheta_{eit} \sim IIDN(0, \sigma_{\vartheta_e}^2)$, and two options for SAR parameter a_η are considered, $a_\eta \in \{0.5, 0.9\}$.

- SAR process is CWD (See Pesaran and Tosetti, 2009, for proof).

Construction of Factor Loadings

- Two sets of factor loadings are considered, $\gamma = \mathbf{0}$ (no unobserved common factor) and $\gamma \neq \mathbf{0}$. Under the latter we set $\gamma_1 = \gamma_2 = 0.5$ and generate the remaining factor loadings randomly as $\gamma_i \sim IIDN(-0.6, 1)$ for $i = 3, \dots, N$.
- In total we have 4 experiments (zero or nonzero factor loadings and high or low spatial dependence of errors). We report experiments with high spatial dependence of errors ($a_{\eta} = 0.9$) since experiments with low spatial dependence are very similar.

Specification of Auxiliary Regressions

- We consider three different augmentations:
 - (i) by dominant unit and arithmetic cross section averages $\{x_{1,t-l}, \bar{x}_{t-l}\}_{l=0}^p$ with $p = 1$,
 - (ii) augmentation by dominant unit only, $\{x_{1,t-l}\}_{l=0}^p$ with p being the largest integer smaller than $T^{1/3}$, and
 - (iii) augmentation by arithmetic cross section averages $\{\bar{x}_{t-l}\}_{l=0}^p$ with the number of lags p chosen to grow with T in the same way.
- Auxiliary regression for unit $i = 2$ corresponding to augmentation of form (i) is:

$$x_{2t} = c_2 + \phi_{22}x_{2,t-1} + \sum_{l=0}^1 b_{1l}x_{1,t-l} + \sum_{l=0}^1 b_{2l}\bar{x}_{t-l} + \epsilon_{2t}.$$

Table 1a: RMSE of $\hat{\phi}_{22}$ in the experiments with high spatial dependence and zero factor loadings ($a_{\eta} = 0.9$ and $\gamma = 0$).

		Auxiliary regressions are augmented by:								
		$\{x_{1,t-l}, \bar{x}_{t-l}\}_{l=0}^1$			$\{\bar{x}_{t-l}\}_{l=0}^p, p \approx T^{-1/3}$			$\{x_{1,t-l}\}_{l=0}^p, p \approx T^{-1/3}$		
N \ T		50	100	200	50	100	200	50	100	200
25		16.87	10.25	6.76	16.87	10.45	6.96	15.67	9.54	6.36
50		16.13	9.85	6.68	16.57	10.11	6.74	15.59	9.45	6.42
75		16.13	10.23	6.47	16.34	10.32	6.59	15.84	10.06	6.43
100		15.99	10.17	6.44	16.27	10.29	6.61	15.53	9.91	6.45
200		16.04	9.92	6.53	16.27	10.20	6.60	15.55	9.99	6.46

Table 1b: Size ($\times 100$) (5% level, $H_0 : \phi_{22} = 0.50$). in experiments with high spatial dependence and zero factor loadings ($a_{\eta} = 0.9$ and $\gamma = 0$).

		Auxiliary regressions are augmented by:								
		$\{x_{1,t-l}, \bar{x}_{t-l}\}_{l=0}^1$			$\{\bar{x}_{t-l}\}_{l=0}^p, p \approx T^{-1/3}$			$\{x_{1,t-l}\}_{l=0}^p, p \approx T^{-1/3}$		
N \ T		50	100	200	50	100	200	50	100	200
25		9.50	7.20	6.75	8.75	6.50	6.55	7.15	5.25	5.80
50		8.55	6.60	6.10	8.40	6.15	5.40	6.80	5.15	4.45
75		8.85	7.00	6.25	7.60	6.25	5.70	7.15	6.15	5.45
100		8.85	6.55	6.00	8.30	6.60	5.80	6.90	5.80	5.40
200		8.60	6.55	5.60	8.25	6.25	5.20	6.90	6.35	4.75

Table 1c: Power ($\times 100$) (5% level, $H_1 : \phi_{22} = 0.60$) in experiments with high spatial dependence and zero factor loadings ($a_{\eta} = 0.9$ and $\gamma = 0$).

