MULTIVARIATE LEAST SQUARES ESTIMATION OF GLOBAL VAR MODELS

Daniel Ngugi[1]

Pan African University, Kenya.

Olusanya Olubusoye[2]**.**

University of Ibadan, Nigeria.

Busoye2001@yahoo.com

Paul Weke[3]

University of Nairobi, Kenya.

pweke@uonbi.ac.ke

Abstract

In this paper, we propose a multivariate least squares estimation procedure to estimate Global Vector Autoregressive (GVAR) models and show that it leads to consistent and asymptotically normal estimates of the parameters. We also provide computationally simple conditions that guarantee that the GVAR model is stable.

Key words

Global Vector Autoregressive, GVAR, VARY*

1. Introduction

The GVAR model combines VAR models for several countries. The different VAR models for each country are linked by inclusion of a foreign variable which is constructed as a weighted average of endogeneous variables in other countries. The estimation strategy follows the suggestion of Pesaran, Schuermann and Weiner (2002), hereafter PSW, who estimated the model on a country by country basis ignoring the endogeneity of the foreign variable. The approach in this paper will be based on the argument that as the number of countries in the sample gets smaller ($N \rightarrow 0$), the foreign variable becomes strongly endogeneous. More so, the conditions for ‘weak exogeneity’ as assumed in PSW (2002) might not be satisfied in many empirical settings, e.g. when using trade weights and there remain important trading partners even as the number of countries in the sample increases. Furthermore, in many situations, the asymptotic guidance should be derived keeping the number of countries fixed (N fixed, T→∞).

As a result, it is of interest to be able to estimate the model consistently taking the endogeneity of the foreign variables into account. We provide a relatively simple multivariate least squares estimation procedure and show that it is consistent and asymptotically normal. In the next section we present the model and discuss the conditions under which the GVAR model is stable. Section 3 then outlines the estimation procedure and provides the asymptotic results. Finally, section 4 offers conclusions. Proofs of the claims made in the paper are contained in the appendix.

2. Model

Formulation of each country VAR*Y model

In this model each country will be linked with the others by including foreign-specific variables. Since all countries are modeled individually, then in each country VARY* model, country specific variables are related to deterministic variables such as time trend (t) and a set of country specific foreign variables. Each country will be modeled as a VARY* model as shown below

$$y_{it} = a_{i0} + a_{i1}t + \phi_{i1}y_{i,t-1} + \cdots + \phi_{ip}y_{i,t-p} + \Lambda_{i0}y^*{}_{it} + \Lambda_{i1}y^*{}_{i,t-1} + \cdots + \Lambda_{iq}y^*{}_{i,t-q} + u_{it} \quad \ldots .1$$

Where

$t = 1, \ldots, T$

$i = 1, \ldots, N$

y_{it} is a $k_i \times 1$ vector of country specific domestic variables

$y^*{}_{it}$ is the $k^*{}_i \times 1$ vector of foreign variables specific to country i

ϕ_{ip} is a $k_i \times k_i$ matrix of coefficients associated to lagged domestic variables

Λ_{i0} is a $k_i \times k^*{}_i$ matrices of coefficients related to contemporaneous foreign variables

Λ_{ij} is a $k_i \times k^*{}_i$ matrices of coefficients related to the lagged foreign variables $(j = 1, \dots, q)$

a_{i0} is a $k_i \times 1$ vector of fixed intercepts

a_{i1} is a $k_i \times 1$ vector of coefficients of the deterministic time trend

u_{it} is a $k_i \times 1$ vector of country specific shocks assumed to be serially uncorrelated with a zero mean and a non-singular covariance matrix. Specifically $u_{it} \sim (0, \Sigma_u)$

The domestic variables and foreign variables are grouped as

$$Z_{it} = \left(y_{it}, y^*{}_{it}\right) \dots \dots .2$$

Each country model in (i) is then written as

$$A_{i0} Z_{it} = a_{i0} + a_{i1} t + A_{i1} Z_{i,t-1} + \cdots + A_{i1} Z_{i,t-p} + u_{it} \dots \dots 3$$

where it is assumed that $p = q$ for ease of computation.

In equation (3)

$$A_{i0} = (I_{ki}, -\Lambda_{i0}) \dots \dots 4$$

$$\begin{matrix} A_{i1} = (\phi_{i1}, \Lambda_{i1}) \\ \vdots \\ A_{ip} = (\phi_{ip}, \Lambda_{ip}) \end{matrix} \quad \dots \dots \dots \dots \dots 5$$

And the A_{ip} coefficient matrices are all of size $k_i \times (k_i + k^*{}_i)$. Equation 3 can be treated like a VAR (p) model by multiplying throughout by $A_{i0}{}^{-1}$.

The GVAR model

To examine the endogeneity of the foreign variable y^*_{it}, we need to solve the entire (global) model. Stacking over the countries model can be written as

$$y_t = a_0 + a_1 t + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \Lambda_0 w y_t + \Lambda_1 w y_{t-1} + \cdots + \Lambda_p w y_{t-p} + u_{t \dots\dots 6}$$

Where

$y_t \; is \; Nk \times 1$

a_0 is $Nk \times 1$

a_1 is $Nk \times 1$

$\Phi_1 \ldots \Phi_p$ is $Nk \times Nk$

$y_{t-1} \ldots y_{t-p}$ is $Nk \times 1$

$\Lambda_0, \Lambda_1, \ldots, \Lambda_p$ is $Nk \times Nk$

The solution of the stacked model is obtained as

$$y_t = (I_{kN} - \Lambda_0 w)^{-1}\left(a_0 + a_1 t + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \Lambda_1 w y_{t-1} + \cdots + \Lambda_p w y_{t-p} + u_t\right) \ldots .7$$

Provided the innovations u_t are independent in the time dimension, the endogeneity of the regressors wy_t follows from

$$E(wy_t u_t) = w(I_{kN} - \Lambda_0 w)^{-1} E(u_t u_t') \ldots \ldots \ldots \ldots .8$$

Pesaran et al. (2002) assume that the weight matrices w_{ij} are diagonal with

$w_{ij} = diag\left(w_{ij}^1, \ldots, w_{ij}^k\right)$ and that $\sum_{j=0}^{N}\left(w_{ij}^m\right)^2 \to 0$, as $N \to \infty, for\ all\ i\ and\ m$

However, this implies that asymptotically the foreign variables have no explanatory power in the model. Asymptotic properties of such model should not be used as small sample guidance for our estimators if we actually expect some degree of cross sectional dependence in our model.

The assumption $\sum_{j=0}^{N}\left|w_{ij}^m\right| \leq c < \infty, for\ all\ i\ and\ m$, where the constant c does not depend on the sample size N. This is clearly a weaker assumption but it turns out to be powerful enough to allow us derive asymptotic properties of our model.

Assumptions

The general assumptions that are maintained/applied throughout this section are listed below. The importance of such assumptions is also explained.

Assumption 1

The disturbances u_{it} are generated from

$$u_t = R_{t,N}\ \eta_t$$

Where $\eta_t = (\eta_{1t}, \dots, \eta_{Nt})\ is\ Nk \times 1\ with\ \eta_{it}\ being\ a\ k \times 1\ vector\ of\ innovations$ and

a) The innovations η_{mit} are totally independent (w.r.t $i, t\ and\ m\ indexes$) and have uniformly bounded absolute $4 + \delta$ moments for some $\delta > 0$

b) The sequence of $Nk \times Nk$ matrices $R_{t,N}$ has uniformly bounded absolute row sums i.e. denoting $r_{ij,t,N}$ the ij^{th} element of $R_{t,N}$, it holds that

$$\sum_{j=1}^{Nk} \left| r_{ij,t,N} \right| \leq k_r < \infty$$

Where the constant k_r does not depend on T or N.

This assumption allows for a general heterogeneity structure within a given time period. However, it imposes the restriction that the disturbances at different time periods are independent. Part (a) is a standard restriction required for deriving asymptotic results, while part (b) guarantees that the amount of heterogeneity in the disturbances is asymptotically limited as the number of countries in the sample increases.

