

Panel VAR Models with Spatial Dependence*

Jan Mutl
Institute of Advanced Studies
Stumpergasse 56
A-1060 Vienna
Austria
mutl@ihs.ac.at

January 30, 2009

Abstract

I consider a panel vector autoregressive (panel VAR) model with cross-sectional dependence of the model disturbances that can be characterized by a first order spatial autoregressive process. I describe and discuss several alternative estimation strategies. First, I consider a computationally simple three-step procedure. Its first step consists of applying an instrumental variable estimation that ignores the spatial correlation of the disturbances. In the second step, the estimated disturbances are used to infer the degree of spatial correlation. The method suggested in this paper is a multivariate extension of the spatial generalized moments estimation. The final step of the procedure uses transformed data and applies standard techniques for estimation of panel vector-autoregressive models. When a quasi-maximum likelihood procedure is used, the procedure essentially becomes a constrained likelihood estimation. When the last step uses moment estimation, the procedure is then a multivariate spatial generalized method of moments estimation. Finally, as an alternative, I describe a full (quasi) maximum likelihood estimation method. I conclude by comparing the small-sample performance of the various estimation strategies in a Monte Carlo study.

Keywords: spatial panel vector-autoregression, spatial PVAR, multivariate, dynamic panel data model, spatial GM, spatial Cochrane-Orcutt transformation, constrained maximum likelihood estimation

JEL Codes: C13, C31, C33

*I am grateful to Michael Binder, Harry Kelejian and Ingmar Prucha for helpful advice and comments. However, all errors and omissions are my responsibility.

1 Introduction

Vector autoregressive (VAR) models are extensively used in econometric applications in a wide variety of fields. The extension to panel data represents an interesting challenge due to the likely presence of cross-sectional heterogeneity. In this paper I tackle the issue by considering a panel VAR model with fixed time dimension T . When the cross-sectional dimension is fixed, one has to parsimoniously parameterize the correlations across cross-sectional units in order to avoid the incidental parameters problem. In this paper I follow the spatial econometrics literature and study a first order spatial autocorrelation model with a known spatial weighting matrix.

The panel spatial autocorrelation model is a generalization of spatial econometric models that include the single equation models, e.g., Cliff and Ord (1973, 1981), and the simultaneous equation models, such as Whittle (1954), Anselin (1988) or Kelejian and Prucha (1998, 1999 and 2004). Lee (2004) provides formal large sample results for a (quasi) maximum likelihood estimator of a static single cross-section model. Extensions to panel data with single equation include Lee and Yu (2008) who extend the formal results for the maximum likelihood estimation for static panel data models, and Kapoor et al. (2007) who introduce and derive formal large sample results for the spatial generalized moments method.

On the other hand, the current paper extends the panel VAR literature to allow for cross-sectional dependence of the model disturbances; for models with homogeneous disturbances see, e.g., Binder et al. (2005) for the quasi maximum likelihood (QML) and minimum distance (MD) estimators, or Arellano and Bond (1991), Ahn and Schmidt (1995) or Arellano and Bover (1995) for the generalized method of moments (GMM) approach in a single equation framework.

The next section will specify the model and state the assumptions maintained throughout the paper. Section XX describes the various estimation procedures, while Section XX presents the results from a Monte Carlo study comparing small-sample performance of these estimators. Section XX then concludes.

2 The Panel VAR Model

In this section I specify the model and discuss the main assumptions that will be maintained throughout. The specification adopted here uses the spatial autoregressive framework with known spatial weighting matrix to capture the heteroscedasticity in the data. Hence it replaces the assumption that the disturbances of the model are independent among units and as such can be viewed as an alternative to other approaches such a principle component models.

The model under consideration can be expressed as a first order panel VAR model:

$$\begin{aligned} \mathbf{y}_{it} &= \mathbf{\Phi} \mathbf{y}_{i,t-1} + \mathbf{u}_{it}, \\ \mathbf{u}_{it} &= \lambda \sum_{j=1}^N \mathbf{w}_{ij} \mathbf{u}_{jt} + (\mathbf{I}_m - \mathbf{\Phi}) \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{it} \end{aligned} \tag{2.1}$$

where the first subscript $i \in \{1, \dots, N\}$ refers to the cross-sectional dimension and the second subscript $t \in \{1, \dots, T\}$ refers to the time dimension of the panel of observations $\{\mathbf{y}_{it}\}_{1 \leq i \leq N, 1 \leq t \leq T}$.

I also allow the model to contain more than one equation and so the observations \mathbf{y}_{it} , the individual-specific effects $\boldsymbol{\mu}_i$ and the disturbances \mathbf{u}_{it} and $\boldsymbol{\varepsilon}_{it}$ are $m \times 1$ vectors and the known weighting parameters \mathbf{w}_{ij} , the unknown model parameters $\boldsymbol{\Phi}$ and the identity matrix \mathbf{I}_m are all $m \times m$ matrices. The degree of spatial autocorrelation is captured by the scalar parameter λ . Note that I restrict the individual effects to be of the form $(\mathbf{I}_m - \boldsymbol{\Phi})\boldsymbol{\mu}_i$ so that when the model contains units roots (for example when $\boldsymbol{\Phi} = \mathbf{I}_m$), the trending behavior remains the same as in the stationary case.

