Consistent Estimation of Global VAR Models
.
PRELIMINARY DRAFT, PLEASE DO NOT QUOTE

Jan Mutl*
Johann Wolfgang Goethe Universität
Box 79, Mertonstrasse 17
60054 Frankfurt am Main
Germany

February 14, 2007

Abstract

In this paper, I propose an instrumental variable (IV) estimation procedure to estimate global VAR (GVAR) models and show that it leads to consistent and asymptotically normal estimates of the parameters. I also provide computationally simple conditions that guarantee that the GVAR model is stable. Finally, I illustrate the procedure using both real and artificial data and document the extent of the endogeneity bias that is present in the estimation procedure commonly utilized in the literature.

JEL classification: C31, C32, C33

Keywords: Global VAR, GVAR, consistent estimation, instrumental variables.

*Email for correspondence: mutl@wiwi.uni-frankfurt.de

1 Introduction[†]

Vector autoregressions (VAR) have become a part of a standard tool box of any empirical economist. VARs are used for both model estimation and evaluation, as well as for a-theoretical data analysis (establishments of so called 'stylized facts'). In practical applications in macroeconomics, VAR models are often estimated using data for a particular cross-sectional unit (typically a country), ignoring any possible international linkages. In fact, when international linkages are present, the standard estimates will be biased (omitted variable bias). The literature that combines several VARs into a panel VAR model[1] assumes that the regressors do not include any contemporaneous endogenous variables and hence also suffers from the same criticism.

As an answer to these challenges, there is a growing volume of empirical literature that combines VAR models for several countries into so called global VAR (GVAR) model.[2] The different VAR models for each country are linked by inclusion of a foreign variable which is constructed as a weighted average of endogenous variables in other countries. The estimation strategy follows the suggestion of Pesaran, Schuermann and Weiner (2002) and estimate the model on a country-by-country basis ignoring the endogeneity of the foreign variable. This approach is based on the argument that as the number of countries in the sample grows ($N \to \infty$), the foreign variable becomes 'weakly exogenous'. In this paper I argue that:

- The 'weak exogeneity' concept leads to asymptotic results that do not serve as a useful small sample guidance.
- The 'weak exogeneity' might not be satisfied in many empirical settings.
- In many situations, the asymptotic guidance should be derived keeping the number of countries fixed (N fixed, $T \to \infty$).

As a result, it becomes necessary to estimate the model consistently taking the endogeneity of the foreign variables into account. I provide a relatively simple instrumental variable procedure and show that it is consistent

[†] I thank ...

[1] See e.g. Binder, Pesaran and Hsiao (2002), or Binder, Mutl, Pesaran and Hsiao (2002).

[2] Pesaran Schuermann and Weiner (2002), Pesaran, Smith and Smith (2005), Dees, di Maurio, Pesaran and Smith (2004), Pesaran and Smith (2006) to mention a few.

and asymtotically normal. I also examine the extent of endogeneity bias using a small Monte Carlo study as well as real data.

In the next section I present the model, explicitly state assumptions under which I derive the large sample results and discuss the conditions under which the GVAR model is stable. Section 3 then outlines the estimation procedure and provides the asymptotic results. Section 4 the presents our estimation and simulation results and Section 5 conclude. Proofs of the claims made in the paper are contained in the appendix.

2 Model

Consider the following global VAR model as proposed by Pesaran et al. (2002). There are N countries and for each country the following vector autoregressive model is assumed to hold:

$$\mathbf{x}_{it} = \mathbf{a}_{i0} + \mathbf{a}_{i1}t + \mathbf{\Phi}_i \mathbf{x}_{i,t-1} + \mathbf{\Lambda}_{i0}\mathbf{x}^*_{it} + \mathbf{\Lambda}_{i1}\mathbf{x}^*_{i,t-1} + \boldsymbol{\varepsilon}_{it}, \tag{2.1}$$

where $\mathbf{x}_{it}$ is a $k \times 1$ vector of endogenous variables in a country i, at time t, $\mathbf{a}_{i0}$ and $\mathbf{a}_{i1}$ are $k \times 1$ vector of parameters, $\mathbf{\Phi}_i$, $\mathbf{\Lambda}_{i0}$, and $\mathbf{\Lambda}_{i1}$ are $k \times k$ matrices of parameters, $\boldsymbol{\varepsilon}_{ij}$ is a $k \times 1$ vector of innovations, and

$$\mathbf{x}^*_{it} = \sum_{j=1}^{N} \mathbf{W}_{ij}\mathbf{x}_{jt}, \tag{2.2}$$

is so called foreign variable which is constructed as a (country specific) weighted average of endogenous variables in other countries where $\mathbf{W}_{ij}$ are $k \times k$ matrices of observable weights. Pesaran et al. propose to estimate the model on a country-by-country basis, arguing that as $N \to \infty$, under reasonable (in their view) assumptions, $Cov\left[\left(\sum_{j=1}^{N} \mathbf{W}_{ij}\mathbf{x}_{it}\right), \boldsymbol{\varepsilon}_{it}\right] \to 0$ and this is what is then referred to as weak exogeneity.

In this paper, I examine the amount of endogeneity bias that can be expected in small samples. I also argue that the assumptions that guarantee that the weak endogeneity holds are too restrictive, in which case there is also an asymptotic bias. I suggest a simple alternative instrumental variable procedure that is consistent under full endogeneity.

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

$$\mathbf{x}_t = \mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{\Lambda}_0\mathbf{W}\mathbf{x}_t + \mathbf{\Lambda}_1\mathbf{W}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t, \tag{2.3}$$

where $(m = 0, 1)$:

$$\begin{aligned} \mathbf{x}_t &= \left(\mathbf{x}_{1t}', ..., \mathbf{x}_{Nt}'\right)', \\ \mathbf{a}_m &= \left(\mathbf{a}_{1m}', ..., \mathbf{a}_{Nm}'\right)', \\ \mathbf{\Phi} &= diag\left(\mathbf{\Phi}_1, ..., \mathbf{\Phi}_N\right), \\ \mathbf{\Lambda}_m &= diag\left(\mathbf{\Lambda}_{1m}, ..., \mathbf{\Lambda}_{Nm}\right), \\ \mathbf{W} &= \left(\mathbf{W}_{ij}\right)_{j=1,..N}^{i=1,..,N}, \\ \boldsymbol{\varepsilon}_t &= \left(\boldsymbol{\varepsilon}_{1t}', ..., \boldsymbol{\varepsilon}_{Nt}'\right)'. \end{aligned} \tag{2.4}$$

The solution of the stacked model is obtained (I will show later that this expression is well defined, based on an explicit set of assumptions) as

$$\mathbf{x}_t = \left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1}\left(\mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{\Phi}\mathbf{x}_{t-1} + \mathbf{\Lambda}_1\mathbf{W}\mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t\right). \tag{2.5}$$

Provided that the the innovations $\boldsymbol{\varepsilon}_t$ are independent in the time dimension, the endogeneity of the regressors $\mathbf{W}\mathbf{x}_t$ follows from

$$E\left(\mathbf{W}\mathbf{x}_t\boldsymbol{\varepsilon}_t\right) = \left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1} E\left(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t'\right). \tag{2.6}$$

Pesaran et al. assume that the weight matrices $\mathbf{W}_{ij}$ are diagonal with $\mathbf{W}_{ij} = diag\left(w_{ij}^1, .., w_{ij}^k\right)$ and that

$$\sum_{j=0}^{N}\left(w_{ij}^m\right)^2 \to 0, \text{ as } N \to \infty, \text{ for all } i \text{ and } m. \tag{2.7}$$

This implies that asymptotically the foreign variables have no explanatory power in the model! Asymptotic properties of such model should not be used as a small sample guidance for our estimators if we actually expect some degree of cross-sectional dependence in our model. A more reasonable assumption is to require some limit on the amount of the cross-sectional interdependence in the model but leave some room for cross-sectional dependence to survive even in the limit. A typical assumption in the spatial econometrics literature is to require that

$$\sum_{j=0}^{N}\left|w_{ij}^m\right| \leq c < \infty, \text{ for all } i \text{ and } m,$$

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 to derive asymptotic properties of our model.