Assumption 2

a) The sequence of the weight matrices w has uniformly bounded absolute row and column sums i.e. denoting $w_{ij,qm}$, the (q, m)-th element of w_{ij}, it holds that $\sum_{j=1}^{N} \sum_{m=1}^{k} \left| w_{ij,qm} \right| \leq k_w < \infty$ Where the constant k_w does not depend on $T\ or\ N$ and the choice of indexes $i\ and\ q$. But can partially depend on other parameters of the model.

b) The sequences of matrices $(I_{kN} - \Lambda_0 w)^{-1}$ and $\left[I_{kN} - (I_{kN} - \Lambda_0 w)^{-1} (\Phi_1 + \Lambda_1 w)^{-1} \dots \left(\Phi_p + \Lambda_p w \right)^{-1} \right]^{-1}$ are well defined (the inverses exist) and have uniformly bounded absolute row and column sums.

c) The parameter space is uniformly bounded i.e. the matrices $\Phi_1, \dots, \Phi_p, \Lambda_0, \dots, \Lambda_p$ have uniformly bounded absolute row sums and the vectors $a_0\ and\ a_1$ have elements uniformly bounded in absolute value.

Assumption 2 guarantees that the degree of international interactions in the data does not explode as the sample size (number of countries) increases.

The existence of the inverses in the above assumption will be guaranteed by the following assumptions that imposes stability of the process in both $N \; and \; T$ dimensions (i.e. assumptions 3 and 4)

Notation

Let A be any square $n \times n$ matrix with real entries. We denote its spectral radius as

$\rho(A) \coloneqq max\{|\lambda| \colon \lambda \; is \; an \; eigen \; value \; of \; A\}$

Assumption 3

The spectral radius of $\Lambda_0 w$ is uniformly less than one i.e. $\rho(\Lambda_0 w\ ) \leq k < 1$ where the constant k does not depend on $N \; or \; T$

Assumption 4

The spectral radius of $(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w)$ and of $(I_{kN} - \Lambda_0 w)^{-1}\left(\Phi_1 + \Phi_p + \Lambda_1 w + \Lambda_p w\right)$ are uniformly less than one.

Assumption 5

The initial observations y_o are drawn from $y_0 = R_0 u$

Where

a) The innovations collected in the $Nk \times 1$ vector u are totally independent of each other as well as of innovations $\eta_t \; for \; t > 0$ and the elements of u have uniformly bounded absolute $4 + \delta$ moments for some $\delta > 0$
b) The sequence of $Nk \times Nk$ matrices R_0 has uniformly bounded absolute row sums i.e. $\sum_{j=1}^{Nk} \left| r_{ij,0} \right| \leq k_0 < \infty$ where the constant k_0 does not depend on $N \; and \; T$.

This assumption is about the initial starting values of the process and it will enable us demonstrate that the observable process is a well-defined transformation of the underlying innovations.

The following lemma will be useful in testing the stability of a GVAR model.

Lemma 1

Let A, B and C be square matrices with same dimensions and let $\|A\|$ and $\|B\|$ be less than one for some matrix norm. Then the matrix $S = \sum_{n=0}^{\infty} A^n \, CB^n$ is well defined and

$$vec(S) = [I - (B' \otimes A)]^{-1} vec(C)$$

Furthermore, the finite sum $S_t = \sum_{n=0}^{t} A^n \, CB^n$ can be expressed as

$$S_t = S - A^{t+1} S B^{t+1}$$

Stability conditions of the GVAR process

Inspecting the solution of the global model given in equation 7, it follows that to determine whether the GVAR model is stable, it is not sufficient to examine the stability of the country-by-country models separately, ignoring the endogeneity of y_{it}^* i.e. to examine the Eigen values of $\Phi_i \; and \; \Lambda_i$. Instead the stability of the global model is determined by the spectral radius of

$$(I_{kN} - \Lambda_0 w)^{-1} \left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right) \ldots .9$$

Hence it does not suffice to impose stability of each country model (i.e. require that $\rho(\Phi) < 1$). Accounting for the autocorrelation in the foreign variable i.e. imposing that $\rho\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right) < 1$) is also not sufficient.

Instead, the stability of the process also depends on the strength of the contemporaneous global links in the models (i.e. on the parameters collected in Λ_0) and it must be determined by the spectral radius of the entire matrix

$$(I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p + \Lambda_1 + \cdots + \Lambda_p) \ldots .10$$

In general when both $N\ and\ T$ are allowed to tend to infinity, the claim that this is sufficient is not straight forward and is demonstrated in the following proposition.

Proposition 1

Under the assumptions 1-5, y_t has well defined uniformly bounded $4 + \delta$ moments for some $\delta > 0$. Furthermore, if $a_1 = 0$, then in the limit as $T \to \infty$, y_T converges in quadratic means to a random variable y_∞ which has well defined finite absolute $4 + \delta$ moments for some $\delta > 0$ with

$$E(y_\infty) = \left[I_{kN} - (I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p + \Lambda_1 + \cdots + \Lambda_p)\right]^{-1}(I_{kN} - \Lambda_0 w)^{-1} a_0 \ldots 11$$

If additionally, $\lim_{T\to\infty} E(u_t u_t') = \Omega_u$, we have

$$vech[VC(y_\infty)] = \{I_{N^2k^2} - [A(I_{kN} - \Lambda_0 w)^{-1} \otimes A(I_{kN} - \Lambda_0 w)^{-1}]\}^{-1}.Dvech(\Omega_u) \ldots .12$$

Where

$A = (I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w) \ldots .13$ and D is a duplication matrix such that

$$vec(\Omega_u) = Dvech(\Omega_u)$$

Proposition 2

Assume that the maximum absolute row sums of W are less or equal to k_w, i.e. $\|w\|_1 \leq k_w$. Suppose that

$$\|\Phi\|_1 + k_w \left(\|\Lambda_0\|_1 + \cdots + \|\Lambda_p\|_1\right) < 1$$

Then the spectral radius of $(I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w)$ is less than one.

3. Multivariate least squares estimation

The GVAR framework

We can build a simple version of our GVAR model from each country models represented by equation 1 as follows.

We collect all the domestic variables of all the countries to create the global vector

$$y_t = \begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{Nt} \end{pmatrix} \dots \dots \dots \dots \dots 14$$

Which is a $k \times 1$ vector containing all endogeneous variables, where $k = \sum_{i=1}^{N} k_i$. Following the step that gives rise to equation 48 and the one above, we obtain the identity

$$Z_{it} = w_i y_t \dots \dots \dots \dots .. 15$$

For $i = 1, \dots, N, where\ w_i$ is a country-specific link matrix of dimensions $(k_i + k_i^*) \times k$ constructed on the basis of trade weights. This identity allows writing each country model in terms of the global vector in 78. By substituting 79 in 48, we obtain

$$A_i w_i y_{it} = a_{i0} + a_{i1} t + B_{i1} w_i y_{i,t-1} + \dots + B_{ip} w_i y_{i,t-p} \dots \dots \dots \dots .16$$

The individual country models are then stacked, yielding the model for all the variables in the global model y_t to obtain

$$G y_t = a_0 + a_1 t + \sum_{j=1}^{p} H_j \, y_{t-j} + u_t \dots \dots \dots \dots \dots \dots \dots .. 17$$

Where

$$G = \begin{pmatrix} A_{1,0} w_1 \\ \vdots \\ A_{N,0} w_N \end{pmatrix}, H_j = \begin{pmatrix} A_{1,j} w_1 \\ \vdots \\ A_{N,j} w_N \end{pmatrix}, a_0 = \begin{pmatrix} a_{1,0} \\ \vdots \\ a_{N,0} \end{pmatrix}, a_1 = \begin{pmatrix} a_{1,1} \\ \vdots \\ a_{N,1} \end{pmatrix}, u_t = \begin{pmatrix} u_{1,t} \\ \vdots \\ a_{N,t} \end{pmatrix}$$

Pre-multiplying equation 17 by G^{-1} yields an autoregressive representation of the GVAR (p) model shown below

$$y_t = b_0 + b_1 t + \sum_{j=1}^{p} F_j \, y_{t-j} + \varepsilon_t \dots \dots \dots \dots \dots 18$$

Where

$$F_j = G^{-1} H_j, b_0 = G^{-1} a_0, b_1 = G^{-1} a_1 \ and \ \varepsilon_t = G^{-1} u_t$$

Equation 18 can be treated like any other VAR equation of order p.