Stacking across individuals we obtain

$$\begin{aligned}\mathbf{y}_t &= (\mathbf{I}_N \otimes \boldsymbol{\Phi}) \mathbf{y}_{t-1} + \mathbf{u}_t, \\ \mathbf{u}_t &= \lambda \mathbf{W} \mathbf{u}_t + [\mathbf{I}_N \otimes (\mathbf{I}_m - \boldsymbol{\Phi})] \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t,\end{aligned}\tag{2.2}$$

where

$$\begin{aligned}\mathbf{y}_t &= (\mathbf{y}'_{1t}, \dots, \mathbf{y}'_{Nt})', & \boldsymbol{\mu} &= (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_N)' \\ & \text{\scriptsize } mN \times 1 & & \text{\scriptsize } mN \times 1 \\ \mathbf{u}_t &= (\mathbf{u}'_{1t}, \dots, \mathbf{u}'_{Nt})', & \boldsymbol{\varepsilon}_t &= (\boldsymbol{\varepsilon}'_{1t}, \dots, \boldsymbol{\varepsilon}'_{Nt})' \\ & \text{\scriptsize } mN \times 1 & & \text{\scriptsize } mN \times 1\end{aligned}\tag{2.3}$$

and the $mN \times mN$ weighting matrix \mathbf{W} is

$$\mathbf{W} = \begin{pmatrix} \mathbf{w}_{11} & \cdots & \mathbf{w}_{1N} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{N1} & \cdots & \mathbf{w}_{NN} \end{pmatrix}_{mN \times mN}\tag{2.4}$$

Solving for the disturbance terms yields¹

$$\mathbf{u}_t = (\mathbf{I}_{mN} - \lambda \mathbf{W})^{-1} ([\mathbf{I}_N \otimes (\mathbf{I}_m - \boldsymbol{\Phi})] \boldsymbol{\mu} + \boldsymbol{\varepsilon}_t).\tag{2.5}$$

To facilitate identification of the model, I assume that there is no spatial correlation across equations, that is each $m \times m$ matrix \mathbf{w}_{ij} is diagonal. However, the model allows for contemporaneous correlation across equations in different cross-sections because the variance-covariance matrix of the error terms $\boldsymbol{\varepsilon}_{it}$ is left unrestricted.²

2.1 Random vs. Fixed Effects Specification

Allowing for individual effects without any additional restrictions, leads to an incidental parameters problem. As the time dimension of the panel is fixed, one cannot consistently estimate a general form of the individual-specific effects with a finite number of observations per parameter. To resolve this problem, there are two options. Either to assume that there is a well-behaved distribution (e.g. with finite fourth moments) from which the individual-specific effects are generated (the random effects specification), or transform the data to

¹Note that this assumes that the inverse of $(\mathbf{I}_{mN} - \lambda \mathbf{W})$ exists. This will indeed be the case under the assumptions maintained in the paper.

²There is cross-equation correlation for a single cross-section and since the cross-sections are spatially correlated, the error terms in different equations for different cross-sections will be contemporaneously correlated.

obtain specification that does not contain the individual-specific effects (the fixed effect specification). The usual approach in the fixed effect specification is to first-difference the data; see the argument in Hsiao, Pesaran and Tahmiscioglu (2002) who show in a univariate context that the QML estimator is invariant to the choice of the transformation matrix that eliminates the individual-specific effects. The argument is readily extended to the multivariate setting in this paper. However, the fixed effect specification and first-differencing does not eliminate the incidental parameter problem unless we assume that the spatial weighting matrices are constant over time. Hence the choice between fixed and random effects specification depends on which of the two assumptions (constant weighting matrix or existence of the distribution that generates the individual-specific effects) is more appropriate.

In this paper I use the transformed likelihood approach (e.g. fixed effects specification). Nevertheless, the initial estimation procedure I suggest in this paper, uses a random effects specification. In particular, in the second step of the procedure is a spatial generalized moments method that uses estimated disturbances from the first step. The spatial GM estimation provides estimates of the degree of spatial autocorrelation in the disturbances, as well as an estimates of the variance covariance matrices of both the independent innovations and the individual effects. Hence an extension of the likelihood approach to include individual effects would be straightforward.

2.2 Initial Disturbances Specification

Instead of conditioning on initial observations, I explicitly treat the initial conditions when defining the likelihood function. There are several assumptions one can make. Since the data is not observed beyond 0, the initial observations \mathbf{y}_0 are just equal to the initial disturbances (which now potentially include all the lagged effects), i.e.

$$\mathbf{y}_0 = \mathbf{u}_0. \quad (2.6)$$

I assume that \mathbf{u}_0 is spatially correlated and is generated by

$$\mathbf{u}_0 = \lambda \mathbf{W} \mathbf{u}_0 + \boldsymbol{\mu} + \boldsymbol{\xi}, \quad (2.7)$$

where $\boldsymbol{\xi} = (\boldsymbol{\xi}'_1, \dots, \boldsymbol{\xi}'_N)'$ with each $\boldsymbol{\xi}_i$ being is an $m \times 1$ vector of independently (of $\boldsymbol{\mu}_j$ and $\boldsymbol{\varepsilon}_{jt}$) and identically distributed initial random disturbances.

Hence the initial observations in first differences are

$$\begin{aligned} \Delta \mathbf{y}_1 &= (\mathbf{I}_N \otimes \boldsymbol{\Phi}) \mathbf{y}_0 + \mathbf{u}_1 - \mathbf{y}_0 \\ &= \mathbf{u}_1 - [\mathbf{I}_N \otimes (\mathbf{I}_m - \boldsymbol{\Phi})] \mathbf{u}_0 \\ &= (\mathbf{I}_{mN} - \lambda \mathbf{W})^{-1} (\boldsymbol{\varepsilon}_1 - [\mathbf{I}_N \otimes (\mathbf{I}_m - \boldsymbol{\Phi})] \boldsymbol{\xi}). \end{aligned} \quad (2.8)$$

I denote by $\boldsymbol{\Psi}$ the variance covariance matrix of the initial observation in first differences ($\Delta \mathbf{y}_1$) after the spatial autocorrelation is removed, i.e.

$$\boldsymbol{\Psi} = VC((\mathbf{I}_{mN} - \lambda \mathbf{W}) \Delta \mathbf{y}_1), \quad (2.9)$$

where $VC(\cdot)$ stands for variance covariance matrix. Given that $\boldsymbol{\Phi} \neq \mathbf{I}_m$, we have that

$$\boldsymbol{\Psi} = \boldsymbol{\Omega}_\varepsilon + (\mathbf{I}_m - \boldsymbol{\Phi}) \cdot VC(\boldsymbol{\xi}_j) \cdot (\mathbf{I}_m - \boldsymbol{\Phi}'), \quad (2.10)$$

and hence Ψ is unconstrained and the entries in it will enter as additional parameters into the likelihood function. In the pure unit root case ($\Phi = \mathbf{I}_m$), the variance of the initial innovations is constrained to be $\Psi = \Omega_\varepsilon$.