It also has to be noted that at least some practical applications use data in which the number of time series is larger than the number of cross-sections. Furthermore, the general statement of the GVAR model allows for the slope coefficients to vary across the cross-sections. Both of these observations suggest that it would be of interest to derive the asymptotic distribution of the estimators holding N constant. In this case the asymptotic (with respect to N) weak endogeneity argument no longer applies.

2.1 Assumptions

Here I spell out explicitly the general assumptions that are maintained throughout the paper.

Assumption 1 *The disturbances ε_{it} are generated from*

$$\boldsymbol{\varepsilon}_t = \mathbf{R}_{t,N} \boldsymbol{\eta}_t, \tag{2.8}$$

where $\boldsymbol{\eta}_t = (\boldsymbol{\eta}_{1t}, ..., \boldsymbol{\eta}_{Nt})$ where $\boldsymbol{\eta}_{Nt} = (\eta_{1it}, ..., \eta_{kit})'$ is a $k \times 1$ vector of innovations and:

(a) *The innovations η_{mit} are totally independent (with respect to 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 $\mathbf{R}_{t,N}$ has uniformly bounded absolute row sums, i.e. denoting $r_{ij,t,N}$ the ij-th element of $\mathbf{R}_{t,N}$ it holds that*

$$\sum_{j=1}^{Nk} |r_{ij,t,N}| \leq k_r < \infty, \tag{2.9}$$

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

Assumption 1 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. The 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. The following assumption then guarantees that the degree of international interactions in the data does not explode as the sample size (number of countries) increases:

Assumption 2 **(a)** *The sequence of the weight matrices* $\mathbf{W}$ *(I drop the indexation by the sample size for convenience) has uniformly bounded absolute row and column sums, i.e.*

$$\sum_{j=1}^{N} \left| w_{ij}^{m} \right| \leq k_w < \infty, \tag{2.10}$$

where the constant k_w *does not depend on* T *or* N *(but can potentially depend on other parameters of the model).*

(b) *Furthermore, the sequences of matrices* $(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1}$ *and* $\left[\mathbf{I}_{kN} - (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} (\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W})\right]^{-1}$ *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* $\mathbf{\Phi}$*,* $\mathbf{\Lambda}_0$*, and* $\mathbf{\Lambda}_1$ *have uniformly bounded absolute row sums and the vectors* $\mathbf{a}_0$ *and* $\mathbf{a}_1$ *have elements uniformly bounded in absolute value.*

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. However the absolute summability is still an additional condition. It proves to be useful to define the following notation. Let $\mathbf{A}$ be any square $n \times n$ matrix with real entries. I denote its spectral radius as

$$\rho(\mathbf{A}) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } \mathbf{A}\}. \tag{2.11}$$

Assumption 3 *The spectral radius of* $(\mathbf{\Lambda}_0 \mathbf{W})$ *is uniformly less than one, i.e.* $\rho(\mathbf{\Lambda}_0 \mathbf{W}) \leq k < \infty$*, where the constant* k *does not depend on* N *or* T*.*

Assumption 4 *The spectral radius of* $(\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W})$ *and of* $(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} (\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W})$ *are uniformly less than one.*

Finally to be able to demonstrate that the observable process is a well-defined transformation of the underlying innovations, we need an assumption about the initial starting values of the process:

Assumption 5 *The initial observations* $\mathbf{x}_0$ *are drawn from*

$$\mathbf{x}_0 = \mathbf{R}_{0,N}\boldsymbol{\xi}, \tag{2.12}$$

where

(a) *The innovations collected in the* $Nk \times 1$ *vector* $\boldsymbol{\xi}$ *are totally independent of each other as well as of innovations* $\boldsymbol{\eta}_t$ *for* $t > 0$ *and the elements of* $\boldsymbol{\xi}$ *have uniformly bounded absolute* $4 + \delta$ *moments for some* $\delta > 0$.

(b) *The sequence of* $Nk \times Nk$ *matrices* $\mathbf{R}_{0,N}$ *has uniformly bounded absolute row sums, i.e.*

$$\sum_{j=1}^{Nk} |r_{ij,0,N}| \leq k_0 < \infty, \tag{2.13}$$

where the constant k_0 *does not depend on* N *and* T.

Of course the above assumption would be satisfied if the data generating process is stable and the initial observations were drawn from the stationary distribution of the process, see e.g. Proposition 1 below.

[**To be added: discussion of random and/or fixed effects**]

2.2 Stationarity Conditions

Inspecting the solution to the global model given in (2.5), it follows that to determine whether the model is stationary, it is not sufficient to examine the stationarity of the country-by-country models separately, ignoring the endogeneity of $\mathbf{x}^*_{it}$, i.e. to examine the eigenvalues of $\boldsymbol{\Phi}_i$ (and $\boldsymbol{\Lambda}_1$). Instead, the stationarity of the global model is determined by the spectral radius of

$$(\mathbf{I}_{kN} - \boldsymbol{\Lambda}_0\mathbf{W})^{-1} (\boldsymbol{\Phi} + \boldsymbol{\Lambda}_1\mathbf{W}). \tag{2.14}$$

Hence it does not suffice to impose stability of each country model (i.e. require that $\rho(\boldsymbol{\Phi}) < 1$). Accounting for the autocorrelation in the foreign variable (i.e. imposing that $\rho(\boldsymbol{\Phi} + \boldsymbol{\Lambda}_1\mathbf{W}) < 1$) is also not sufficient. Instead, the stability of the process also depends on the strength of the contemporaneous global links in the model (i.e. on the parameters collected in $\boldsymbol{\Lambda}_0$) and it must be determined by the spectral radius of the entire matrix $(\mathbf{I}_{kN} - \boldsymbol{\Lambda}_0\mathbf{W})^{-1} (\boldsymbol{\Phi} + \boldsymbol{\Lambda}_1)$. In general when both N and T are allowed to tend to infinity, the claim that this is sufficient is not straightforward and is demonstrated in the proof of the following proposition:

Proposition 1 *Under Assumptions 1-5,* $\mathbf{x}_t$ *has well defined uniformly bounded absolute* $4+\delta$ *moments for some* $\delta > 0$*. Furthermore, if* $\mathbf{a}_1 = \mathbf{0}$*, then in the limit as* $T \to \infty$*,* $\mathbf{x}_T$ *converges in quadratic means to a random variable* $\mathbf{x}_\infty$ *which has well defined finite absolute* $4+\delta$ *moments for some* $\delta > 0$ *with*

$$E(\mathbf{x}_\infty) = \left[\mathbf{I}_{kN} - (\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W})^{-1}(\mathbf{\Phi} + \mathbf{\Lambda}_1)\right]^{-1}(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W})^{-1}\mathbf{a}_0.$$

If additionally $\lim_{T\to\infty} E(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t') = \mathbf{\Omega}_\varepsilon$*, we have*

$$\begin{aligned} vech\left[VC(\mathbf{x}_\infty)\right] &= \left\{\mathbf{I}_{N^2k^2} - \left[\mathbf{A}(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W})^{-1} \otimes \mathbf{A}(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W})^{-1}\right]\right\}^{-1} \\ &\quad \cdot \mathbf{D} \cdot vech(\mathbf{\Omega}_\varepsilon), \end{aligned}$$

where $\mathbf{D}$ *is a duplication matrix such that* $vec(\mathbf{\Omega}_\varepsilon) = \mathbf{D} \cdot vech(\mathbf{\Omega}_\varepsilon)$*.*

Proof: See the Appendix.