		Auxiliary regressions are augmented by:								
		$\{x_{1,t-l}, \bar{x}_{t-l}\}_{l=0}^1$			$\{\bar{x}_{t-l}\}_{l=0}^p, p \approx T^{-1/3}$			$\{x_{1,t-l}\}_{l=0}^p, p \approx T^{-1/3}$		
$N \setminus T$		50	100	200	50	100	200	50	100	200
25		25.85	32.95	50.10	22.80	30.85	48.80	18.45	23.00	33.25
50		24.50	31.30	49.45	22.30	28.60	45.45	18.90	22.30	37.90
75		23.60	33.00	48.15	22.00	31.55	45.10	20.10	27.05	39.45
100		22.95	31.10	46.55	20.55	29.25	44.75	19.25	27.05	39.15
200		23.50	29.70	47.30	22.35	27.10	45.30	20.95	25.35	42.15

Table 2a: RMSE of $\hat{\phi}_{22}$ in the experiments with high spatial dependence and nonzero factor loadings ($a_{\eta} = 0.9$ and $\gamma \neq 0$).

		Auxiliary regressions are augmented by:								
		$\{x_{1,t-l}, \bar{x}_{t-l}\}_{l=0}^1$			$\{\bar{x}_{t-l}\}_{l=0}^p, p \approx T^{-1/3}$			$\{x_{1,t-l}\}_{l=0}^p, p \approx T^{-1/3}$		
N \ T		50	100	200	50	100	200	50	100	200
25		15.52	9.94	6.46	14.33	11.28	11.38	17.17	11.54	7.79
50		15.24	9.28	6.26	15.05	13.61	13.79	17.55	11.32	8.20
75		15.39	9.64	6.34	15.25	14.19	14.61	17.52	11.88	8.32
100		15.43	9.61	6.18	15.32	14.51	15.21	17.63	12.00	8.30
200		15.35	8.94	6.29	16.09	15.92	15.77	18.41	11.32	8.55

Table 2b: Size ($\times 100$) (5% level, $H_0 : \phi_{22} = 0.50$). in experiments with high spatial dependence and nonzero factor loadings ($a_{\eta} = 0.9$ and $\gamma \neq 0$).

		Auxiliary regressions are augmented by:								
		$\{x_{1,t-l}, \bar{x}_{t-l}\}_{l=0}^1$			$\{\bar{x}_{t-l}\}_{l=0}^p, p \approx T^{-1/3}$			$\{x_{1,t-l}\}_{l=0}^p, p \approx T^{-1/3}$		
N \ T		50	100	200	50	100	200	50	100	200
25		8.50	7.75	6.30	9.50	15.95	40.15	9.05	9.95	9.25
50		8.05	5.00	5.25	12.10	27.80	57.30	9.10	8.75	10.40
75		8.60	6.60	6.30	13.35	31.25	65.40	9.70	10.35	11.85
100		8.80	6.75	5.60	13.80	33.90	69.00	9.80	10.55	11.05
200		8.60	5.70	6.15	16.20	42.50	71.75	10.85	8.55	12.30

Table 2c: Power ($\times 100$) (5% level, $H_1 : \phi_{22} = 0.60$) in experiments with high spatial dependence and nonzero factor loadings ($a_{\eta} = 0.9$ and $\gamma \neq 0$).

		Auxiliary regressions are augmented by:								
		$\{x_{1,t-l}, \bar{x}_{t-l}\}_{l=0}^1$			$\{\bar{x}_{t-l}\}_{l=0}^p, p \approx T^{-1/3}$			$\{x_{1,t-l}\}_{l=0}^p, p \approx T^{-1/3}$		
N \ T		50	100	200	50	100	200	50	100	200
25		23.00	33.50	49.40	7.80	6.40	6.55	25.55	38.70	59.00
50		22.60	31.75	48.70	6.70	6.15	9.80	24.90	37.05	61.50
75		24.45	32.30	50.25	7.10	7.25	10.75	26.35	39.15	63.65
100		23.65	33.40	48.20	6.10	7.15	13.40	25.95	40.95	63.15
200		25.00	30.50	50.80	6.70	9.75	16.75	28.25	38.15	64.75

Modelling Returns on Equity Futures

- We estimate impact of S&P weekly returns on the remaining stock market indices across the globe, where we assume that S&P is dominant unit.
- $N = 26$, and $T = 308$ observations of weekly returns (06 January 2003 - 24 November 2008)
- The equity series refer to futures contracts downloaded from Datastream. Daily returns were calculated allowing for contract rollovers. Weekly returns were calculated from daily returns by summing working days (Monday to Friday).
- It is assumed that \mathbf{x}_t is given by the following VAR model, $\mathbf{x}_t = \mathbf{d} + \Phi \mathbf{x}_{t-1} + \mathbf{u}_t$, and $\mathbf{u}_t = \delta \mathbf{e}_t + \boldsymbol{\eta}_t$, where $\boldsymbol{\eta}_t = \mathbf{R} \boldsymbol{\varepsilon}_t$ is CWD and assumptions of Theorem 2 holds.