In general, if the eigenvalues of Φ are inside the unit circle, one could make further assumptions on the ξ disturbances and express Ψ in terms of Φ and Ω_ε . In particular, since in this case the data generating process is stationary and, therefore, one could assume that it has started in an infinite past. This would imply that the initial observations \mathbf{y}_0 are drawn from the limiting stationary distribution of the process, e.g. that:

$$\mathbf{y}_0 = (\mathbf{I}_{mN} - \lambda \mathbf{W})^{-1} \sum_{j=0}^{\infty} (\mathbf{I}_N \otimes \Phi)^{j-1} ([\mathbf{I}_N \otimes (\mathbf{I}_m - \Phi)] \boldsymbol{\mu} + \boldsymbol{\varepsilon}_{-j}) \quad (2.11)$$

Therefore, the initial observations in first differences are

$$\Delta \mathbf{y}_1 = (\mathbf{I}_{mN} - \lambda \mathbf{W})^{-1} \sum_{j=0}^{\infty} (\mathbf{I}_N \otimes \Phi)^{j-1} \Delta \boldsymbol{\varepsilon}_{-j} \quad (2.12)$$

and

$$\begin{aligned} VC[(\mathbf{I}_{mN} - \lambda \mathbf{W}) \Delta \mathbf{y}] &= \mathbf{I}_N \otimes VC \left(\sum_{j=0}^{\infty} \Phi^{j-1} \Delta \boldsymbol{\varepsilon}_{i,-j} \right) \\ &= \mathbf{I}_N \otimes VC \left(\boldsymbol{\varepsilon}_{i0} + (\mathbf{I}_m - \Phi) \sum_{j=0}^{\infty} \Phi^j \boldsymbol{\varepsilon}_{i,-j-1} \right) \\ &= \mathbf{I}_N \otimes \left[\Omega_\varepsilon + (\mathbf{I}_m - \Phi) \left(\sum_{j=0}^{\infty} \Phi^j \Omega_\varepsilon \Phi^j \right) (\mathbf{I}_m - \Phi') \right], \end{aligned} \quad (2.13)$$

or given the definition of Ψ , we have

$$\Psi = \left[\Omega_\varepsilon + (\mathbf{I}_m - \Phi) \left(\sum_{j=0}^{\infty} \Phi^j \Omega_\varepsilon \Phi^j \right) (\mathbf{I}_m - \Phi') \right]. \quad (2.14)$$

Hence, I distinguish three assumptions on how the elements of Ψ are determined:

- (UR) In the pure unit root case ($\Phi = \mathbf{I}_m$), we have to set $\Psi = \Omega_\varepsilon$.
- (IOR) When all of the eigenvalues of Φ are inside the unit circle, we could impose an additional assumption and restrict the elements of Ψ to be a function of Ω_ε and Φ , i.e. or rewriting the expression in the equation above:

$$vec \Psi = \mathbb{D}vec \Omega_\varepsilon + [(\mathbf{I}_m - \Phi) \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I} - (\Phi \otimes \Phi)]^{-1} \mathbb{D}vec \Omega_\varepsilon, \quad (2.15)$$

where \mathbb{D} is a duplication matrix such that $vec \Omega_\varepsilon = \mathbb{D}vec \Omega_\varepsilon$.

- (NR) No restrictions are placed on elements of Ψ (other than imposing that Ψ is symmetric and strictly positive definite matrix).

In all of the cases, we have that the variance covariance matrix of the first difference of the initial observations is

$$E(\Delta \mathbf{y}_1 \Delta \mathbf{y}_1') = (\mathbf{I}_{mN} - \lambda \mathbf{W})^{-1} \Psi (\mathbf{I}_{mN} - \lambda \mathbf{W}')^{-1}. \quad (2.16)$$

2.3 Maintained Assumptions

In order to guarantee that the model and the estimation procedure is well defined, I maintain the following assumptions about the disturbances and the spatial weighting matrices.

Assumption 1 *The disturbance vectors $\boldsymbol{\varepsilon}_{it}$ are identically and independently (of $\boldsymbol{\varepsilon}_{jt}$ for $j \neq i$) distributed with zero mean, and finite absolute $4 + \delta$ moments for some $\delta > 0$. Furthermore, the vector $\boldsymbol{\varepsilon}_{it}$ has a finite positive-definite variance matrix $\boldsymbol{\Omega}_\varepsilon$.*

The above assumption is needed to ensure that the observable data, which is a transformation of the $\boldsymbol{\varepsilon}_{it}$ process, has a well-defined asymptotic properties.

The next two assumptions ensure that the weighting matrices do not 'explode' as the sample size increases.

Assumption 2 *The matrices $(\mathbf{I}_{mN} - \gamma\mathbf{W})$ are nonsingular for all $|\gamma| < 1/\rho(\mathbf{W})$, where $\rho(\cdot)$ denotes the spectral radius of a matrix. Furthermore, the parameter λ also satisfies $|\lambda| < 1/\rho(\mathbf{W})$.*

Assumption 3 *The row and column sums of the matrices \mathbf{W} and $(\mathbf{I}_{mN} - \lambda\mathbf{W})^{-1}$ are uniformly bounded in absolute value.*

3 Estimation

The model can be estimated using a variety of approaches. Straightforward least squares estimation of the first differences of the observations on its lagged values is not consistent because the error term $\Delta\mathbf{u}_t$ is correlated with the explanatory variable $\Delta\mathbf{y}_{t-1}$. However, based on results in Mutl (2006), there is an alternative instrumental variable (IV) estimation that leads to a consistent estimates of the spatially correlated disturbances. Given an initial estimator of the slope coefficients, we can then use a spatial generalized moments estimation (spatial GM) to obtain a consistent estimator of the spatial parameter λ ; e.g. use the moment conditions based on the estimated disturbances:

$$\hat{\mathbf{u}}_t = \mathbf{y}_t - \left(\mathbf{I}_N \otimes \hat{\boldsymbol{\Phi}}_{IV} \right) \mathbf{y}_{t-1} \quad (3.1)$$

where $\hat{\boldsymbol{\Phi}}_{IV}$ is the IV estimator of $\boldsymbol{\Phi}$. Generalizing the univariate results in Kapoor et al. (2007), it then follows that this two stage procedure leads to a consistent estimator of λ .