The asymptotic results in the above proposition can be useful in specifying the initial distribution of the initial values of the process $\mathbf{x}_0$. Of course in the presence of deterministic time trends ($\mathbf{a}_1 \neq \mathbf{0}$), the limiting moments of $\mathbf{x}_T$ only exist when appropriately normalizing by $T^{-\frac{3}{2}}$, see the discussion in Hamilton (1994), Chapter 16.

I now examine the sufficient conditions for stability in more detail. Note that for any matrix norm, the spectral radius $\rho(\mathbf{A})$ is smaller than the norm $\|\mathbf{A}\|$ (e.g. Theorem 5.6.9. in Horn and Johnson, 1985). Hence using the submultiplicative property of the matrix norm, we have that

$$\begin{aligned} \rho\left[(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W})^{-1}\mathbf{\Phi}\right] &\leq \left\|(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W})^{-1}(\mathbf{\Phi} + \mathbf{\Lambda}_1\mathbf{W})\right\| \\ &\leq \left\|(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W})^{-1}\right\| \cdot \|\mathbf{\Phi} + \mathbf{\Lambda}_1\mathbf{W}\|. \end{aligned} \tag{2.15}$$

Convenient matrix norms can be, for example, the maximum absolute row sum of a matrix defined as

$$\|\mathbf{A}\|_1 = \max_{1\leq i\leq n} \sum_{j=1}^{n} |a_{ij}|, \tag{2.16}$$

or the spectral norm

$$\|\mathbf{A}\|_2 = \max_{1\leq i\leq n} \left\{\sqrt{\lambda} : \lambda \text{ is an eigenvalue of } \mathbf{A}'\mathbf{A}\right\}, \tag{2.17}$$

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 that the inverse $(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1}$ can be expanded as an infinite sum. Therefore, (any) norm of $(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1}$ can be bounded from above by

$$\left\| (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \right\| \leq \sum_{s=0}^{\infty} \left(\|\mathbf{W}\| \cdot \|\mathbf{\Lambda}_0\| \right)^s . \tag{2.18}$$

Often the weight matrices are row normalized. In this case we have that $\|\mathbf{W}\|_1 = 1$ and hence

$$\begin{aligned} \left\| (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \right\|_1 &\leq \sum_{s=0}^{\infty} \|\mathbf{\Lambda}_0\|_1^s \\ &= \frac{1}{1 - \|\mathbf{\Lambda}_0\|_1} \\ &= \frac{1}{1 - \max_{1 \leq i \leq N} \left\{ \|\mathbf{\Lambda}_{i0}\|_1 \right\}}. \end{aligned} \tag{2.19}$$

Note to satisfy Assumption 3 (in the case of $\|\mathbf{W}\|_1 = 1$) we can, for example, require that $0 \leq \max_{1 \leq i \leq N} \left\{ \|\mathbf{\Lambda}_{i0}\|_1 \right\} < 1$. However, if there are global feedbacks in the model, we have $\max_{1 \leq i \leq N} \left\{ \|\mathbf{\Lambda}_{i0}\|_1 \right\} > 0$ and hence

$$\frac{1}{1 - \max_{1 \leq i \leq N} \left\{ \|\mathbf{\Lambda}_{i0}\|_1 \right\}} > 1. \tag{2.20}$$

In this case the requirement that $\|\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W}\|_1 < 1$ (which is a stronger requirement than $\rho(\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W}) < 1$) does not necessarily guarantees that the process is stable.[3]

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

Proposition 2 *Assume that the maximum absolute row sums of* $\mathbf{W}$ *are less or equal to* k_w*, i.e.* $\|W\|_1 \leq k_w$*. Suppose that*

$$\|\mathbf{\Phi}\|_1 + k_w \left(\|\mathbf{\Lambda}_0\|_1 + \|\mathbf{\Lambda}_1\|_1 \right) < 1. \tag{2.21}$$

Then the spectral radius of $(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} (\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W})$ *is less than one.*

[3] This is motivated by the fact that the requirement $\|\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W}\|_1 < 1$ is a sufficient condition for $\rho(\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W}) < 1$.

Proof: see Appendix.

The above proposition provides a simpler alternative to checking the eigenvalues of the entire matrix $(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W})^{-1}(\mathbf{\Phi} + \mathbf{\Lambda}_1)$. 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 $|\mathbf{\Phi}| + |\mathbf{\Lambda}_{i0}| + |\mathbf{\Lambda}_{i1}|$ are less than one. Note however that the above proposition provides only a sufficient condition for stability. Necessary condition is that the spectral radius of $(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W})^{-1}(\mathbf{\Phi} + \mathbf{\Lambda}_1)$ is less than one.

3 Estimation Procedure and Large Sample Results

The stacked model can be written compactly as

$$\mathbf{x}_t = \mathbf{\Lambda}_0\mathbf{W}\mathbf{x}_t + \underset{Nk\times 4Nk}{\boldsymbol{\theta}} \cdot \underset{4Nk\times 1}{\mathbf{Z}_t} + \boldsymbol{\varepsilon}_t, \tag{3.1}$$

with

$$\boldsymbol{\theta} = \sum_{i=1}^{N}\left(\mathbf{E}_i^N \otimes \boldsymbol{\theta}_i\right), \qquad \boldsymbol{\theta}_i = \left[\begin{array}{cccc} \mathbf{a}_{i0} & \mathbf{a}_{i1} & \mathbf{\Phi}_i & \mathbf{\Lambda}_{i1} \end{array}\right], \tag{3.2}$$

and

$$\mathbf{Z}_t = \sum_{i=1}^{N}\left(\mathbf{e}_i^N \otimes \mathbf{Z}_{it}\right), \qquad \mathbf{Z}_{it} = \left[\boldsymbol{\iota}_k', \boldsymbol{\iota}_k' t, \mathbf{x}_{i,t-1}', \mathbf{x}_{i,t-1}^{*\prime}\right]', \tag{3.3}$$

where I denote by where $\mathbf{E}_{ij}^N$ is an $N \times N$ matrix of zeros with an entry of one at the ij-th position, by $\mathbf{e}_i^N$ a $N \times 1$ vector of zeros with an entry of one at the i-th position and by $\boldsymbol{\iota}_k$ a $k \times 1$ vector of ones. Note that using this notation, the model for each country can be written as

$$\mathbf{x}_{it} = \mathbf{\Lambda}_{i0}\mathbf{x}_{it}^* + \underset{k\times 4k}{\boldsymbol{\theta}_i} \cdot \underset{4k\times 1}{\mathbf{Z}_{it}} + \boldsymbol{\varepsilon}_{it}. \tag{3.4}$$