- The following *conditional* models are estimated

$$x_{it} = c_i + \sum_{\ell=1}^{k_{oi}} \phi_{ii,\ell} x_{i,t-\ell} + \sum_{\ell=0}^{k_{di}} \beta_{i\ell} x_{1,t-\ell} + \sum_{\ell=1}^{k_{si}} \delta_{i\ell} \bar{x}_{wi,t-\ell} + \epsilon_{it},$$

for $i = 2, \dots, N$, and for $i = 1$ (S&P), we estimate the following *marginal* model,

$$x_{1t} = c_1 + \sum_{\ell=1}^{k_{o1}} a_{\ell} x_{1,t-\ell} + \sum_{\ell=1}^{k_{s1}} b_{\ell} \bar{x}_{w1,t-\ell} + \epsilon_{1t},$$

where $\bar{x}_{wit} = \sum_{j=1}^N w_{ij} x_{jt}$ represents the spatial weighted average.

- Truncation lags were chosen according to SBC criterion with the maximum lag set to 4.

Unit	SP Cont. Eff.		SP Lag Eff.		Own Lag Eff.		Spatial Lag Eff.		\bar{R}^2	
	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio		
Dominant Unit										
S&P	-	-	-	-	0.563	3.6	-0.543	-4.3	0.067	
Advanced European Countries										
AEX	1.086	24.3	0.441	4.5	-0.538	-4.0	0.469	2.4	0.695	
BEL	0.946	21.8	0.718	6.4	-0.046	-0.5	-0.241	-1.8	0.665	
CAC	0.959	28.0	0.328	4.5	-0.274	-2.5	0.027	0.2	0.749	
DAX	1.089	26.9	0.461	4.3	-0.294	-3.8	0.061	0.6	0.753	
FTSE	0.833	27.4	0.292	3.2	-0.181	-2.4	-0.062	-0.8	0.751	
FOX	0.900	20.7	0.612	4.7	-0.209	-2.2	-0.208	-1.8	0.638	
GRX	0.880	13.8	0.616	3.6	-0.095	-1.6	0.225	1.6	0.461	
IBE	0.857	20.9	0.433	4.1	0.034	0.4	-0.303	-2.8	0.651	
KFX	0.837	18.6	0.603	4.9	-0.114	-1.7	-0.111	-0.9	0.590	
MIB	0.852	25.6	0.326	4.5	-0.328	-2.7	0.185	1.4	0.722	
OBX	0.928	15.5	0.649	5.1	0.049	0.8	-0.334	-2.2	0.510	
OMX	0.969	23.2	0.337	3.8	-0.190	-2.5	0.052	0.5	0.656	
PSI	0.619	14.9	0.325	3.7	0.036	0.6	0.036	0.4	0.504	
SMI	0.855	22.8	0.211	2.7	-0.268	-3.4	0.038	0.4	0.685	

Unit	SP Cont. Eff.		SP Lag Eff.		Own Lag Eff.		Spatial Lag Eff.		\bar{R}^2
	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	Coef.	t-ratio	
Other Advanced Countries									
TSX	0.751	21.9	0.161	2.3	-0.095	-1.5	0.054	0.7	0.633
NK	0.919	17.4	0.323	2.8	-0.108	-1.7	0.107	0.8	0.539
ASX	0.600	16.5	0.284	3.7	-0.115	-1.8	0.019	0.2	0.503
Latin America									
BRX	1.210	18.7	0.327	1.8	-0.109	-1.9	0.107	0.7	0.571
Emerging Europe									
HUX	0.831	13.3	0.399	3.0	-0.018	-0.2	0.193	1.2	0.458
POX	0.898	13.1	0.180	1.2	-0.023	-0.4	0.100	0.7	0.370
SAX	0.756	12.9	0.217	2.2	-0.152	-2.5	-0.031	-0.3	0.361
Emerging Asia									
HKX	0.707	11.7	0.713	4.5	-0.206	-3.2	0.003	0.0	0.391
KOX	0.768	11.2	0.050	0.4	-0.551	-5.7	1.047	5.9	0.409
SIX	0.718	14.3	1.127	7.2	-0.386	-4.0	-0.122	-0.9	0.493
TWX	0.717	10.4	0.772	4.2	-0.380	-4.2	0.007	0.0	0.358