Finally, in the last step of the proposed estimation procedure, we can use the spatial Cochrane-Orcutt transformation and write the model as³

$$(\mathbf{I}_{mN} - \lambda\mathbf{W}) \Delta\mathbf{y}_t = (\mathbf{I}_{mN} - \lambda\mathbf{W}) (\mathbf{I}_N \otimes \boldsymbol{\Phi}) \Delta\mathbf{y}_{t-1} + \Delta\boldsymbol{\varepsilon}_t. \quad (3.2)$$

If λ is known, the transformed model can be estimated with standard techniques, such as the quasi-maximum likelihood (QML) method in Binder, et al. (2005) or a multivariate

³It would also be possible to use the full spatial panel GLS tranformation since the spatial GM procedure also provides estimates of $\boldsymbol{\Omega}_\varepsilon$ and $\boldsymbol{\Omega}_\mu$. Nevertheless, the tranformed likelihood approach on a model after the spatial Cochrane-Orcutt transformation does not depend on the variance of the individual effects.

extension of the generalized method of moments (GMM) approach of Arellano and Bond (1991), Ahn and Schmidt (1995), or Arellano and Bover (1995).

An alternative to the above procedure is to use the maximum likelihood function of the entire model and obtain a QML estimates. Given the computational complexity of this approach, is important to have reasonable initial estimates. Hence even if one ultimately employs the full likelihood approach, it is of interest to study the properties of the initial estimators. In the following, I first define the IV estimator and then discuss the spatial GM estimator of the spatial parameter. Finally, I define the full as well as the constrained QML procedures and show that the spatial parameter is a nuisance parameter.

3.1 Initial Estimation

Unlike the transformed likelihood approach, the initial estimators are based on moment conditions that involve (lagged) levels of the endogenous variable. Therefore, their large (as well as small) sample propererties are not independent of the distribution of the individual effects. Similarly, the spatial GM procedure is based on estimated levels of the disturbances and directly uses a random effects assumption. Hence I maintain the following assumption:

Assumption 4 *The disturbance vectors $\boldsymbol{\mu}_i$ are identically and independently (of $\boldsymbol{\varepsilon}_{jt}$ and $\boldsymbol{\mu}_j$) distributed with zero mean, and finite absolute $4 + \delta$ moments for some $\delta > 0$. Furthermore, the vector $\boldsymbol{\mu}_{it}$ has a finite positive-definite variance matrix $\boldsymbol{\Omega}_\mu$.*

3.1.1 Instrumental Variable Estimator of the Slope Coefficients

To be able to define the IV estimator, it turns out to be convenient to stack the model differently. Our model is:

$$\Delta \mathbf{y}_{it} = \boldsymbol{\Phi} \Delta \mathbf{y}_{i,t-1} + \Delta \mathbf{u}_{it} \quad (3.3)$$

where $\Delta \mathbf{y}_{it}$ and $\Delta \mathbf{u}_{it}$ are $m \times 1$ vectors. After taking transpose and staking the observations at different times for a given cross-section, we have

$$\begin{pmatrix} \Delta \mathbf{y}'_{i1} \\ \vdots \\ \Delta \mathbf{y}'_{iT} \end{pmatrix}_{T \times m} = \begin{pmatrix} \Delta \mathbf{y}'_{i0} \\ \vdots \\ \Delta \mathbf{y}'_{i,T-1} \end{pmatrix}_{T \times m} \boldsymbol{\Phi}'_{m \times m} + \begin{pmatrix} \Delta \mathbf{u}'_{i1} \\ \vdots \\ \Delta \mathbf{u}'_{iT} \end{pmatrix}_{T \times m} \quad (3.4)$$

or with the obvious notation

$$\Delta \mathbf{Y}_i = \Delta \mathbf{Y}_{i,-1} \boldsymbol{\Phi}' + \Delta \mathbf{U}_i \quad (3.5)$$

Stacking the cross-sections yields

$$\Delta \mathbf{Y} = \Delta \mathbf{Y}_{-1} \boldsymbol{\Phi}' + \Delta \mathbf{U} \quad (3.6)$$

where $\Delta \mathbf{Y} = (\Delta \mathbf{Y}'_1, \dots, \Delta \mathbf{Y}'_N)'$, $\Delta \mathbf{Y}_{-1} = (\Delta \mathbf{Y}'_{1,-1}, \dots, \Delta \mathbf{Y}'_{N,-1})'$ and $\Delta \mathbf{U} = (\Delta \mathbf{U}'_1, \dots, \Delta \mathbf{U}'_N)'$.

We define the IV estimator of $\boldsymbol{\Phi}$ as

$$\widehat{\boldsymbol{\Phi}}_{IV} = \left[\widehat{\mathbf{Z}}' \widehat{\mathbf{Z}} \right]^{-1} \widehat{\mathbf{Z}}' \Delta \mathbf{Y} \quad (3.7)$$

where $\widehat{\mathbf{Z}} = \mathbf{P}_H \Delta \mathbf{Y}$ with $\mathbf{P}_H = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$ where \mathbf{H} is vector of instruments used for $\Delta \mathbf{Y}_{-1}$. I suggest the use of the instruments $\mathbf{H} = \mathbf{Y}_{-2} = (\mathbf{Y}'_{1,-2}, \dots, \mathbf{Y}'_{N,-2})'$ where $\mathbf{Y}_{i,-2} = (\mathbf{y}_{i,-1}, \dots, \mathbf{y}_{i,T-2})'$. However, any instruments that satisfy the following conditions lead to consistent estimates of the spatially correlated disturbances.