Given Assumption 3, the inverse of $(\mathbf{I}_{Nk} - \mathbf{\Lambda}_0\mathbf{W})$ exists (cf. Lemma 5.6.10 and Corollary 5.6.16 in Horn and Johnson, 1985) and the solution to the global model is then

$$\mathbf{x}_t = (\mathbf{I}_{Nk} - \mathbf{\Lambda}_0\mathbf{W})^{-1}(\boldsymbol{\theta}\mathbf{Z}_t + \boldsymbol{\varepsilon}_t). \tag{3.5}$$

Based on the discussion in Amemyia (1986), ideal instruments for $\mathbf{x}_t^* = \mathbf{W}\mathbf{x}_t$ would then be $\mathbf{W}\left(\mathbf{I}_{Nk} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1}\boldsymbol{\theta}\mathbf{Z}_t$. Observe that we can expand the inverse $\left(\mathbf{I}_{Nk} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1}$ by its infinite sum approximation (see e.g. Corollary 5.6.16 in Horn and Johnson, 1985):

$$\left(\mathbf{I}_{Nk} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1} = \sum_{s=0}^{\infty}\left(\mathbf{\Lambda}_0\mathbf{W}\right)^s. \tag{3.6}$$

When $\mathbf{\Lambda}_0$ and $\boldsymbol{\theta}$ are scalars, the optimal instruments for $\mathbf{W}\mathbf{x}_t$would be $\mathbf{W}\mathbf{Z}_t$, $\mathbf{W}^2\mathbf{Z}_t$, ... However, in the general case of a VAR model, the instrument set is more complicated.

Note that we can write

$$\mathbf{\Lambda}_0 = \sum_{i=1}^{N}\left(\mathbf{E}_{ii}^N \otimes \mathbf{\Lambda}_{i0}\right), \qquad \mathbf{W} = \sum_{l=1}^{k}\sum_{m=1}^{k}\left(\mathcal{W}_{lm} \otimes \mathbf{E}_{lm}^k\right), \tag{3.7}$$

where the $N \times N$ matrices $\mathcal{W}_{lm}$ are weights that relate the m-th foreign variable in the l-th equation of the domestic system.

The solution to the model implies then that the foreign variable is

$$\begin{aligned}
\mathbf{x}_t^* &= \sum_{p=1}^{k}\sum_{q=1}^{k}\left(\mathcal{W}_{pq} \otimes \mathbf{E}_{pq}^k\right)\sum_{s=0}^{\infty}\left[\sum_{i=1}^{N}\left(\mathbf{E}_{ii}^N \otimes \mathbf{\Lambda}_{i0}\right)\sum_{l=1}^{k}\sum_{m=1}^{k}\left(\mathcal{W}_{lm} \otimes \mathbf{E}_{lm}^k\right)\right]^s & (3.8)\\
&\quad \cdot\left[\sum_{i=1}^{N}\left(\mathbf{E}_{ii}^N \otimes \boldsymbol{\theta}_i\right)\right]\mathbf{Z}_t + \left[\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right]^{-1}\boldsymbol{\varepsilon}_t & (3.9)\\
&= \sum_{p=1}^{k}\sum_{q=1}^{k}\sum_{s=0}^{\infty}\sum_{i=1}^{N}\sum_{n_{11}=1}^{k}\sum_{n_{12}=1}^{k}\sum_{n_{13}=1}^{N}\sum_{n_{14}=1}^{N}\cdots\sum_{n_{s1}=1}^{k}\sum_{n_{s2}=1}^{k}\sum_{n_{s3}=1}^{N}\sum_{n_{s4}=1}^{N}\\
&\quad \left[\mathcal{W}_{pq}\left(\mathbf{E}_{n_{13}n_{14}}^N\mathcal{W}_{n_{11}n_{12}} \cdot \ldots \cdot \mathbf{E}_{n_{s3}n_{s4}}^N\mathcal{W}_{n_{s1}n_{s2}}\right)\mathbf{E}_{ii}^N\right.\\
&\quad \left.\otimes\mathbf{E}_{pq}^k\left(\mathbf{\Lambda}_{i0}\mathbf{E}_{n_{11}n_{12}} \cdot \ldots \cdot \mathbf{\Lambda}_{i0}\mathbf{E}_{n_{s1}n_{s2}}\right)\boldsymbol{\theta}_i\right]\mathbf{Z}_t\\
&\quad + \left[\mathbf{I}_{kN} - \left(\mathbf{I}_N \otimes \mathbf{\Lambda}_0\right)\mathbf{W}\right]^{-1}\boldsymbol{\varepsilon}_t\\
&= \sum_{p=1}^{k}\sum_{q=1}^{k}\sum_{s=0}^{\infty}\sum_{i=1}^{N}\sum_{n_{11}=1}^{k}\sum_{n_{12}=1}^{k}\sum_{n_{13}=1}^{N}\sum_{n_{14}=1}^{N}\cdots\sum_{n_{s1}=1}^{k}\sum_{n_{s2}=1}^{k}\sum_{n_{s3}=1}^{N}\sum_{n_{s4}=1}^{N}\\
&\quad \left[\mathbf{I}_N \otimes \mathbf{E}_{pq}^k\left(\mathbf{\Lambda}_{i0}\mathbf{E}_{n_{11}n_{12}}^k \cdot \ldots \cdot \mathbf{\Lambda}_{i0}\mathbf{E}_{n_{s1}n_{s2}}^k\boldsymbol{\theta}_i\right)\right]\\
&\quad \cdot\left[\mathcal{W}_{pq}\left(\mathbf{E}_{n_{13}n_{14}}^N\mathcal{W}_{n_{11}n_{12}} \cdot \ldots \cdot \mathbf{E}_{n_{s3}n_{s4}}^N\mathcal{W}_{n_{s1}n_{s2}}\mathbf{E}_{ii}^N\right) \otimes \mathbf{I}_k\right]\mathbf{Z}_t\\
&\quad + \left[\mathbf{I}_{kN} - \left(\mathbf{I}_N \otimes \mathbf{\Lambda}_0\right)\mathbf{W}\right]^{-1}\boldsymbol{\varepsilon}_t.
\end{aligned}$$

To facilitate manageable notation, we associate a single number, say m to a given values of the indexes $p, q, s, i, n_{11}, ..., n_{s4}$ and denote a matrix of unknown parameters

$$\mathbf{\Upsilon}_m = \mathbf{E}_{pq}^k \mathbf{\Lambda}_{i0} \mathbf{E}_{n_{11}n_{12}}^k \cdot ... \cdot \mathbf{\Lambda}_{i0} \mathbf{E}_{n_{s1}n_{s2}}^k \boldsymbol{\theta}_i, \tag{3.10}$$

and an observed matrix of transformed data

$$\mathbf{H}_m = \left[\left(\mathbf{E}_{n_{13}n_{14}}^N \mathcal{W}_{n_{11}n_{12}} \cdot ... \cdot \mathbf{E}_{n_{s3}n_{s4}}^N \mathcal{W}_{n_{s1}n_{s2}} \mathbf{E}_{ii}^N \right) \otimes \mathbf{I}_k \right]. \tag{3.11}$$