Equation 18 can be re-written as

$$y_t = v + F_1 y_{t-1} + \cdots + F_p y_{t-p} + \varepsilon_t \ldots \ldots .. 19$$

The estimator

It is assumed that a time series $y_1, \ldots, y_T$ of the y variables is available, that is, we have a sample of size T for each of the k variables for the same sample period. In addition p pre-sample values for each variable,$y_{-p+1}, \ldots, y_0$ are assumed to be available. Partitioning a multiple time series in sample and pre-sample values is convenient in order to define:-

$$Y = y_1, \ldots, y_T \ , (K \times T)$$

$$B = \left(v, A_1, \ldots, A_p\right), K \times (Kp \times 1)$$

$$Z_t = \begin{bmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{bmatrix}, ((Kp + 1) \times 1\ )$$

$$Z = (Z_0, \ldots, Z_{T-1}), ((Kp + 1) \times T$$

$$U = (u_1, \ldots, u_T)\ , (K \times T) \ldots \ldots \ldots \ldots \ldots \ldots \ldots .. 20$$

$$y = vec\ (Y)\ , KT \times 1$$

$$\beta = vec\ (B)\ , \left((K^2 p + K) \times 1\right)$$

$$b = vec\ (B')\ , \left((K^2 p + K) \times 1\right)$$

$$u = vec\ (U)\ , (KT \times 1)$$

Here vec is the column stacking operator as defined in appendix A5.

Using the notation for $t = 1, \ldots, T$, the GVAR (p) model in (19) can be written compactly as

$$Y = BZ + U \ldots \ldots \ldots .. 21$$

Or

$$vec\ (Y) = vec\ (BZ) + vec\ (U)$$

$$= (Z' \otimes I_K) vec\ (B) + vec\ (U) \ldots \ldots \ldots ..22$$

Or

$$y = (Z' \otimes I_K)\beta + u \ldots \ldots \ldots \ldots \ldots ..23$$

Note that the covariance matrix of u is

$$\sum_u = I_T \otimes \sum_u \ldots \ldots \ldots \ldots \ldots \ldots .24$$

where $\otimes$ is the Kronecker product as defined in appendix A1.

The multivariate LS estimation (or GLS estimation) of β means to choose the estimator that minimizes

$$S(\beta) = u'(I_T \otimes \textstyle\sum_u)^{-1} u = u'(I_T \otimes \textstyle\sum_u^{-1})\, u$$

$$= [y - (Z' \otimes I_K)\beta]'(I_T \otimes \textstyle\sum_u^{-1})\, [y - (Z' \otimes I_K)\beta]$$

$$= vec\ (Y - BZ)'(I_T \otimes \textstyle\sum_u^{-1})\, vec\ (Y - BZ)$$

$$= tr[(Y - BZ)' \textstyle\sum_u^{-1} (Y - BZ)] \ldots \ldots \ldots \ldots \ldots 25$$

In order to find the minimum of this function we note that

$$S(\beta) = y'(I_T \otimes \textstyle\sum_u^{-1})\, y + \beta'(Z \otimes I_K)(I_T \otimes \textstyle\sum_u^{-1})\, (Z' \otimes I_K)\beta - 2\beta'(Z \otimes I_K)(I_T \otimes \textstyle\sum_u^{-1})\, y$$

$$= y'(I_T \otimes \textstyle\sum_u^{-1})\, y + \beta'(ZZ' \otimes \textstyle\sum_u^{-1})\beta - 2\beta'(Z \otimes \textstyle\sum_u^{-1})\, y \ldots \ldots \ldots \ldots \ldots 26$$

Hence

$$\frac{\partial S(\beta)}{\partial \beta} = 2(ZZ' \otimes \textstyle\sum_u^{-1})\beta - 2(Z \otimes \textstyle\sum_u^{-1})\, y \ldots \ldots \ldots \ldots \ldots \ldots ..27$$

Equating to zero gives the normal equations

$$(ZZ' \otimes \textstyle\sum_u^{-1})\beta = (Z \otimes \textstyle\sum_u^{-1})\, y \ldots \ldots \ldots \ldots \ldots .28$$

This implies that

$$\hat{\beta} = ((ZZ')^{-1} \otimes \textstyle\sum_u)\,(Z \otimes \textstyle\sum_u^{-1})\,y$$

$$= ((ZZ')^{-1} Z \otimes I_K)y \ldots\ldots\ldots\ldots\ldots\ldots\ldots..29$$

The Hessian of $S(\beta)$

$$\frac{\partial^2 S}{\partial\beta\partial\beta'} = 2(ZZ' \otimes \sum\nolimits_u^{-1}) \ldots\ldots\ldots\ldots\ldots 30$$

is positive definite which confirms that $\hat{\beta}$ is indeed a minimizing vector. It has to be assumed that ZZ' is non-singular for these results to hold.

The multivariate LS estimator $\hat{\beta}$ obtained above is identical to the OLS estimator obtained by minimizing

$$\bar{S}(\beta) = u'u = [y - (Z' \otimes I_K)\beta]'[y - (Z' \otimes I_K)\beta] \ldots\ldots 31$$

This result is due to Zellner (1962) who showed that GLS and LS estimation in a multiple equation model are identical if the regressors in all equations are the same.

The LS estimator can also be written as

$$\hat{\beta} = ((ZZ')^{-1} Z \otimes I_K)[(Z' \otimes I_K)\beta + u]$$

$$= \beta + ((ZZ')^{-1} Z \otimes I_K)u \ldots\ldots\ldots\ldots\ldots..32$$

Or

$$vec\,(\hat{B}) = \hat{\beta} = ((ZZ')^{-1} Z \otimes I_K) vec\,(Y)$$

$$= vec(YZ'(ZZ')^{-1}) \ldots\ldots\ldots\ldots\ldots\ldots 33$$

Thus

$$\hat{B} = YZ'(ZZ')^{-1}$$

$$= (BZ + U)Z'(ZZ')^{-1}$$

$$= B + UZ'(ZZ')^{-1} \ldots\ldots\ldots 34$$

Another possibility to write the LS estimator is

$$\hat{b} = vec(\hat{B}') = (I_K \otimes (ZZ')^{-1}Z)vec(Y') \dots\dots\dots\dots\dots\dots .35$$

If we let b'_k be the k^{th} row of B, that is, b_k contains all the parameters of the k^{th} equation. Obviously $b' = (b'_1, \dots, b'_k)$.

If we let $y_{(k)} = (y_{k1}, \dots, y_{kT})'$ be the time series available for the k^{th} variable, so that

$$vec(Y') = \begin{bmatrix} y_{(1)} \\ \vdots \\ y_{(K)} \end{bmatrix} \dots\dots\dots\dots\dots 36$$

then $\hat{b}_k = (ZZ')^{-1}Zy_{(K)}$ is the OLS estimator of the model

$$y_{(K)} = Z'^{b_k} + u_{(k)} \dots\dots\dots .37$$

where $u_k = (u_{k1}, \dots, u_{kT})$ and $\hat{b}' = (\hat{b}'_1, \dots, \hat{b}'_k)$.

Comparing equation (35) and equation (37), it is easy to see that the GLS estimator is equivalent to OLS estimator of each of the K equations in model (19) separately.

3.4 Asymptotic properties of GLS estimators

Consistency and asymptotic normality are established if the following results hold:

$\Gamma \coloneqq plim \frac{ZZ\prime}{T} \dots\dots\dots\dots .38$ exist and is non-singular and

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{T} vec\ (u_t Z'_{t-1}) = \frac{1}{\sqrt{T}} vec\ (UZ') = \frac{1}{\sqrt{T}} (Z \otimes I_K)u \xrightarrow{d} N(0, \Gamma \otimes \textstyle\sum_u)\ \ as\ T \to \infty \dots\dots\dots\dots .39$$

where $\xrightarrow{d}$ denotes convergence in distribution.