Assumption 5 *The instrument matrix \mathbf{H} has a full column rank.*

Assumption 6 *The instruments satisfy the following:*

1. $p \lim \frac{1}{N} \mathbf{H}'\mathbf{H} = \mathbf{Q}_{HH}$ where \mathbf{Q}_{HH} is finite and nonsingular;
2. $p \lim \frac{1}{N} \mathbf{H}'\Delta \mathbf{Y} = \mathbf{Q}_{HY}$ where \mathbf{Q}_{HY} is finite and has a full column rank;
3. *The instruments \mathbf{H} can be expressed as $\mathbf{H} = \mathbf{F}(\boldsymbol{\varsigma}_1, \dots, \boldsymbol{\varsigma}_m)$ where each $\boldsymbol{\varsigma}_i$ is a $NT \times 1$ vector of identically and independently distributed random variables and \mathbf{F} is an $N \times N$ nonstochastic absolutely summable matrix. Furthermore, each $\boldsymbol{\varsigma}_i$ is independent of $\boldsymbol{\varepsilon}_{it}$.*

The first two assumptions guarantee that the instruments are not degenerate and that they are asymptotically correlated with the variables they replace. The last assumption implies that the instruments are not correlated with the error terms and that a central limit theorem for triangular arrays of quadratic forms can be applied. Given these additional assumptions we can assert that the IV estimation produces consistent estimates:

Proposition 1 *Given the model and assumptions 1-7, the IV estimator is consistent and the rate of convergence is $N^{-1/2}$; that is $\widehat{\boldsymbol{\Phi}}_{IV} = \boldsymbol{\Phi} + O_p(N^{-1/2})$.*

Remark: The rate of convergence is important for consistency of estimation λ (the degree of spatial correlation in the residuals) in the the second step of the procedure.

Note that our suggested instruments meet the required conditions. By backward substitution we can eliminate the lagged dependent variables and express the instruments as a function of lagged disturbance terms and lagged explanatory variables. It is easily verified that

$$\Delta \mathbf{y}_t = (\mathbf{I}_{mN} - \lambda \mathbf{W})^{-1} \left(\sum_{j=0}^{t-2} (\mathbf{I}_N \otimes \boldsymbol{\Phi})^{j-1} \Delta \boldsymbol{\varepsilon}_{t-j} + (\mathbf{I}_N \otimes \boldsymbol{\Phi})^{t-1} [\boldsymbol{\varepsilon}_1 - (\mathbf{I}_m - \boldsymbol{\Phi}) \boldsymbol{\xi}] \right) \quad (3.8)$$

and hence we have that

$$\mathbf{H} = (\mathbf{Y}'_{1,-2}, \dots, \mathbf{Y}'_{N,-2})' \quad (3.9)$$

$$= \mathbf{F} \cdot [\boldsymbol{\varepsilon}_1 - (\mathbf{I} - \boldsymbol{\Phi}) \boldsymbol{\xi}, \Delta \boldsymbol{\varepsilon}_2, \dots, \Delta \boldsymbol{\varepsilon}_{T-2}]' \quad (3.10)$$

where

$$\mathbf{F} = \left[\mathbf{I}_T \otimes (\mathbf{I}_{mN} - \lambda \mathbf{W}')^{-1} \right] \left[\mathbf{I}_T \otimes (\mathbf{I}_{mN} - \lambda \mathbf{W})^{-1} \right]$$

Our assumptions on the spatial weighting matrices imply that the $mNT \times mNT$ matrix \mathbf{F} is absolutely summable.

3.1.2 Estimation of the Degree of Spatial Autocorrelation

The second step in the proposed estimation procedure is to use moment conditions based on the estimated disturbances:

$$\hat{\mathbf{u}}_t = \mathbf{y}_t - \left(\mathbf{I}_N \otimes \hat{\Phi}_{IV} \right) \mathbf{y}_{t-1} \quad (3.11)$$

where $\hat{\Phi}_{IV}$ is the IV estimators of Φ . Kelejian and Prucha (1999) show consistency of a similar two stage procedure for univariate single cross-section model with spatial lags in both the dependent variable as well as the error term. Kapoor et al. (2007) extend the results for a univariate static panel model. Note that both of these papers consider nonstochastic exogenous variables and hence their results are not directly applicable to the panel VAR model considered here. However, Mutl (2006) contains a straightforward extension of their proofs for univariate panel autoregressive models. Hence we conjecture that the spatial GM procedure will also be consistent in a multivariate setting (under an appropriate set of assumptions).

To be able to describe the multivariate version of the spatial GM estimation procedure, it proves to be convenient to stack the model differently. It is also possible to impose more structure on the innovations $\boldsymbol{\varepsilon}_{it}$ and, in particular, consider that they are generated from a two-way error component model. Recall that the disturbances of the model are generated from

$$\mathbf{u}_{it} = \lambda \sum_{j=1}^N \mathbf{w}_{ij} \mathbf{u}_{jt} + \boldsymbol{\nu}_{it}. \quad (3.12)$$

We now assume that the innovations have a two-way error components structure

$$\boldsymbol{\nu}_{it} = \boldsymbol{\mu}_i + \boldsymbol{\varepsilon}_{it}, \quad (3.13)$$

where the elements of the $m \times 1$ vectors $\boldsymbol{\mu}_i$ and $\boldsymbol{\varepsilon}_{it}$ are independent with $E(\boldsymbol{\mu}_i \boldsymbol{\mu}_i') = \boldsymbol{\Omega}_\mu$ and $E(\boldsymbol{\varepsilon}_{it} \boldsymbol{\varepsilon}_{it}') = \boldsymbol{\Omega}_\varepsilon$. We can now stack the disturbances and innovations over the different cross-sections. In contrast to Section 1, we define

$$\begin{aligned} \tilde{\mathbf{u}}_t &= (\mathbf{u}_{1t}, \dots, \mathbf{u}_{Nt})', & \tilde{\boldsymbol{\nu}}_t &= (\boldsymbol{\nu}_{1t}, \dots, \boldsymbol{\nu}_{Nt})'. \end{aligned} \quad (3.14)$$