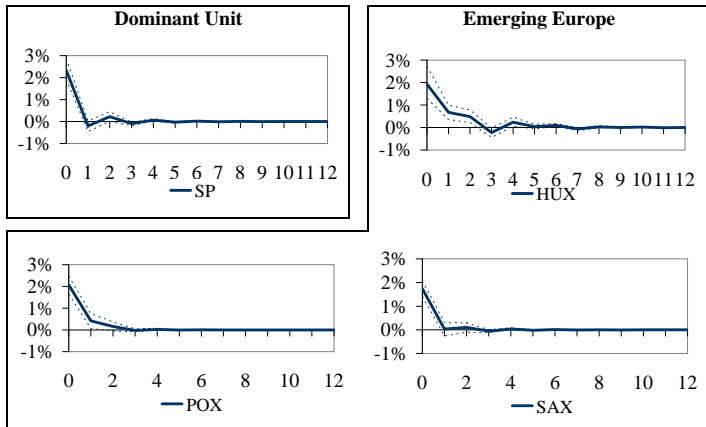


Figure 1: GIRF of a 1 s.d. shock to weekly S&P return and 90% confidence bounds: Impact on S&P, HUX (Hungary), POX (Poland) and SAX (Slovakia).

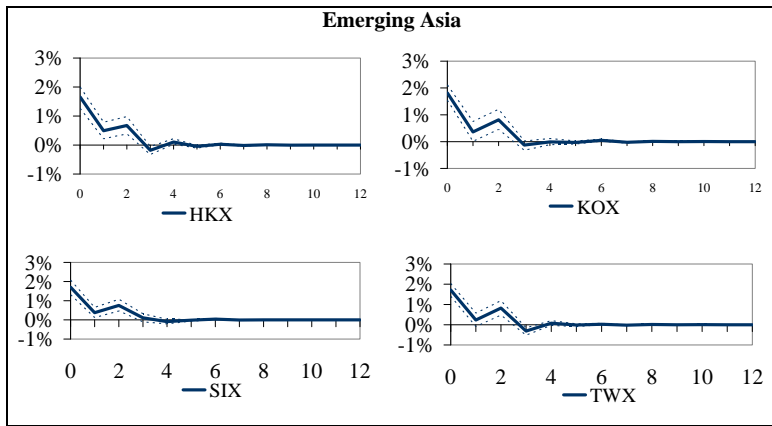


Figure 2: GIRF of a 1 s.d. shock to weekly S&P return and 90% confidence bounds: Impact on HKX (Hong Kong), KOX (Korea), SIX (Singapore), TWX (Taiwan).

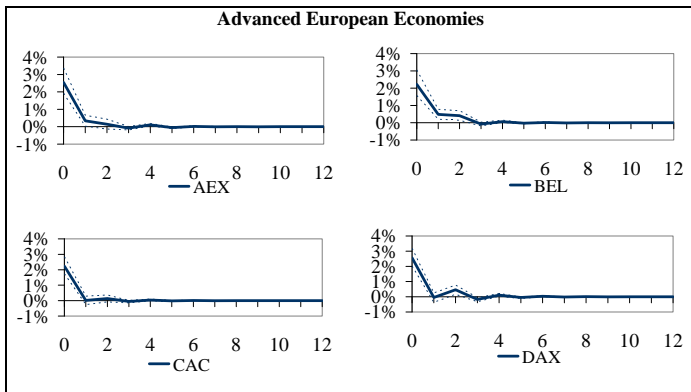


Figure 3: GIRF of a 1 s.d. shock to weekly S&P return and 90% confidence bounds: Impact on AEX (Netherlands), BEL (Belgium), CAC (France), DAX (Germany).