Using this simplified notation, the foreign variable becomes

$$\begin{aligned} \mathbf{x}_t^* &= \sum_m \left(\mathbf{I}_N \otimes \mathbf{\Upsilon}_m \right) \mathbf{H}_{tm} \mathbf{Z}_t + \left[\mathbf{I}_{kN} - \left(\mathbf{I}_N \otimes \mathbf{\Lambda}_0 \right) \mathbf{W} \right]^{-1} \boldsymbol{\varepsilon}_t \qquad (3.12) \\ &= \sum_m \left(\mathbf{Z}_t' \mathbf{H}_{tm}' \otimes \mathbf{I}_{kN} \right) \left(\mathbf{I}_N \otimes \mathbf{K}_{kN} \otimes \mathbf{I}_k \right) \left(vec \mathbf{I}_N \otimes \mathbf{I}_{k^2} \right) vec \mathbf{\Upsilon}_m \\ &\quad + \left[\mathbf{I}_{kN} - \left(\mathbf{I}_N \otimes \mathbf{\Lambda}_0 \right) \mathbf{W} \right]^{-1} \boldsymbol{\varepsilon}_t. \end{aligned}$$

Thus a valid set of instrument for $\mathbf{x}_t^*$ can consist of independent columns of $\mathbf{H}_t = \left[\widetilde{\mathbf{H}}_{t,m_1}, .., \widetilde{\mathbf{H}}_{t,m_n} \right]$ where

$$\widetilde{\mathbf{H}}_{tm} = \left(\mathbf{Z}_t' \mathbf{H}_m' \otimes \mathbf{I}_{kN} \right) \left(\mathbf{I}_N \otimes \mathbf{K}_{kN} \otimes \mathbf{I}_k \right) \left(vec \mathbf{I}_N \otimes \mathbf{I}_{k^2} \right), \tag{3.13}$$

and where $m_1, ..., m_n$ are some selected indexes corresponding to a set of values of the indexes $p, q, s, i, n_{11}, ..., n_{s4}$ in the expression (3.9). Based on the arguments in Kelejian and Prucha (1998), at least the quadratic approximation should be used and hence at the minimum the instruments should contain terms for which s is at least 2. Denote the set of stacked instruments by $\mathbf{H} = (\mathbf{H}_1', ..., \mathbf{H}_T')'$. In the first step of the procedure, the projected values of $\mathbf{x}^* = (\mathbf{x}_1^{*\prime}, ..., \mathbf{x}_T^{*\prime})'$are calculated as

$$\begin{aligned} \widehat{\mathbf{x}^*} &= \mathbf{P}_H \mathbf{x}^*, \qquad (3.14) \\ \mathbf{P}_H &= \mathbf{H}' \left(\mathbf{H} \mathbf{H}' \right)^{-1} \mathbf{H}. \end{aligned}$$

In the second step, we regress $\mathbf{x}_t$ on the predicted values of the endogenous variables and on the exogenous variables. This amounts to estimating country-by-country regressions using the predicted instead of the true values of the foreign variable). Note that the model for country i can be written as

$$\mathbf{x}_{it} = \mathbf{\Lambda}_{i0} \mathbf{x}_{it}^* + \boldsymbol{\theta}_i \mathbf{Z}_{it} + \boldsymbol{\varepsilon}_{it}, \tag{3.15}$$

and hence the 2SLS estimator is

$$\left(\widetilde{\boldsymbol{\Lambda}}_{i0}, \widetilde{\boldsymbol{\theta}}_i\right) = \left[\sum_{t=1}^{T} \mathbf{x}_{it} \left(\widehat{\mathbf{x}}_{it}^{*\prime}, \mathbf{Z}_{it}^{\prime}\right)\right] \left[\sum_{t=1}^{T} \begin{pmatrix} \mathbf{x}_{it}^{*} \\ \mathbf{Z}_{it} \end{pmatrix} \left(\widehat{\mathbf{x}}_{it}^{*\prime}, \mathbf{Z}_{it}^{\prime}\right)\right]^{-1}. \tag{3.16}$$

To be able to state conveniently large sample results, I now restrict attention to a model without deterministic time trend, i.e. to the case $\mathbf{a}_1 = 0$. In this case, the matrix of weakly exogenous regressors at time t for country i becomes

$$\mathbf{Z}_{it} = \left[\boldsymbol{\iota}_k^{\prime}, \mathbf{x}_{i,t-1}^{\prime}, \mathbf{x}_{i,t-1}^{\prime *}\right]^{\prime}.$$

It proves to be convenient to work with the model stacked over the time periods. Note that the model without deterministic trends can be rewritten as

$$\mathbf{x}_t = \sum_{i=1}^{N} \left(\mathbf{E}_{ii}^{N} \otimes \boldsymbol{\Lambda}_{i0}\right) \mathbf{x}_t^{*} + \sum_{i=1}^{N} \left(\mathbf{E}_{ii}^{N} \otimes \boldsymbol{\Lambda}_{i1}\right) \mathbf{x}_{t-1}^{*} + \sum_{i=1}^{N} \left(\mathbf{E}_{ii}^{N} \otimes \boldsymbol{\Phi}_i\right) \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t. \tag{3.17}$$

After vectorizing the right-hand side, we obtain

$$\begin{aligned} \mathbf{x}_t \quad = \quad & \sum_{i=1}^{N} \left(\mathbf{x}_t^{*\prime} \otimes \mathbf{I}_{mN}\right) \left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right) \left(vec\mathbf{E}_{ii}^{N} \otimes \mathbf{I}_{m^2}\right) vec\boldsymbol{\Lambda}_{i0} \qquad (3.18) \\ & + \sum_{i=1}^{N} \left(\mathbf{x}_{t-1}^{*\prime} \otimes \mathbf{I}_{mN}\right) \left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right) \left(vec\mathbf{E}_{ii}^{N} \otimes \mathbf{I}_{m^2}\right) vec\boldsymbol{\Lambda}_{i1} \\ & + \sum_{i=1}^{N} \left(\mathbf{x}_{t-1}^{\prime} \otimes \mathbf{I}_{mN}\right) \left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right) \left(vec\mathbf{E}_{ii}^{N} \otimes \mathbf{I}_{m^2}\right) vec\boldsymbol{\Phi}_i + \boldsymbol{\varepsilon}_t, \end{aligned}$$

where K_{mN} is a $mN \times mN$ commutation matrix (see e.g. Magnus and Neudecker, 1988, chapter 3.7).