We additionally define the notation for the spatial lag as $\bar{\mathbf{u}}_{it} = \sum_{j=1}^N \mathbf{w}_{ij} \mathbf{u}_{jt}$ and $\bar{\boldsymbol{\nu}}_{it} = \sum_{j=1}^N \mathbf{w}_{ij} \boldsymbol{\nu}_{jt}$. The stacked spatially lagged disturbances and innovations hence become

$$\bar{\tilde{\mathbf{u}}}_t = (\bar{\mathbf{u}}_{1t}, \dots, \bar{\mathbf{u}}_{Nt})', \quad \bar{\tilde{\boldsymbol{\nu}}}_t = (\bar{\boldsymbol{\nu}}_{1t}, \dots, \bar{\boldsymbol{\nu}}_{Nt})'. \quad (3.15)$$

The multivariate version of the spatial GM estimation is based on the following moment conditions (see the Appendix B for their derivation):

$$E \frac{1}{N(T-1)} \tilde{\boldsymbol{\nu}}' \mathbf{Q}_0 \tilde{\boldsymbol{\nu}} = \boldsymbol{\Omega}_\varepsilon, \quad (3.16)$$

$$E \frac{1}{N(T-1)} \bar{\tilde{\boldsymbol{\nu}}}^j \mathbf{Q}_0 \tilde{\boldsymbol{\nu}} = N^{-1} \sum_{i=1}^N \sum_{j=1}^N \mathbf{w}_{ij} \boldsymbol{\Omega}_\varepsilon \mathbf{w}'_{ij},$$

$$E \frac{1}{N(T-1)} \bar{\tilde{\boldsymbol{\nu}}}^j \mathbf{Q}_0 \tilde{\boldsymbol{\nu}} = N^{-1} \sum_{i=1}^N \mathbf{w}_{ii} \boldsymbol{\Omega}_\varepsilon,$$

$$\begin{aligned}
E \frac{1}{N} \tilde{\boldsymbol{\nu}}' \mathbf{Q}_1 \tilde{\boldsymbol{\nu}} &= \boldsymbol{\Omega}_1, \\
E \frac{1}{N} \bar{\tilde{\boldsymbol{\nu}}}' \mathbf{Q}_1 \bar{\tilde{\boldsymbol{\nu}}} &= N^{-1} \sum_{i=1}^N \sum_{j=1}^N \mathbf{w}_{ij} \boldsymbol{\Omega}_1 \mathbf{w}'_{ij}, \\
E \frac{1}{N} \bar{\tilde{\boldsymbol{\nu}}}' \mathbf{Q}_1 \tilde{\boldsymbol{\nu}} &= N^{-1} \sum_{i=1}^N \mathbf{w}_{ii} \boldsymbol{\Omega}_1,
\end{aligned}$$

where \mathbf{Q}_0 and \mathbf{Q}_1 are the within and between transformation matrices defined as

$$\mathbf{Q}_0 = (\mathbf{I}_T - \frac{1}{T} \mathbf{J}_T) \otimes \mathbf{I}_N, \quad \mathbf{Q}_1 = \frac{1}{T} \mathbf{J}_T \otimes \mathbf{I}_N, \quad (3.17)$$

and $\boldsymbol{\Omega}_1 = \boldsymbol{\Omega}_\varepsilon + T \cdot \boldsymbol{\Omega}_\mu$.

To express the above moment conditions in terms of the disturbances $\tilde{\mathbf{u}}$ we note that

$$\tilde{\boldsymbol{\nu}} = \tilde{\mathbf{u}} - \lambda \bar{\tilde{\mathbf{u}}}, \quad \text{and} \quad \bar{\tilde{\boldsymbol{\nu}}} = \bar{\tilde{\mathbf{u}}} - \lambda \bar{\bar{\tilde{\mathbf{u}}}}, \quad (3.18)$$

where $\bar{\bar{\tilde{\mathbf{u}}}} = (\bar{\bar{\tilde{\mathbf{u}}}}_1', \dots, \bar{\bar{\tilde{\mathbf{u}}}}_T')'$ with $\bar{\bar{\tilde{\mathbf{u}}}}_t = (\bar{\bar{\tilde{\mathbf{u}}}}_{1t}, \dots, \bar{\bar{\tilde{\mathbf{u}}}}_{Nt})'$ and $\bar{\tilde{\mathbf{u}}}_{it} = \sum_{j=1}^N \mathbf{w}_{ij} \bar{\tilde{\mathbf{u}}}_{jt}$. The six moment conditions then can be written as

$$\boldsymbol{\Gamma} \cdot [\lambda, \lambda^2, \text{vec}(\boldsymbol{\Omega}_\varepsilon)', \text{vec}(\boldsymbol{\Omega}_1)']' - \boldsymbol{\gamma} = \mathbf{0}, \quad (3.19)$$

where

$$\boldsymbol{\Gamma} = \begin{bmatrix} \gamma_{11}^0 & \gamma_{12}^0 & \gamma_{13}^0 & 0 \\ \gamma_{21}^0 & \gamma_{22}^0 & \gamma_{23}^0 & 0 \\ \gamma_{31}^0 & \gamma_{32}^0 & \gamma_{33}^0 & 0 \\ \gamma_{11}^1 & \gamma_{12}^1 & 0 & \gamma_{13}^1 \\ \gamma_{21}^1 & \gamma_{22}^1 & 0 & \gamma_{23}^1 \\ \gamma_{31}^1 & \gamma_{32}^1 & 0 & \gamma_{33}^1 \end{bmatrix}, \quad \boldsymbol{\gamma} = \begin{bmatrix} \gamma_1^0 \\ \gamma_2^0 \\ \gamma_3^0 \\ \gamma_1^1 \\ \gamma_2^1 \\ \gamma_3^1 \end{bmatrix}, \quad (3.20)$$

with ($i = 1, 2$)