The theorem due to Mann and Wald (1943) shows that these results are true under suitable conditions for u_t, if y_t is a stable GVAR (p). For instance, the conditions stated in the following definition are sufficient.

Definition 1: standard white noise

A white noise process $u_t = (u_{1t}, \dots, u_{kt})'$ is called standard white noise if the u_t are continous random vectors satisfying $E(u_t) = 0$, $\sum_u = E(u_t u'_t)$ is non-singular, u_t and u_s are independent for $s \neq t$, and, for some finite constant c

$$E\left|u_{it}u_{jt}u_{kt}u_{mt}\right| \leq c \; for \; i,j,k,m = 1, \dots, K \; and \; all \; t \; \dots\dots\dots.40$$

Condition (40) means that all fourth moments exist and are bounded. It is obvious that if the u_t are normally distributed (Gaussian), then they satisfy the moment requirements.

With this definition it is easy to state conditions for consistency and asymptotic normality of the LS estimator. The following lemma is important for proving the large sample results.

Lemma 2

If y_t is a stable, K-dimensional GVAR (p) process in equation (19) with standard white noise residuals u_t then results (38) and (39) hold.

Proof

see theorem 8.2.3 of Fuller (1976, p340)

The Lemma holds also for other definitions of standard white noise. For example, the convergence result in equation (39) follows from a CLT for martingale differences or martingale differences arrays (See proposition D1) by noting that $\omega_t = vec(u_t Z'_{t-1})$ is a martingale difference sequence under quite general conditions. The convergence result in (38) may then be obtained from the weak Law of Large Numbers (see proposition B1).

We now formally state the asymptotic properties of the LS estimator.

Proposition 3:Asymptotic properties of the LS estimator

Let y_t be a stable, K-dimensional GVAR (p) process as in (19) with standard white noise residuals, $\hat{B} = YZ'(ZZ')^{-1}$ is the LS estimator of the GVAR coefficients B. Then

$$plim \; \hat{B} = B \; \dots\dots\dots\dots.41$$

And

$$\sqrt{T}\left(\hat{\beta}-\beta\right)=\sqrt{T}\ vec(\hat{B}-B)\xrightarrow{d}N(0,\Gamma^{-1}\otimes\sum\nolimits_{u})\ \ldots\ldots\ldots..42$$

Or equivalently

$$\sqrt{T}\left(\hat{b}-b\right)=\sqrt{T}vec\left(\hat{B}'-B'\right)\xrightarrow{d}N\left(0,\sum\nolimits_{u}\otimes\Gamma^{-1}\right)\ldots\ldots\ldots\ldots..43$$

Where $\Gamma \coloneqq plim\frac{ZZ\prime}{T}$

3.5 Asymptotic properties of the white noise covariance estimators

In order to assess the asymptotic dispersion of the LS estimator, we need to know the matrices Γ and $\sum_u$. From (38) a consistent estimator for Γ is

$$\hat{\Gamma}=\frac{ZZ'}{T}\ldots\ldots\ldots\ldots\ldots\ldots44$$

Because $\sum_u = E(u_t u'_t)$, a plausible estimator for the covariance matrix is

$$\tilde{\sum}_u=\frac{1}{T}\sum_{t=1}^{T}\hat{u}_t\hat{u}'_t=\frac{1}{T}\hat{U}\hat{U}'=\frac{1}{T}\left(Y-\hat{B}Z\right)\left(Y-\hat{B}Z\right)'$$

$$=\frac{1}{T}[Y-YZ'(ZZ')^{-1}Z][Y-YZ'(ZZ')^{-1}Z]'$$

$$=\frac{1}{T}Y[I_T-Z'(ZZ')^{-1}Z][I_T-Z'(ZZ')^{-1}Z]'$$

$$=\frac{1}{T}Y(I_T-Z'(ZZ')^{-1}Z)Y'\ldots\ldots\ldots\ldots\ldots\ldots\ldots.45$$

An adjustment for degrees of freedom is desired because in a regression with fixed, non-stochastic regressors this leads to an unbiased estimator of the covariance matrix. Thus

$$\hat{\sum}_u=\frac{T}{T-Kp-1}\tilde{\sum}_u\ldots\ldots\ldots\ldots..46$$

may be considered.

Proposition 4: Asymptotic properties of the white noise covariance estimators

Let y_t be a stable, K-dimensional GVAR (p) process as in (19) with standard white noise innovations and let $\bar{B}$ be an estimator of the VAR coefficients B so that $\sqrt{T}vec(\bar{B}-B)$ converges in distribution. Using the symbols from (20) suppose that

$\frac{\overline{\sum_u}=(Y-\bar{B}Z)(Y-\bar{B}Z)'}{T-c}$ … … … … … … … … .47 where c is a fixed constant. Then

$$plim\sqrt{T}\left(\frac{\overline{\sum_u}-UU'}{T}\right)=0 \text{ … … … … … } 48$$

Corollary 1

Under the conditions of proposition 4

$$plim\widehat{\sum}_u = plim\widetilde{\sum}_u = plim\frac{UU'}{T} = \sum\nolimits_u \text{ … … … … } 49$$

Proof

By proposition 3 it suffices to show that $plim\frac{UU'}{T}=\sum_u$, which follows from proposition D3 (4) because

$$E\left(\frac{1}{T}UU'\right)=\frac{1}{T}\sum_{t=1}^{T}E(u_t u'_t)=\sum\nolimits_u \text{ … … … … . .} 50$$

And

$$Var\left(\frac{1}{T}vec(UU')\right)=\frac{1}{T^2}\sum_{t=1}^{T}Var[vec(u_t u'_t)] \leq \frac{T}{T^2}g \to 0 \; as \; T \to \infty \text{ … .} 51$$

where g is the constant upper bound for $Var[vec(u_t u'_t)]$. This bound exists because the fourth moments of u_t are bounded by definition 1. ■

4. Conclusion

Although the endogeneity of the foreign variables is normally taken into account in the empirical implementations of the GVAR models when constructing impulse responses, the researchers have ignored it when estimating the model. This is common to other existing literature. In this paper we have argued that GVAR models should be estimated by assuming the exogeneity of the foreign variable. We have showed that a simple multivariate least squares estimation has desirable asymptotic properties and that it is easily implementable. This paper also provides easy to check stability conditions.

Appendix

A: Vectors and Matrices

A1) Eigen values and -vectors- characteristic values and vectors

The Eigen values or characteristic values or characteristic roots of an $(m \times m)$ square matrix A are the roots of the polynomial in λ given by $\det(A - \lambda I_m)$ or $\det(\lambda I_m - A)$. The determinant is sometimes called the characteristic polynomial of A. A number λ_i is an eigen value of A if the columns of $(A - \lambda I_m)$ are linearly dependent. Consequently there exists a $(m \times 1)$ vector $v_i \neq 0$ such that

$(A - \lambda_i I_m)v_i = 0$ or $Av_i = \lambda_i v_i$

A vector with this property is an eigen vector or characteristic vector of A associated with the eigen value λ_i.

Rule: All Eigen values of the $(m \times m)$ matrix A have modulus less than 1 if and only if $\det(I_m - Az)$ has no roots in and on the complex unit circle.

A2) The Trace

The trace of a $(m \times m)$ square matrix $A = (a_{ij})$ is the sum of its diagonal elements.

$$tr\ A = a_{11} + a_{22} + \cdots + a_{mm}$$

A3) Definite matrices and Quadratic forms

Let A be a symmetric $(m \times m)$ matrix and x an $(m \times 1)$ vector. The function $x'Ax$ is called the quadratic form in x. The symmetric matrix A or the corresponding quadratic form is positive definite if $x'Ax > 0$ for all $m - vectyors\ x \neq 0$.

Rule: $A = (a_{ij})$ is positive definite if and only if its principle minors are positive.

A4) The Kronecker product

Let $A = (a_{ij})$ and $B = (b_{ij})$ be $(m \times n)$ and $(p \times q)$ matrices respectively. The $(mp \times nq)$ matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{bmatrix}$$

is the Kronecker product or direct product of A and B.