Stacking over time periods leads to

$$\mathbf{x} = \mathbf{Y}\boldsymbol{\theta} + \boldsymbol{\varepsilon}, \tag{3.19}$$

where $\mathbf{x} = (\mathbf{x}_1', ..., \mathbf{x}_T')'$, $\mathbf{Y} = (\mathbf{Y}_1', ..., \mathbf{Y}_T')'$ and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', ..., \boldsymbol{\varepsilon}_T')'$ with

$$\mathbf{Y}_t = \begin{bmatrix} \left(\mathbf{x}_t^{*\prime} \otimes \mathbf{I}_{mN}\right)\left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right)\left(vec\mathbf{E}_{11}^N \otimes \mathbf{I}_{m^2}\right) : \\ \vdots \\ \left(\mathbf{x}_t^{*\prime} \otimes \mathbf{I}_{mN}\right)\left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right)\left(vec\mathbf{E}_{NN}^N \otimes \mathbf{I}_{m^2}\right) : \\ \left(\mathbf{x}_{t-1}^{*\prime} \otimes \mathbf{I}_{mN}\right)\left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right)\left(vec\mathbf{E}_{11}^N \otimes \mathbf{I}_{m^2}\right) : \\ \vdots \\ \left(\mathbf{x}_{t-1}^{*\prime} \otimes \mathbf{I}_{mN}\right)\left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right)\left(vec\mathbf{E}_{NN}^N \otimes \mathbf{I}_{m^2}\right) : \\ \left(\mathbf{x}_{t-1}' \otimes \mathbf{I}_{mN}\right)\left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right)\left(vec\mathbf{E}_{11}^N \otimes \mathbf{I}_{m^2}\right) : \\ \vdots \\ \left(\mathbf{x}_{t-1}' \otimes \mathbf{I}_{mN}\right)\left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right)\left(vec\mathbf{E}_{NN}^N \otimes \mathbf{I}_{m^2}\right) \end{bmatrix}, \tag{3.20}$$

where : denotes horizontal stacking, and

$$\boldsymbol{\theta} = \begin{bmatrix} (vec\boldsymbol{\Lambda}_{10})' : & : (vec\boldsymbol{\Lambda}_{N0})' : \\ (vec\boldsymbol{\Lambda}_{11})' : & : (vec\boldsymbol{\Lambda}_{N1})' : \\ (vec\boldsymbol{\Phi}_1)' : & : (vec\boldsymbol{\Phi}_N)' \end{bmatrix}'. \tag{3.21}$$

Note that the 2SLS estimator can be equivalently written as

$$\widehat{\boldsymbol{\theta}}_{2SLS} = \left(\widehat{\mathbf{Y}}'\mathbf{Y}\right)^{-1}\widehat{\mathbf{Y}}'\mathbf{x}, \tag{3.22}$$

where $\widehat{\mathbf{Y}}$ is the same as $\mathbf{Y}$ except that $\mathbf{x}_t^*$ in the definition of $\mathbf{Y}_t$ is replaced by $\widehat{\mathbf{x}}_t^*$. Observe that $\widehat{\mathbf{Y}}$ is hence

$$\widehat{\mathbf{Y}} = \left[\begin{pmatrix} \widehat{\mathbf{x}_1^*}' \\ \\ \widehat{\mathbf{x}_T^*}' \end{pmatrix}, \begin{pmatrix} \mathbf{x}_0' \\ \\ \mathbf{x}_{T-1}' \end{pmatrix}, \begin{pmatrix} \mathbf{x}_0^{*\prime} \\ \\ \mathbf{x}_{T-1}^{*\prime} \end{pmatrix} \otimes \mathbf{I}_{mN}\right]\mathbf{E},$$

where I define the transformation matrices $\mathbf{E}$ as

$$\mathbf{E} = \left(\mathbf{E}_1, ..., \mathbf{E}_N\right), \tag{3.23}$$

where

$$\mathbf{E}_j = \left(\mathbf{I}_N \otimes \mathbf{K}_{mN} \otimes \mathbf{I}_m\right)\left(vec\mathbf{E}_{jj}^N \otimes \mathbf{I}_{m^2}\right). \tag{3.24}$$

The asymptotic distribution of the estimator depends on the choice of instruments. To fix ideas, I assume that the instruments are chosen so that asymptotically they perfectly aproximate the expectations of the dependent variable:[4]

Assumption 6 *The instruments collected in* $\mathbf{H}$ *are such that*

$$p\lim_{T\to\infty}(NT)^{-1}\widehat{\mathbf{Y}}'\mathbf{Y} = \lim_{T\to\infty}(NT)^{-1}E(\mathbf{Y})'E(\mathbf{Y}) = \mathbf{\Xi}, \tag{3.25}$$

where Ξ *is invertable,*

$$p\lim_{T\to\infty}\widehat{\mathbf{Y}}, \tag{3.26}$$

and

$$\sqrt{NT}\widehat{\mathbf{Y}}'\boldsymbol{\varepsilon} - \sqrt{NT}E(\mathbf{Y})'\boldsymbol{\varepsilon} = o_p(1). \tag{3.27}$$

The theorem below summarizes the main asymptotic results:

Theorem 1 *Under Assumptions 1-6 and if the limit*

$$\mathbf{\Sigma}_{Y\varepsilon} = \lim_{T\to\infty}E(\mathbf{Y})\mathbf{R}E(\boldsymbol{\eta}\boldsymbol{\eta}')\mathbf{R}'E(\mathbf{Y}), \tag{3.28}$$

exists we have that

$$\sqrt{NT}\left(\widehat{\boldsymbol{\theta}}_{2SLS} - \boldsymbol{\theta}\right) \xrightarrow{D} N(0, \mathbf{\Psi}) \;\; \text{as } T \to \infty, \tag{3.29}$$

where

$$\mathbf{\Psi} = \mathbf{\Xi}\mathbf{\Sigma}_{Y\varepsilon}\mathbf{\Xi}'. \tag{3.30}$$

Proof: See the Appendix.

4 Empirical Illustration

[To be added: application of the procedure to Pesaran et. al data]
[To be added: Monte Carlo simulations results]

[4]See e.g. the series type efficient IV estimator introduced in Kelejian, Prucha and Yuzefovich (2003).

A Appendix

The following lemma is useful in evaluating infinite sums of sequences of matrices:

Lemma A1 *Let* $\mathbf{A}$, $\mathbf{B}$ *and* $\mathbf{C}$ *be square matrices with same dimensions and let* $\|\mathbf{A}\|$ *and* $\|\mathbf{B}\|$ *be less than one for some matrix norm. Then the matrix* $\mathbf{S} = \sum_{n=0}^{\infty} \mathbf{A}^n \mathbf{C} \mathbf{B}^n$ *is well defined and*

$$vec\,(\mathbf{S}) = \left[\mathbf{I} - (\mathbf{B}' \otimes \mathbf{A})\right]^{-1} vec\,(\mathbf{C})\,. \tag{A.1}$$

Furthermore, the finite sum $\mathbf{S}_t = \sum_{n=0}^{t} \mathbf{A}^n \mathbf{C} \mathbf{B}^n$ *can be expressed as*

$$\mathbf{S}_t = \mathbf{S} - \mathbf{A}^{t+1} \mathbf{S} \mathbf{B}^{t+1}. \tag{A.2}$$

Proof: We have that

$$\|\mathbf{S}_{t+1}\| - \|\mathbf{S}_t\| = \|\mathbf{A}^n \mathbf{C} \mathbf{B}^n\| \leq \|\mathbf{A}\|^n \|\mathbf{C}\| \|\mathbf{B}\|^n \to 0, \tag{A.3}$$

and hence the series $\|\mathbf{S}_t\|$ is Cauchy and converges to, say $\|\mathbf{S}\|$. By Theorem 5.6.15 in Horn and Johnson it must be that the entries in $\mathbf{S}_t$ converge to the entries in $\mathbf{S}$. To derive the expression for $\mathbf{S}$, note that

$$\begin{aligned} \mathbf{ASB} = \mathbf{A} & \left(\sum_{n=0}^{\infty} \mathbf{A}^n \mathbf{C} \mathbf{B}^n\right) \mathbf{B} = \left(\sum_{n=1}^{\infty} \mathbf{A}^n \mathbf{C} \mathbf{B}^n\right) \\ = & \left(\sum_{n=0}^{\infty} \mathbf{A}^n \mathbf{C} \mathbf{B}^n\right) - \mathbf{C} = \mathbf{S} - \mathbf{C}. \end{aligned} \tag{A.4}$$

After vectorizing and solving for $vec\,(\mathbf{S})$ we obtain the claim in the Lemma.