$$\gamma_{11}^i = \frac{1}{N(T-1)^{1-i}} \text{vec} \left(E \tilde{\mathbf{u}}' \mathbf{Q}_i \tilde{\mathbf{u}} + E \bar{\tilde{\mathbf{u}}}' \mathbf{Q}_{i,N} \tilde{\mathbf{u}} \right), \quad \gamma_{12}^i = \frac{-1}{N(T-1)^{1-i}} \text{vec} \left(E \bar{\tilde{\mathbf{u}}}' \mathbf{Q}_i \tilde{\mathbf{u}} \right), \quad (3.21)$$

$$\gamma_{21}^i = \frac{1}{N(T-1)^{1-i}} \text{vec} \left(E \bar{\tilde{\mathbf{u}}}' \mathbf{Q}_i \bar{\tilde{\mathbf{u}}} + E \bar{\tilde{\mathbf{u}}}'_N \mathbf{Q}_i \bar{\tilde{\mathbf{u}}}\right), \quad \gamma_{22}^i = \frac{-1}{N(T-1)^{1-i}} \text{vec} \left(E \bar{\tilde{\mathbf{u}}}' \mathbf{Q}_i \bar{\tilde{\mathbf{u}}}\right),$$

$$\gamma_{31}^i = \frac{1}{N(T-1)^{1-i}} \text{vec} \left(E \tilde{\mathbf{u}}' \mathbf{Q}_i \bar{\bar{\tilde{\mathbf{u}}}} + E \bar{\tilde{\mathbf{u}}}' \mathbf{Q}_i \tilde{\mathbf{u}} \right), \quad \gamma_{32}^i = \frac{-1}{N(T-1)^{1-i}} \text{vec} \left(E \bar{\tilde{\mathbf{u}}}' \mathbf{Q}_i \bar{\bar{\tilde{\mathbf{u}}}} \right),$$

$$\gamma_{13}^i = \mathbf{I}_{m^2}, \quad \gamma_1^i = \frac{1}{N(T-1)^{1-i}} \text{vec} (E \tilde{\mathbf{u}}' \mathbf{Q}_i \tilde{\mathbf{u}}),$$

$$\gamma_{23}^i = \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{w}'_{ij} \otimes \mathbf{w}_{ij}), \quad \gamma_2^i = \frac{1}{N(T-1)^{1-i}} \text{vec} \left(E \bar{\tilde{\mathbf{u}}}' \mathbf{Q}_i \bar{\tilde{\mathbf{u}}}\right),$$

$$\gamma_{33}^i = \frac{1}{N} \sum_{i=1}^N (\mathbf{I}_m \otimes \mathbf{w}_{ii}), \quad \gamma_3^i = \frac{1}{N(T-1)^{1-i}} \text{vec} \left(E \tilde{\mathbf{u}}' \mathbf{Q}_i \bar{\bar{\tilde{\mathbf{u}}}} \right).$$

The multivariate spatial GM procedure is based on the sample counterpart of the six moment conditions above. In particular, given an initial estimate, say $\widehat{\Phi}$, of the slope coefficients, we calculate the projected disturbances ($t = p, \dots, T$)

$$\begin{aligned}\widehat{\mathbf{u}}_{it} &= \mathbf{y}_{it} - \widehat{\Phi} \mathbf{y}_{i,t-1}, \\ \widehat{\mathbf{u}}_{it} &= \sum_{j=1}^N \mathbf{W}_{ij} \mathbf{y}_{jt} \widehat{\mathbf{u}}_{jt}, \\ \widehat{\mathbf{u}}_{it} &= \sum_{j=1}^N \mathbf{W}_{ij} \mathbf{y}_{jt} \widehat{\mathbf{u}}_{jt}.\end{aligned}\tag{3.22}$$

Thus the estimated vectors $\widehat{\mathbf{u}}$, $\widehat{\mathbf{u}}$, and $\widehat{\mathbf{u}}$ have dimensions $N(T-p+1) \times m$ where p is the number of lags in the model (in contrast to e.g. $\widetilde{\mathbf{u}}$ which has dimensions $NT \times m$). However, when the PVAR model only has one lag ($p = 1$), as it is for the case considered in this paper, the dimensions do not change.

The sample analogue of the moment conditions is then based on

$$\mathbf{G} \cdot [\lambda, \lambda^2, \text{vec}(\boldsymbol{\Omega}_\varepsilon)', \text{vec}(\boldsymbol{\Omega}_1)']' - \mathbf{g} = \mathbf{0},\tag{3.23}$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_{11}^0 & \mathbf{g}_{12}^0 & \mathbf{g}_{13}^0 & 0 \\ \mathbf{g}_{21}^0 & \mathbf{g}_{22}^0 & \mathbf{g}_{23}^0 & 0 \\ \mathbf{g}_{31}^0 & \mathbf{g}_{32}^0 & \mathbf{g}_{33}^0 & 0 \\ \mathbf{g}_{11}^1 & \mathbf{g}_{12}^1 & 0 & \mathbf{g}_{13}^1 \\ \mathbf{g}_{21}^1 & \mathbf{g}_{22}^1 & 0 & \mathbf{g}_{23}^1 \\ \mathbf{g}_{31}^1 & \mathbf{g}_{32}^1 & 0 & \mathbf{g}_{33}^1 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} \mathbf{g}_1^0 \\ \mathbf{g}_2^0 \\ \mathbf{g}_3^0 \\ \mathbf{g}_1^1 \\ \mathbf{g}_2^1 \\ \mathbf{g}_3^1 \end{bmatrix},\tag{3.24}$$

with ($i = 1, 2$)

$$\begin{aligned}\mathbf{g}_{11}^i &= \frac{1}{N(T-1)^{1-i}} \text{vec} \left(\widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} + \widehat{\mathbf{u}}' \mathbf{Q}_{i,N} \widehat{\mathbf{u}} \right), \quad \mathbf{g}_{12}^i = \frac{-1}{N(T-1)^{1-i}} \text{vec} \left(\widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} \right) \\ \mathbf{g}_{21}^i &= \frac{1}{N(T-1)^{1-i}} \text{vec} \left(\widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} + \widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} \right), \quad \mathbf{g}_{22}^i = \frac{-1}{N(T-1)^{1-i}} \text{vec} \left(\widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} \right), \\ \mathbf{g}_{31}^i &= \frac{1}{N(T-1)^{1-i}} \text{vec} \left(\widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} + \widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} \right), \quad \mathbf{g}_{32}^i = \frac{-1}{N(T-1)^{1-i}} \text{vec} \left(\widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} \right),\end{aligned}\tag{3.25}$$