Rules:-Assuming suitable dimensions for the matrices

1) $A \otimes B \neq B \otimes A$
2) $(A \otimes B)' = A' \otimes B'$
3) $A \otimes (B + C) = (A \otimes B) + (A \otimes C)$
4) $(A \otimes B)(C \otimes D) = AC \otimes BD$
5) If A and B are invertible, then $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$
6) If A and B are square matrices with eigen values $\lambda_A\ and\ \lambda_B$ respectively and associated eigen vectors $v_A\ and\ v_B$ then $\lambda_A \lambda_B$ is an eigen value of $A \otimes B$ with eigen vector $v_A \otimes v_B$
7) If A and B are $(m \times m)$ and $(n \times n)$ square matrices respectively, then $|A \otimes B| = |A|^n |B|^m$
8) If A and B are square matrices $tr\ (A \otimes B) = tr(A) tr(B)$

A5) The vec and vech operators and related matrices

- **the operators**

Let $A = (a_1, \dots, a_n)$ be an $(m \times n)$ matrix with $(m \times 1)$ columns a_i. The vec operator transforms A into an $(mn \times 1)$ vector by stacking the columns, that is

$$vec\ (A) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}$$

Rules:-let A, B and C be matrices with appropriate dimensions then

1) $vec\ (A + B) = vec(A) + vec(B)$
2) $vec(ABC) = (C' \otimes A)vec(B)$
3) $vec\ (AB) = (I \otimes A)vec(B) = (B' \otimes I)vec(A)$
4) $vec(ABC) = (I \otimes AB)vec\ (C) = (C'B' \otimes I)vec(A)$
5) $vec(B')vec(A) = tr(BA) = tr(AB) = vec(A')'vec(B)$
6) $tr(ABC) = vec(A')'(C' \otimes I)vec(B)$

$$= vec(A')'(I \otimes B)vec\ (C)$$
$$= vec\ (B')'(A' \otimes I)vec(C)$$
$$= vec\ (B')'(I \otimes C)vec\ (A)$$
$$= vec(C')'(B' \otimes I)vec(A)$$
$$= vec(C')'(I \otimes A)vec\ (B)$$

The vech operator stacks the elements on and below the main diagonal of a square matrix.

B: Stochastic convergence and asymptotic distributions

B1) Convergence in probability and in distribution

Let $x_{1,}, x_2, \ldots$ or $\{x_T\}, T = 1,2, \ldots$ be a sequence of scalar random variables which are all defined on a common probability space $(\Omega, \mathcal{F}, \Pr)$. The sequence $\{x_T\}$ converges in probability to the random variable x (which is also defined on $(\Omega, \mathcal{F}, \Pr)$) if for every $\varepsilon > 0$

$$\lim_{T \to \infty} \Pr(|x_T - x| > \varepsilon) = 0$$

Or equivalently

$$\lim_{T \to \infty} \Pr(|x_T - x| < \varepsilon) = 1$$

This type of stochastic convergence is abbreviated as

$plim x_T = x$ or $x_T \xrightarrow{p} x$

The sequence $\{x_T\}$ converges almost sure (a.s) or with probability one to the random variable x if for every $\varepsilon > 0$

$$\Pr(\lim_{T \to \infty} |x_T - x| < \varepsilon) = 1$$

This type of convergence is often written as $x_T \xrightarrow{a.s} x$ and is sometimes called strong convergence.

Denoting the distribution functions of x_T and x by F_T and F respectively, the sequence $\{x_T\}$ is said to converge in distribution or weakly or in law to x, if for all real numbers c for which F is continous

$$\lim_{T \to \infty} F_T(c) = F(c)$$

This type of convergence is abbreviated as $x_T \xrightarrow{d} x$.

These concepts of stochastic convergence can be extended to sequences of random vectors (multivariate random variables)

Suppose $\{x_T = (x_{1T}, \dots, x_{KT})'\}, T = 1,2, \dots$ is a sequence of K-dimensional random vectors and $x = (x_1, \dots, x_K)'$ is a K-dimensional random vector. Then

$plim x_T = x$ or $x_T \xrightarrow{p} x$ if $plim x_{kT} = x_k$ for $k = 1, \dots, K$

$x_T \xrightarrow{a.s} x$ if $x_{kT} \xrightarrow{d} x_k$ for $k = 1, \dots, K$

$x_T \xrightarrow{d} x$ if $lim F_T(c) = F(c)$ for all continuity points of F.

In this case F_T and F are the joint distribution functions of x_T and x respectively.

B2) Law of Large Numbers (LLN) and Central Limit Theorems (CLT)

Suppose $\{x_t\}, t = 1,2, \ldots$ is a sequence of zero mean random variables and let Ω_t be an information set available at time t which includes at least $\{x_1, \ldots, x_t\}$ and possibly other random variables. The sequence $\{x_t\}$ is said to be a martingale difference sequence with respect to the sequence Ω_t if $E(x_t/\Omega_{t-1}) = 0$ for all $t = 2,3, \ldots$

It is simply referred to as a martingale difference sequence if $E(x_t) = 0$ for $t = 1,2, \ldots$ and $E(x_t/x_{t-1}, \ldots, x_1) = 0 \; for \; t = 2,3, \ldots$

More generally, a sequence $\{x_t\}$ of K-dimensional vector random variables satisfying $E(x_t) = 0$ for all t and $E(x_t/x_{t-1}, \ldots, x_1) = 0 \; for \; t = 2,3, \ldots$ is a vector martingale difference sequence.

Proposition B1: law of large numbers (LLN)

1) LLN for martingale differences
   Let $\{x_t\}$ be a strict stationary martingale difference sequence with $E|x_t| < \infty$ t=1,2,…, then $\bar{x}_T \xrightarrow{p} 0$
2) LLN for martingale difference arrays
   Let $\{x_t\}$ be a martingale difference array with $E\left|x_{T,t}\right|^{1+\varepsilon} \leq c < \infty \; for \; all \; t \; and \; T$ for some $\varepsilon > 0$ and a finite constant c. Then

$$\bar{x}_T := T^{-1} \sum_{t=1}^{T} x_{T,t} \xrightarrow{p} 0$$

Proposition B2: Central Limit Theorem (CLT)

1) CLT for martingale difference arrays
   Let $\left\{x_{T,t} = (x_{1T,t}, \ldots, x_{kT,t)})'\right\}$ be a K-dimensional martingale difference array with covariance matrices
   $E\left(x_{T,t}x'_{T,t}\right) = x_{Tt}$ such that $T^{-1} \sum_{t=1}^{T} \sum_{Tt} \to \sum, \; where \; \sum \; is \; positive \; definite$
   Moreover suppose that $T^{-1} \sum_{t=1}^{T} x_{T,t}x'_{T,t} \xrightarrow{p} \sum \; and \;\; E(x_{iT,t}x_{jT,t}x_{kT,t}x_{lT,t}) < \infty$ for all $t \; and \; T$ and all $1 \leq i, j, k, l \leq K$. Then

$$\sqrt{T}\bar{x}_T \xrightarrow{d} N(0, \sum)$$

2) CLT for stationary processes

Let $x_t = \mu + \sum_{j=0}^{\infty} \Phi_j u_{t-j}$ be a K-dimensional stationary stochastic process with $E(x_t) = \mu < \infty$, $\sum_{j=0}^{\infty} \|\Phi_j\| < \infty$ and $u_t \sim (0, \Sigma_u)$ iid white noise. Then

$$\sqrt{T}(\bar{x}_T - \mu) \xrightarrow{d} N(0, \sum_{j=-\infty}^{\infty} \Gamma_x(j))$$

where

$$\Gamma_x(j) \coloneqq E\left[(x_t - \mu)(x_{t-j} - \mu)'\right]$$

C: STOCHASTIC INEQUALITIES

Chebyshev

Let X be a non-negative random variable with finite mean μ_X and finite variance σ_X^2. Then, for any $\varepsilon \in \boldsymbol{R}, \varepsilon > \mathbf{0}$

$$P(|X - \mu_X|) > \sqrt{\frac{\sigma_X^2}{\varepsilon}} \leq \varepsilon$$

Holder

Let X and Y be random variables and let $p, q \in \boldsymbol{R}, p > 1, \frac{1}{p} + \frac{1}{q} = 1$. Then

$$E(|XY|) \leq [E(|X|^p)]^{\frac{1}{p}}[E(|Y|^q)]^{\frac{1}{q}}$$

Cauchy-Schwartz

For p=q=2, we have

$$E(|XY|) \leq \sqrt{E|X|^2}\sqrt{E|Y|^2}$$

Lyapunov

For Y=1, we have for $p > 1$

$$E(|X|) \leq [E(|X|^p)]^{\frac{1}{p}}$$

Minkowski

If for some $p \geq 1, E(|X|^p) < \infty\ and\ E(|Y|^p) < \infty$, then

$$E(|X+Y|) \leq [E(|X|^p)]^{\frac{1}{p}}[E(|Y|^p)]^{\frac{1}{p}}$$

Jensen

Let X be a random variable and $f : D \subseteq R \rightarrow R$ be a convex realfunction. Then

$$f\,[E\,(X)] \leq E\,[f\,(X)]$$

By selecting the random variables to be constants, the above inequalities can be applied in the deterministic case as well.