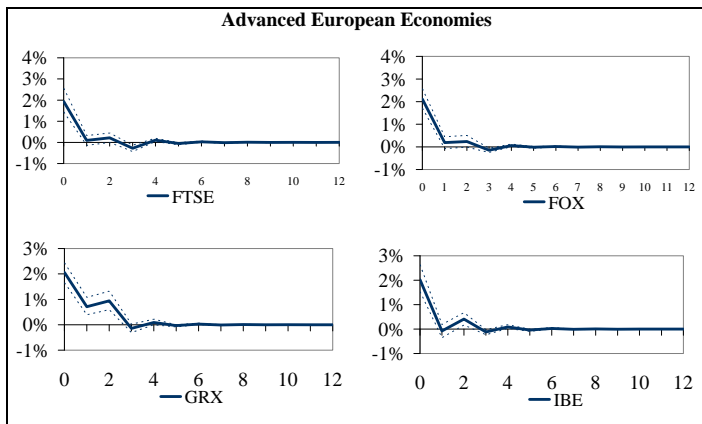


Figure 4: GIRF of a 1 s.d. shock to weekly S&P return and 90% confidence bounds: Impact on FTSE (UK), FOX (Finland), GRX (Greece), IBE (Spain).

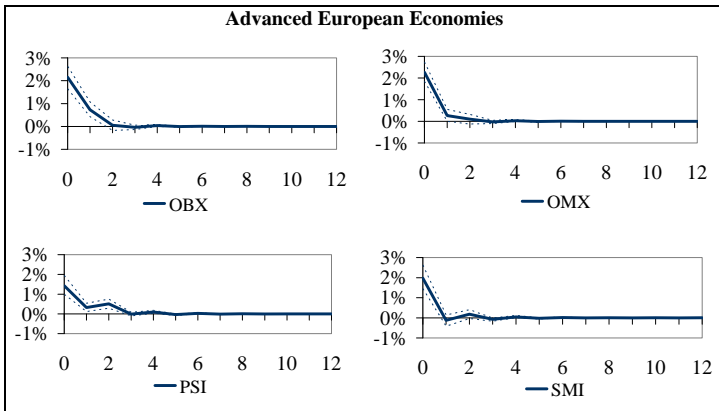


Figure 5: GIRF of a 1 s.d. shock to weekly S&P return and 90% confidence bounds: Impact on OBX (Norway), OMX (Sweden), PSI (Portugal), SMI (Switzerland).

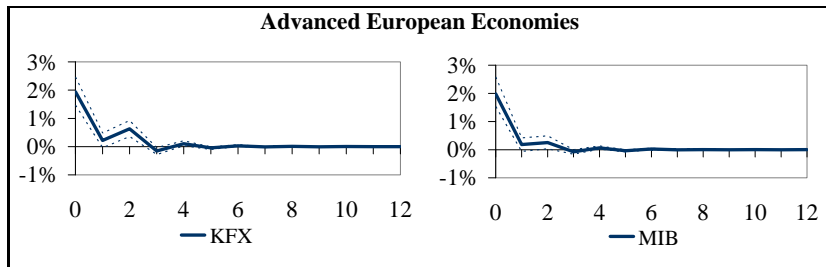


Figure: Figure 6: GIRF of a 1 s.d. shock to weekly S&P return and 90% confidence bounds: Impact on KFX (Denmark), MIB (Italy).

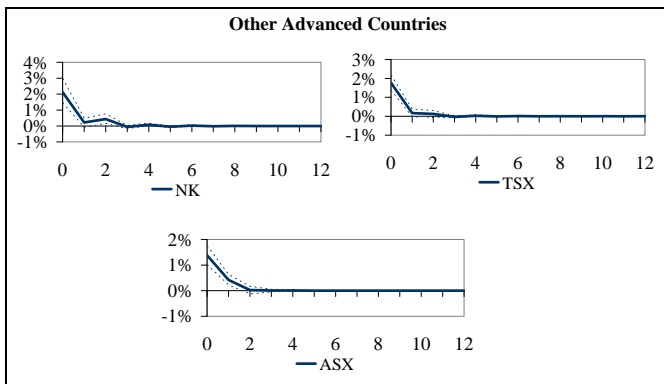


Figure 7: GIRF of a 1 s.d. shock to weekly S&P return and 90% confidence bounds: Impact on NK (Japan), TSX (Canada), ASX (Australia).

Conclusion

- This paper considered the problem of estimation of high dimensional VARs featuring a dominant unit.
- We showed that the asymptotic normality of the cross section augmented least squares estimator continues to hold (once the individual auxiliary regressions are correctly specified).
- How to correctly specify the individual regressions is an important topic, and the correct specification depends on the assumption about the presence of dominant units, observed and unobserved common factors and the (local) spatiotemporal neighborhood effects.
- The framework developed here can be applied to model spatio-temporal dependence. See "Spatial and Temporal Diffusion of House Prices in the UK" by Holly, Pesaran and Yamagata (2009, forthcoming).