To derive the expression for the finite sum, we write

$$\begin{aligned} \mathbf{S}_t &= \mathbf{S} - \sum_{n=t+1}^{\infty} \mathbf{A}^n \mathbf{C} \mathbf{B}^n = \mathbf{S} - \mathbf{A}^{t+1} \left(\sum_{n=0}^{\infty} \mathbf{A}^n \mathbf{C} \mathbf{B}^n\right) \mathbf{B}^{t+1} \\ &= \mathbf{S} - \mathbf{A}^{t+1} \mathbf{S} \mathbf{B}^{t+1}. \end{aligned} \tag{A.5}$$

A.1 Proof of Proposition 1

Given Assumption 3, the matrix $(\mathbf{I} - \mathbf{\Lambda}_0 \mathbf{W})$ is invertible (cf. Lemma 5.6.10 and Corollary 5.6.16 in Horn and Johnson, 1985) and the endogenous variable $\mathbf{x}_t$ can be expressed as

$$\mathbf{x}_t = (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} (\mathbf{a}_0 + \mathbf{a}_1 t + \mathbf{\Phi} \mathbf{x}_{t-1} + \mathbf{\Lambda}_1 \mathbf{W} \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_t) . \tag{A.6}$$

By backward substitution, we then obtain

$$\mathbf{x}_t = \mathbf{b}_{1t} + \mathbf{b}_{2t} + \mathbf{b}_{3t} + \mathbf{b}_{4t}, \tag{A.7}$$

where

$$\begin{aligned}
\mathbf{b}_{1t} &= \sum_{s=0}^{t-1} \left[(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} (\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W}) \right]^s (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \mathbf{a}_0, \\
\mathbf{b}_{2t} &= \sum_{s=0}^{t-1} \left[(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} (\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W}) \right]^s (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \mathbf{a}_1 s, \\
\mathbf{b}_{3t} &= \sum_{s=0}^{t-1} \left[(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} (\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W}) \right]^s (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \boldsymbol{\varepsilon}_s, \\
\mathbf{b}_{4t} &= \left[(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} (\mathbf{\Phi} + \mathbf{\Lambda}_1 \mathbf{W}) \right]^t \mathbf{x}_0.
\end{aligned} \tag{A.8}$$

Given Assumption 2b, we then have $\mathbf{b}_{1t}$ and $\mathbf{b}_{2t}$ have elements uniformly bounded in absolute value. I demonstrate that the sequences of stochastic vectors $\mathbf{b}_{3t}$ and $\mathbf{b}_{4t}$ have elements with uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. The claim in the Proposition then follows from Minkowski's inequality.

Consider the stochastic term $\mathbf{b}_{3t}$:

$$\mathbf{b}_{3t} = \sum_{s=0}^{t-1} \mathbf{A}^s (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \boldsymbol{\varepsilon}_s. \tag{A.9}$$

Note that given Assumption 1, the random vector $\boldsymbol{\eta}_t$ and the sequence of matrices $\mathbf{R}_{t,N}$ satisfy the conditions of Lemma B2 in Mutl (2006). Therefore, the elements of the random vector $\boldsymbol{\varepsilon}_t$ have uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. From Assumption 2, we have that the absolute row sums of $\mathbf{A}^s (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1}$ are uniformly bounded in absolute value.

Hence by repeated application of the Lemma B2 in Mutl (2006), we have that $\mathbf{A}^s (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \boldsymbol{\varepsilon}_s$ has elements with uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$. By Minkowski inequality (XX) we then have that $\mathbf{b}_{3t}$ has elements with uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$.

Next, consider the stochastic term $\mathbf{b}_{4t} = \mathbf{A}^t \mathbf{x}_0$. Again, by Assumption 2, the matrix $\mathbf{A}^t$ has uniformly bounded absolute row sums and hence given Assumption 5, we have by the same Lemma B2 that the elements of $\mathbf{b}_{4t}$ have uniformly bounded absolute $4+\delta$ moments for some $\delta > 0$.

We now turn to the asymptotic moments of $\mathbf{x}_t$ as $t \to \infty$, assuming that $\mathbf{a}_1 = 0$. Using Lemma A1 and Theorem 5.6.12 in Horn and Johnson, it follows that $\mathbf{b}_{1t}$ converges to

$$\begin{aligned} \mathbf{b}_1 &= \lim_{t\to\infty} \mathbf{b}_{1t} = \lim_{t\to\infty} (\mathbf{I}_{kN} - \mathbf{A})^{-1} \left(\mathbf{I}_{kN} - \mathbf{A}^t\right) (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \mathbf{a}_0 \\ &= (\mathbf{I}_{kN} - \mathbf{A})^{-1} (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \mathbf{a}_0. \end{aligned} \tag{A.10}$$

Given Assumption 2b, ot follows that $\mathbf{b}_1$ has elements uniformly bounded in absolute value and it suffices to show that the elements of $\mathbf{b}_{3t}$ and $\mathbf{b}_{4t}$ converge in quadratic means to random variables $\mathbf{b}_3$ and $\mathbf{b}_4$ with finite $4+\delta$ moments (note that trivially by Assumption 1 the elements of $\mathbf{b}_{3t}$ are independent of the elements of $\mathbf{b}_{4t}$).

Denote the matrix $\mathbf{B}_{3s} = \mathbf{A}^s (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} \mathbf{R}_s$ and note that from Assumptions 1 and 2b it follows that

$$\begin{aligned} \sum_{s=0}^{\infty} \mathbf{B}_{3s} &\leq \mathbf{A}^s (\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1} k_r \\ &\leq \mathbf{A}^s k_1 k_r = (\mathbf{I}_{Nk} - \mathbf{A})^{-1} k_1 k_r \leq k_2 k_1 k_r < \infty, \end{aligned} \tag{A.11}$$

where k_r is the uniform bound for absolute row sums of matrices $\mathbf{R}_t$, and k_1 and k_2 are uniform bounds for absolute row sums of matrices $(\mathbf{I}_{kN} - \mathbf{\Lambda}_0 \mathbf{W})^{-1}$ and $(\mathbf{I}_{kN} - \mathbf{A})^{-1}$. Given Assumption 1, the elements of $\mathbf{b}_{3t}$ 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 form Assumption 4 and Theorem 5.6.12 it follows that

$$\lim_{t\to\infty} \mathbf{A}^t = \mathbf{0}, \tag{A.12}$$

and hence given Assumption 5, we have that elements of $\mathbf{b}_{4t}$ converge in quadratic means to zero.