Since the mean of a finite number of non-random variables in **R** may be considered as mathematical expectations, it follows from Holder's inequality that for real numbers $x_i, y_i, p > 1, \frac{1}{p} + \frac{1}{q} = 1$

$$\left|\sum_{i=1}^{m} x_i y_i\right| \leq \left(\sum_{i=1}^{m} |x_i|^p\right)^{\frac{1}{p}} \left(\sum_{i=1}^{m} |y_i|^q\right)^{\frac{1}{q}}$$

Similarly, from Lyapunov's inequality

$$\left|\sum_{i=1}^{m} x_i\right|^p \leq m^{p-1} \sum_{i=1}^{m} |x_i|^p, p \geq 1$$

And by Minkowski's inequality

$$\left|\sum_{i=1}^{m} x_i + y_i\right|^{\frac{1}{p}} \leq \left(\sum_{i=1}^{m} |x_i|^p\right)^{\frac{1}{p}} + \left(\sum_{i=1}^{m} |y_i|^q\right)^{\frac{1}{q}}$$

Note, if $x_i\ and\ y_i$ are random variables, then the last three inequalities hold for all their realizations. As a result we can apply these inequalities also in cases where $x_i\ and\ y_i$ are stochastic. The same also holds for triangle inequality.

D: STOCHASTIC CONVERGENCE AND ASYMPTOTIC DISTRIBUTIONS

D1: Concepts of stochastic convergence

Let $x_{11}, x_2, \dots$ or $\{x_T\}, T = 1,2, \dots$ be a sequence of scalar random variables which are all defined on a common probability space $(\Omega, \mathcal{F}, Pr)$. The sequence $\{x_T\}$ converges in probability to the random variable x (which is also defined on $(\Omega, \mathcal{F}, Pr)$ if for every $\in > 0$

$$\lim_{T\to\infty} Pr(|x_T - x| > \in) = 0$$

Or equivalently

$$\lim_{T\to\infty} Pr(|x_T - x| < \in) = 1$$

This type of stochastic convergence is abbreviated as

$$plim x_T = x \; or \; x_T \overset{p}{\to} x$$

The limit x may be a fixed, non-stochastic real number which is then regarded as a degenerate random variable that takes on one particular value with probability one.

The sequence $\{x_T\}$ converges almost surely (a.s) or with probability one to the random variable x if for every $\in > 0$

$$Pr\left(\lim_{T\to\infty} |x_T - x| < \in\right) = 1$$

This type of stochastic convergence is often written as $x_T \overset{a.s}{\to} x$ and is sometimes known as strong convergence.

The sequence $\{x_T\}$ converges in quadratic mean or mean square error to x, written briefly as $x_T \overset{q.m}{\to} x$, if

$$\lim_{T\to\infty} E(x_T - x)^2 = 0$$

This type of convergence requires that the mean and variance of the $x_T's$ and x exist.

Denoting the distribution functions of x_T *and* x by F_T *and* F, *respectively*, the sequence $\{x_T\}$ is said to converge in distribution or weakly or in law to x, if for all real numbers c for which F is continuous

$$\lim_{T\to\infty} F_T(c) = F(c)$$

This type of convergence is abbreviated as $x_T \overset{d}{\to} x$.

All these concepts of stochastic convergence can be extended to sequences of random vectors (multivariate random variables). Suppose $\{x_T = (x_{1T}, \dots, x_{KT})'\}, T = 1,2, \dots$, is a sequence of K-dimensional random vectors and $x = (x_1, \dots, x_K)'$ is a K-dimensional random vector. Then the following dimensions are used

$$plim x_T = x \text{ or } x_T \overset{p}{\to} x \text{ if } plim x_{kT} = x_k \text{ for } k = 1, \dots, K$$

$$x_T \overset{a.s}{\to} x \text{ if } x_{kT} \overset{a.s}{\to} x_k \text{ for } k = 1, \dots, K$$

$$x_T \overset{q.m}{\to} x \text{ if } lim E[(x_T - x)'(x_T - x)] = 0$$

$$x_T \overset{d}{\to} x \text{ if } lim F_T(c) = F(c) \text{ for all continuity points of } F.$$

Here F_T and F are the joint distribution functions of x_T *and* x *respectively*.

Proposition D1: convergence properties of sequences of random variables

Suppose $\{x_T\}$ is a sequence of K-dimensional random variables. Then the following relations hold:

1. $x_T \overset{a.s}{\to} x \Rightarrow x_T \overset{p}{\to} x \Rightarrow x_T \overset{d}{\to} x$
2. $x_T \overset{q.m}{\to} x \Rightarrow x_T \overset{p}{\to} x \Rightarrow x_T \overset{d}{\to} x$
3. If x is a fixed non-stochastic vector, then
   $x_T \overset{q.m}{\to} x \Leftrightarrow [lim E(x_T) = x \text{ and } lim E\{(x_T - Ex_T)'(x_T - Ex_T)\} = 0]$

4. If x is a fixed, non-stochastic random vector, then $x_T \overset{p}{\to} x \Rightarrow x_T \overset{d}{\to} x$
5. Slutsky's theorem

   If $g: R^K \to R^m$ is a continuous function, then $x_T \overset{p}{\to} x \Rightarrow g(x_T) \overset{p}{\to} g(x)$ $[plimg(x_T) = g(plimx_T)]$

   $x_T \overset{d}{\to} x \Rightarrow g(x_T) \overset{d}{\to} g(x)$ and $x_T \overset{a.s}{\to} x \Rightarrow g(x_T) \overset{a.s}{\to} g(x)$

Proposition D2: properties of convergence in probability and in distribution

Suppose $\{x_T\}$ and $\{y_T\}$ are sequences of $K \times 1$ random vectors, $\{A_T\}$ is a sequence of $K \times K$ random matrices, x is a fixed $K \times K$ matrix

1. If $plimx_T, plimy_T\ and\ plimA_T$ exist, then
   a) $plim(x_T \pm y_T) = plimx_T \pm plimy_T$
   b) $plim(c'x_T) = c'(plimx_T)$
   c) $plimx_T^{\prime y_T} = (plimx_T)'(plimy_T)$
   d) $plimA_Tx_T = plim(A_T)plim(x_T)$
2. If $x_T \overset{d}{\to} x$ and $plim(x_T - y_T) = 0, then\ y_T \overset{d}{\to} x$
3. If $x_T \overset{d}{\to} x$ and $plimy_T = c$, then

   a) $x_T \pm y_T \overset{d}{\to} x \pm c$

   b) $y_T{}'x_T \overset{d}{\to} c'x$
4. If $x_T \overset{d}{\to} x$ and $plimA_T = A,\ then\ A_Tx_T \overset{d}{\to} Ax$
5. If $x_T \overset{d}{\to} x$ and $plimA_T = 0, then\ plimA_Tx_T = 0$

Proposition D3: weak laws of large numbers

1. Khinchen's theorem (Rao, 1973 p112)

   Let$\{x_t\}$ be asequence of iid random variables with $E(x_t) = \mu < \infty$. Then

$$x_T \coloneqq \frac{1}{T}\sum_{t=1}^{T} x_T \overset{p}{\to} \mu$$

2. Let $\{x_t\}$ be a sequence of independent random variables with $E(x_t) = \mu < \infty$ and $E|x_t|^{1+\epsilon} \leq c < \infty \; (t = 1,2,\ldots) for\ some\ \epsilon > 0$ and a finite constant c. Then $\bar{x}_T \xrightarrow{p} \mu$
3. Chebyshev's theorem (Rao, 1973 p112)
   Let $\{x_t\}$ be a sequence of uncorrelated random variables with $E(x_t) = \mu < \infty$ and $lim E(\bar{x}_T - \mu)^2 = 0 \; then\ \bar{x}_T \xrightarrow{p} \mu$
4. Corollary to Chebyshev's theorem
   Let $\{x_t\}$ be a sequence of independent random variables with $E(x_t) = \mu < \infty \; var(x_t) \leq c < \infty \; (t = 1,2,\ldots) for\ some\ finite\ constant\ c, then\ \bar{x}_T \xrightarrow{p} \mu$.

E: Proof of claims

Proof of proposition 1

Given assumption 3, the matrix $(I_{kN} - \Lambda_0 w)$ is invertible (cf lemma 5.6.10 and corollary 5.6.16 in Horn and Johnson (1985) and the endogeneous variable y_t can be expressed as in equation 7, that is,

$$y_t = (I_{kN} - \Lambda_0 w)^{-1}\left(a_0 + a_1 t + \Phi_1 y_{t-1} + \cdots + \Phi_p y_{t-p} + \Lambda_1 w y_{t-1} + \cdots + \Lambda_p w y_{t-p} + u_t\right)$$

By backward substitution, we then obtain

$$y_t = b_{1t} + b_{2t} + b_{3t} + b_{4t}$$

Where

$$b_{1t} = \sum_{s=0}^{t-1} [A]^s \, (I_{kN} - \Lambda_0 w)^{-1} a_0$$

$$b_{2t} = \sum_{s=0}^{t-1} [A]^s \, (I_{kN} - \Lambda_0 w)^{-1} a_1 s$$

$$b_{3t} = \sum_{s=0}^{t-1} [A]^s \, (I_{kN} - \Lambda_0 w)^{-1} u_s$$

$$b_{4t} = [A]^t y_0$$

By assumption (2b), b_{1t} and b_{2t} have elements uniformly bounded in absolute value. We demonstrate that the sequences of the stochastic vectors b_{3t} and b_{4t} have elements with uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. The claim in the proposition will then follow from Minkowski inequality (appendix C).

Consider the stochastic term $b_{3t} = \sum_{s=0}^{t-1} [A]^s \, (I_{kN} - \Lambda_0 w)^{-1} u_s$

Note that given assumption 1, the random vector η_t and the sequence of matrices $R_{t,N}$ satisfy the conditions of lemma B2 in Mutl (2006). Therefore the elements of the random vector u_t have uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. From assumption 2, we have that the absolute row sums of $A^s (I_{kN} - \Lambda_0 w)^{-1}$ are uniformly bounded in absolute value. Hence by repeated application of the lemma B2 in Mutl (2006) we have that $A^s (I_{kN} - \Lambda_0 w)^{-1} u_s$ has elements with uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. By Minikowski inequality, we then have that b_{3t} has elements with uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$.

Next, consider the stochastic term $b_{4t} = [A]^t y_0$.

Again by assumption 2, the matrix A^t has uniformly bounded absolute row sums and hence given assumption 5, we have by the same lemma B2 that the elements of b_{4t} have uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$.

We now turn to the asymptotic moments of y_t as $t \to \infty$, assuming that $a_1 = 0$. Using lemma A1 and theorem 5.6.12 in Horn and Johnson, it follows that b_{1t} converges to

$$b_1 = \lim_{t\to\infty} b_{1t} = \lim_{t\to\infty} (I_{kN} - A)^{-1} \, (I_{kN} - A^t)^{-1} (I_{kN} - \Lambda_0 w)^{-1} a_0$$

$$= (I_{kN} - A)^{-1} (I_{kN} - \Lambda_0 w)^{-1} a_0$$

Given assumption (2b), it follows that b_1 has elements uniformly bounded in absolute value and it suffices to show that the elements of b_{3t} and b_{4t} converge in quadratic means to random variables b_3 and b_4 with finite $4+\delta$ moments for some $\delta > 0$. By assumption 1, the elements of b_{3t} are independent of the elements of b_{4t}.

Denote the matrix $B_{3s} = A^s(I_{kN} - \Lambda_0 w)^{-1} R_s$ and note that from assumption 1 and 2b it follows that

$$\sum_{s=0}^{\infty} \|B_{3s}\|_1 \leq \|A^s(I_{kN} - \Lambda_0 w)^{-1}\|_1 . k_r$$

$$\leq \|A^s\|_1 . k_1 k_r = \|(I_{Nk} - A)^{-1}\|_1 . k_1 k_r$$

$$\leq k_2 k_1 k_r < \infty$$

Where k_r is the uniform bound for absolute row sums of matrices R_t, k_1 *and* k_2 are uniform bounds for absolute row sums of matrices $(I_{kN} - \Lambda_0 w)^{-1}$ and $(I_{kN} - A)^{-1}$.

Given assumption 1, the elements of b_{3t} satisfy conditions of lemma B1 in Mutl (2006) and hence converge in quadratic means to a random variable with uniformly bounded absolute $4 + \delta$ moments for some $\delta > 0$.

Finally, note that from assumption 4 and theorem 5.6.12, it follows that

$$\lim_{t \to \infty} A^t = 0$$

And hence given assumption 5, we have that the elements of b_{4t} converge in quadratic means to zero.

Therefore, the random variable y_∞ is well defined and we have

$$y_\infty = \lim_{t \to \infty} B_{0t} a_0 + \sum_{s=0}^{\infty} A^s (I_{kN} - \Lambda_0 w)^{-1} u_s$$

$$= (I_{kN} - A)^{-1}(I_{kN} - \Lambda_0 w)^{-1} a_0 + \sum_{s=0}^{\infty} A^s (I_{kN} - \Lambda_0 w)^{-1} u_s$$

Hence

$$E(y_\infty) = (I_{kN} - A)^{-1}(I_{kN} - \Lambda_0 w)^{-1} a_0$$

And using the independence of u_t *and* u_s *for* $t \neq s$

$$VC(y_\infty) = \sum_{s=0}^{\infty} A^s (I_{kN} - \Lambda_0 w)^{-1} \Omega_u (I_{kN} - w' \Lambda_0')^{-1} A^{s'}$$

Using lemma 1, we find that

$$vech[VC(y_\infty)] = \{I_{N^2k^2} - [A(I_{kN} - \Lambda_0 w)^{-1} \otimes A(I_{kN} - \Lambda_0 w)^{-1}]\}^{-1} D.vech(\Omega_u)$$

Where D is a duplication matrix.

We now examine the sufficient conditions for stability in more details. Note that, for any matrix norm, the spectral radius $\rho(A)$ is smaller than the norm $\|A\|$ (Horn and Johnson, 1985), hence using the sub-multiplicative property of the matrix norm, we have that

$$\rho[(I_{kN} - \Lambda_0 w)^{-1}\Phi] \leq \left\|(I_{kN} - \Lambda_0 w)^{-1}\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|$$

$$\leq \|(I_{kN} - \Lambda_0 w)^{-1}\|.\left\|\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|$$

Note that from assumption 3 and lemma 5.6.10 in Horn and Johnson (1985), we have by corollary 5.6.16 in Horn and Johnson (1985) that the inverse $(I_{kN} - \Lambda_0 w)^{-1}$ can be expanded as an infinite sum. Therefore, any norm of $(I_{kN} - \Lambda_0 w)^{-1}$ can be bounded above by

$$\|(I_{kN} - \Lambda_0 w)^{-1}\| \leq \sum_{s=0}^{\infty}(\|w\|.\|\Lambda_0\|)^s$$

Often, the weight matrices are row normalized. In this case we have that $\|w\|_1 = 1$ and hence

$$\|(I_{kN} - \Lambda_0 w)^{-1}\|_1 \leq \sum_{s=0}^{\infty}\|\Lambda_0\|_1^s$$

$$= \frac{1}{1 - \|\Lambda_0\|_1}$$

$$= \frac{1}{1 - max_{1\leq i\leq N}\{\|\Lambda_{i0}\|_1\}}$$

To satisfy assumption 3 (in the case of $\|w\|_1 = 1$) we can, for example, require that $0 \leq max_{1\leq i\leq N}\{\|\Lambda_{i0}\|_1\} < 1$. However, if there are global feedbacks in the model we have $max_{1\leq i\leq N}\{\|\Lambda_{i0}\|_1\} > 0$ and hence

$$\frac{1}{1 - max_{1\leq i\leq N}\{\|\Lambda_{i0}\|_1\}} > 1$$

In this case the requirement that $\left\|\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|_1 < 1$ which is a stronger requirement than $\rho\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right) < 1$ does not necessarily guarantee that the process is stable. This is due to the fact that the requirement $\left\|\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|_1 < 1$ is a sufficient condition for $\rho\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right) < 1$.