Therefore the random variable $\mathbf{x}_\infty$ is well defined and we have

$$\begin{aligned}\mathbf{x}_\infty &= \lim_{t\to\infty} \mathbf{B}_{0t}\mathbf{a}_0 + \sum_{s=0}^{\infty} \mathbf{A}^s \left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1} \boldsymbol{\varepsilon}_s \qquad (A.13)\\ &= \left(\mathbf{I}_{kN} - \mathbf{A}\right)^{-1} \left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1} \mathbf{a}_0 + \sum_{s=0}^{\infty} \mathbf{A}^s \left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1} \boldsymbol{\varepsilon}_s.\end{aligned}$$

Hence

$$E\left(\mathbf{x}_\infty\right) = \left(\mathbf{I}_{kN} - \mathbf{A}\right)^{-1} \left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1} \mathbf{a}_0, \qquad (A.14)$$

and using the independence of $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\varepsilon}_s$ for $t \neq s$:

$$VC\left(\mathbf{x}_\infty\right) = \sum_{s=0}^{\infty} \mathbf{A}^s \left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1} \mathbf{\Omega}_\varepsilon \left(\mathbf{I}_{kN} - \mathbf{W}'\mathbf{\Lambda}_0'\right)^{-1} \mathbf{A}^{s\prime}. \qquad (A.15)$$

Finally,using Lemma A1, we find that

$$\begin{aligned}&vech\left[VC\left(\mathbf{x}_\infty\right)\right] \qquad (A.16)\\ =\; &\left\{\mathbf{I}_{N^2k^2} - \left[\mathbf{A}\left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1} \otimes \mathbf{A}\left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1}\right]\right\}^{-1} \mathbf{D} \cdot vech\left(\mathbf{\Omega}_\varepsilon\right),\end{aligned}$$

where $\mathbf{D}$ is a duplication matrix.

A.2 Proof of Proposition 2

Observe that by (2.18) and the assumption in the proposition we have

$$\begin{aligned}\rho\left[\left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1} \mathbf{\Phi}\right] &\le \left\|\left(\mathbf{I}_{kN} - \mathbf{\Lambda}_0\mathbf{W}\right)^{-1}\right\|_1 \cdot \left\|\mathbf{\Phi} + \mathbf{\Lambda}_1\mathbf{W}\right\|_1\\ &\le \left[\sum_{s=0}^{\infty} \left(k_w \left\|\mathbf{\Lambda}_0\right\|_1\right)^s\right] \left[\left\|\mathbf{\Phi}\right\|_1 + k_w \left\|\mathbf{\Lambda}_1\right\|_1\right]\\ &= \frac{\left\|\mathbf{\Phi}\right\|_1 + k_w \left\|\mathbf{\Lambda}_1\right\|_1}{1 - k_w \left\|\mathbf{\Lambda}_0\right\|_1}. \qquad (A.17)\end{aligned}$$

Next note that from the condition in the proposition ($\left\|\mathbf{\Phi}\right\|_1 + k_w \left\|\mathbf{\Lambda}_1\right\|_1 + k_w \left\|\mathbf{\Lambda}_0\right\|_1 < 1$) it follows that $\left\|\mathbf{\Phi}\right\|_1 + k_w \left\|\mathbf{\Lambda}_1\right\|_1 < 1 - k_w \left\|\mathbf{\Lambda}_0\right\|_1$ and thus (observe that the condition also implies that $k_w \left\|\mathbf{\Lambda}_0\right\|_1 < 1$, thus also $1 - k_w \left\|\mathbf{\Lambda}_0\right\|_1 > 0$)

$$\frac{\left\|\mathbf{\Phi}\right\|_1 + k_w \left\|\mathbf{\Lambda}_1\right\|_1}{1 - k_w \left\|\mathbf{\Lambda}_0\right\|_1} < 1, \qquad (A.18)$$

which proves the claim.

A.3 Proof of Theorem 1

[To be added]

References

[1] Amemiya, T. and T.E. MaCurdy, 1986, Instrumental Variable Estimation of an Error-Components Model, *Econometrica*, v. 54, iss. 4, pp. 869-80.

[2] Ahn, S.C., and P. Schmidt, 1995, Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-27.

[3] Anselin, L., 1988, *Spatial Econometrics: Methods and Models* (Kluwer Academic Publishers, Boston).

[4] Arrelano, M., and S.R. Bond, 1991, Some Tests of Specification for Panel Data: Monte Evidence and an Application to Employment Equations, *Review of Economic Studies*, 58, 277-297.

[5] Arrelano, M., and O. Bover, 1995, Another Look at the Instrumental Variable Estimation of Error-Components Models, *Journal of Econometrics*, 68, 29-51.

[6] Binder, M., C. Hsiao, and M.H. Pesaran, 2002, Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration, Working Paper, University of Maryland, College Park.

[7] Binder, M., C. Hsiao, J. Mutl, and M.H. Pesaran, 2002, Computational Issues in the Estimation of Higher-Order Panel Vector Autoregressions, mimeo, University of Maryland, College Park.

[8] Blundell, R., and S.R. Bond, 1998, Initial Conditions and Moment Restrictions in Dynamic Panel Data Models, *Journal of Econometrics*, 82, 135-156.

[9] Dees, S., F. di Maurio, H.M. Pesaran and L.V. Smith, 2004, Exploring the International Linkages of the Euro Area: a Global VAR Analysis, *IEPR Working Paper*, 04.6.

[10] Hamilton, J.D., 1994, *Time Series Analysis*, Princeton University Press.

[11] Horn, R.A. and C.R. Johnson, 1985, *Matrix Analysis*, Cambridge University Press.

[12] Kapoor, M. H.H. Kelejian and I.R.Prucha, 2003, Panel Data Models with Spatially Correlated Error Components, Department of Economics, University of Maryland, College Park, mimeo.

[13] Kelejian, H.H. and I.R. Prucha, 1998, A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances, *Journal of Real Estate Finance and Economics* 17, 99-121.

[14] Kelejian, H.H. and I.R. Prucha, 1999, A generalized moments estimator for the autoregressive parameter in a spatial model, *International Economic Review* 40, 509-533.

[15] Kelejian, H.H. and I.R. Prucha, 2001a, On the asymptotic distribution of the Moran I test statistic with applications, *Journal of Econometrics* 104, 219-257.

[16] Kelejian, H.H. and I.R. Prucha, 2003, Specificaiton and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroscedastic Disturbances, Department of Economics, University of Maryland, College Park, mimeo.

[17] Kelejian, H.H., I.R. Prucha and Y. Yuzefovich, 2003,Instrumental Variable Estimation of a Spatial Autorgressive Model with Autoregressive Disturbances: Large and Small Sample Results, University of Maryland, College Park, mimeo.

[18] Magnus, J.R. and H. Neudecker, 1988, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, John Wiley & Sons.

[19] Mutl, J., 2005, *Dynamic Panel Data Models with Spatially Correlated Disturbances*, PhD Thesis, University of Maryland, College Park.

[20] Pesaran, H.M., T. Shuermann and S.M. Weiner, 2002, MOdeling Regional Interdependencies using a Global Error-Correcting Macroeconometric Model, *Wharton Working Paper Series*, 01-38-B.

[21] Pesaran, H.M., L.V. Smith and R.P. Smith, 2005, What if the UK Had Joined thr Euro in 1999? An Empirical Investigation Using a Global VAR, *CES-Ifo Working Paper*, No. 1477.

[22] Pesaran, H.M. and R. Smith, 2006, Macroeconometric Modelling with a Global Perspective, *IEPR Working Paper*, No. 06.43.