The following proposition provides a sufficient condition under which the process is stable.

Proof of proposition 2

Observe by equation $\|(I_{kN} - \Lambda_0 w)^{-1}\| \leq \sum_{s=0}^{\infty}(\|w\|.\|\Lambda_0\|)^s$ (in the proof above) and the assumption in the proposition we have that

$$\rho\left[(I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p)\right] \leq \|(I_{kN} - \Lambda_0 w)^{-1}\|_1 \left\|\left(\Phi_1 + \cdots + \Phi_p + \Lambda_1 w + \cdots + \Lambda_p w\right)\right\|_1$$

$$\leq [\textstyle\sum_{s=0}^{\infty}(k_w\|\Lambda_0\|_1)^s]\left[\left\|\Phi_1 + \cdots + \Phi_p\right\|_1 k_w \left\|\Lambda_1 + \cdots + \Lambda_p\right\|_1\right]$$

$$= \frac{\left\|\Phi_1 + \cdots + \Phi_p\right\|_1 + k_w\left\|\Lambda_1 + \cdots + \Lambda_p\right\|_1}{1 - k_w\|\Lambda_0\|_1}$$

Next, note that from the condition in the proposition

$(\left\|\Phi_1 + \cdots + \Phi_p\right\|_1 + k_w\left\|\Lambda_1 + \cdots + \Lambda_p\right\|_1 + k_w\|\Lambda_0\|_1 < 1)$ it follows that

$$\left\|\Phi_1 + \cdots + \Phi_p\right\|_1 + k_w\left\|\Lambda_1 + \cdots + \Lambda_p\right\|_1 < 1 - k_w\|\Lambda_0\|_1$$

And thus, observe that the condition also implies that

$$k_w\|\Lambda_0\|_1 < 1, \textit{thus also } 1 - k_w\|\Lambda_0\|_1 > 0$$

Hence,

$\frac{\left\|\Phi_1 + \cdots + \Phi_p\right\|_1 + k_w\left\|\Lambda_1 + \cdots + \Lambda_p\right\|_1}{1 - k_w\|\Lambda_0\|_1} < 1$ which proves the claim.

The above proposition provides a simpler alternative to checking the eigen values of the entire matrix $(I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p + \Lambda_1 + \cdots + \Lambda_p)$. Note that, when the weights are normalized to add up to one, we have $k_w = 1$ and it suffices to check whether for all country models it holds that the row sums of

$|\Phi_1| + \cdots + |\Phi_p| + |\Lambda_{io}| + |\Lambda_{i1}| + \cdots + |\Lambda_{ip}|$ are less than one.

However, the above proposition provides only a sufficient condition for stability. Necessary condition is that the spectral radius of

$(I_{kN} - \Lambda_0 w)^{-1}(\Phi_1 + \cdots + \Phi_p + \Lambda_1 + \cdots + \Lambda_p)$ is less than one.

Proof of proposition 3

Using equation (34)

$$plim\,(\hat{B} - B) = plim\,(\frac{UZ'}{T})plim(\frac{ZZ'}{T})^{-1} = 0$$

By Lemma 2, because 39 implies $plim\ \frac{UZ'}{T} = 0$. Thus the consistency of $\hat{B}$ is established.

Using equation (32)

$$\sqrt{T}(\hat{\beta} - \beta) = \sqrt{T}((ZZ')^{-1}Z \otimes I_K)u$$

$$= ((\frac{1}{T}ZZ')^{-1} \otimes I_K)\frac{1}{\sqrt{T}}\ (Z \otimes I_K)u \ldots\ldots\ldots\ldots 32$$

By proposition D2 (4) of appendix D, $\sqrt{T}(\hat{\beta} - \beta)$ has the same asymptotic distribution as

$$\left[plim(\frac{1}{\sqrt{T}}ZZ')^{-1} \otimes I_K\right]\frac{1}{\sqrt{T}}(Z \otimes I_K)u = (\Gamma^{-1} \otimes I_K)\frac{1}{\sqrt{T}}(Z \otimes I_K)u \ldots\ldots 33$$

Hence the asymptotic distribution of $\sqrt{T}(\hat{\beta} - \beta)$ is normal by Lemma 2 and the covariance matrix is

$(\Gamma^{-1} \otimes I_K)(\Gamma \otimes \sum_u)\ (\Gamma^{-1} \otimes I_K) = \Gamma^{-1} \otimes \sum_u \ldots\ldots\ldots 34$ ■

Proof of proposition 4

$$\frac{1}{T}(Y - \bar{B}Z)(Y - \bar{B}Z)' = (B - \bar{B})\left(\frac{ZZ'}{T}\right)(B - \bar{B})' + (B - \bar{B})\frac{ZU'}{T} + \frac{UZ'}{T}(B - \bar{B})' + \frac{UU'}{T}$$

Under the conditions of the proposition, $plim(B - \bar{B}) = 0$. Hence by Lemma 2

$plim(B - \bar{B})\frac{ZU'}{\sqrt{T}} = 0 \ldots\ldots\ldots\ldots 41$ and $plim\left[(B - \bar{B})\frac{ZZ'}{T}\sqrt{T}(B - \bar{B})'\right] = 0 \ldots\ldots 42$ (see appendix C1)

Thus $plim\sqrt{T}\left[\frac{(Y-\bar{B}Z)(Y-\bar{B}Z)'}{T}-\frac{UU'}{T}\right]=0$ … … … … … .43

The proposition follows by noting that as $T \to \infty$, then $\frac{T}{T-c} \to 1$

The proposition covers both $\hat{\Sigma}_u$ *and* $\tilde{\Sigma}_u$ This implies that both $\hat{\Sigma}_u$ *and* $\tilde{\Sigma}_u$ have the same asymptotic properties as the estimator

$$\frac{UU'}{T}=\frac{1}{T}\sum_{t=1}^{T}u_t u'_t$$

which is based on the unknown true residuals and therefore not feasible in practice.

In particular, if $\sqrt{T}vec(\frac{UU'}{T}-\Sigma_u)$ converges in distribution,

$\sqrt{T}vec(\hat{\Sigma}_u-\Sigma_u)$ *and* $\sqrt{T}vec(\tilde{\Sigma}_u-\Sigma_u)$ will have the same limiting distribution (see proposition D2 of appendix D1).

References

Horn R. A and C. R. Johnson (1985). Matrix Analysis. Cambridge: Cambridge University Press

Mann H.B. and Wald A. (1943). “On the statistical treatment of linear difference equations, Econometrica, vol. 11, pp.173-220.

Mutl J. (2006). Dynamic Panel Data Models with Spatially Autocorrelated Disturbances. Ph.D. thesis, University of Maryland, College Park.

Pesaran H. M., Shuermann T. and Weiner S. M. (2002). Modelling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model. Wharton Working Paper Series, 01-38-B

Rao M. 1973. *Abstract Martingales and Ergodic Theory*. Multivariate Analysis.

Zellner A. (1962): An Efficient Method of Estimating Seemingly Unrelated Regression Equations and Tests of Aggregation Bias, Journal of the American Statistical Association, 57, 500-509.