$$\begin{aligned}\mathbf{g}_{13}^i &= \mathbf{I}_{m^2}, \quad \mathbf{g}_1^i = \frac{1}{N(T-1)^{1-i}} \text{vec} \left(\widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} \right), \\ \mathbf{g}_{23}^i &= \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N (\mathbf{w}'_{ij} \otimes \mathbf{w}_{ij}), \quad \mathbf{g}_2^i = \frac{1}{N(T-1)^{1-i}} \text{vec} \left(\widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} \right), \\ \mathbf{g}_{33}^i &= \frac{1}{N} \sum_{i=1}^N (\mathbf{I}_m \otimes \mathbf{w}_{ii}), \quad \mathbf{g}_3^i = \frac{1}{N(T-1)^{1-i}} \text{vec} \left(\widehat{\mathbf{u}}' \mathbf{Q}_i \widehat{\mathbf{u}} \right).\end{aligned}$$

The spatial GM procedure then maximizes the objective function (XX) subject to the fact that the matrices Ω_ε and Ω_1 have to be strictly positive definite. This is easily implemented by the variance covariance matrices with their Cholesky decompositions and maximizing only with respect to the $m(m+1)/2$ free parameters in each of the two $m \times m$ variance covariance matrices.

3.2 Quasi Maximum Likelihood (QML) Estimation

The likelihood function for the panel VAR model is easily derived under the assumption that $\varepsilon_{it} \sim N(0, \Omega_\varepsilon)$ where Ω_ε is the $m \times m$ variance-covariance matrix of ε_{ti} . I specify the exact distribution of the initial observations as in Binder et al. (2001) and derive the QML function taking this into account. We can define the $mNT \times 1$ vector

$$\Delta\eta = (\Delta\mathbf{y}'_1, \Delta\mathbf{u}'_2, \dots, \Delta\mathbf{u}'_T)'. \quad (3.26)$$

Recall that the variance of the initial observations is given by

$$E(\Delta\mathbf{y}_1\Delta\mathbf{y}'_1) = (\mathbf{I}_{mN} - \lambda\mathbf{W})^{-1} (\mathbf{I}_N \otimes \Psi) (\mathbf{I}_{mN} - \lambda\mathbf{W}')^{-1}. \quad (3.27)$$

Similarly, we also have that

$$\begin{aligned} E(\Delta\mathbf{u}_t\Delta\mathbf{u}'_t) &= 2(\mathbf{I}_{mN} - \lambda\mathbf{W})^{-1} (\mathbf{I}_N \otimes \Omega_\varepsilon) (\mathbf{I}_{mN} - \lambda\mathbf{W}')^{-1}, \\ E(\Delta\mathbf{u}_t\Delta\mathbf{u}'_{t-1}) &= -(\mathbf{I}_{mN} - \lambda\mathbf{W})^{-1} (\mathbf{I}_N \otimes \Omega_\varepsilon) (\mathbf{I}_{mN} - \lambda\mathbf{W}')^{-1}. \end{aligned} \quad (3.28)$$

Thus, we then have that $E(\Delta\eta) = \mathbf{0}$ and

$$VC(\Delta\eta) = (\mathbf{I}_{mNT} - \lambda\mathbf{W})^{-1} (\mathbf{I}_N \otimes \Sigma) (\mathbf{I}_{mNT} - \lambda\mathbf{W}')^{-1}, \quad (3.29)$$

where $\mathbf{W} = \mathbf{I}_T \otimes \mathbf{W}$, and

$$\Sigma = \begin{pmatrix} \Psi & -\Omega_\varepsilon & & 0 \\ -\Omega_\varepsilon & 2\Omega_\varepsilon & & \\ & & \ddots & -\Omega_\varepsilon \\ 0 & & -\Omega_\varepsilon & 2\Omega_\varepsilon \end{pmatrix}, \quad (3.30)$$

with Ψ being a $m \times m$ symmetric matrix of parameters (under Assumption NR). The (NR) specification leaves the variance-covariance matrix of the initial observations unrestricted - e.g. there are $m(m+1)/2$ free additional parameters.

The likelihood function for the entire sample is then

$$\begin{aligned} L_N(\boldsymbol{\theta}) &= -\frac{mNT}{2} \log(2\pi) - \frac{N}{2} \ln |\Sigma| + \ln |\mathbf{I}_{mNT} - \lambda\mathbf{W}| \\ &\quad - \frac{1}{2} \text{tr} [(\mathbf{I}_N \otimes \mathbf{R}') (\mathbf{I}_{mNT} - \lambda\mathbf{W}) (\mathbf{I}_N \otimes \Sigma^{-1}) (\mathbf{I}_{mNT} - \lambda\mathbf{W}') (\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}], \end{aligned} \quad (3.31)$$

where $\boldsymbol{\theta} = (\text{vech}\Psi', \text{vech}\Omega'_\varepsilon, \text{vec}\Phi', \lambda)$ is the vector of parameters. The $mT \times mT$ matrix \mathbf{R} is defined as

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}_m & & & 0 \\ -\Phi & \mathbf{I}_m & & \\ & & \ddots & \\ 0 & & -\Phi & \mathbf{I}_m \end{pmatrix} \quad (3.32)$$

and the matrix \mathbf{S} is

$$\mathbf{S} = (\Delta \mathbf{y}'_1, \dots, \Delta \mathbf{y}'_T) \cdot (\Delta \mathbf{y}'_1, \dots, \Delta \mathbf{y}'_T)'$$

Under the (IOR) or the (UR) conditions, the vector of parameters is composed of only $\boldsymbol{\vartheta} = (\text{vech}\boldsymbol{\Omega}'_\varepsilon, \text{vec}\boldsymbol{\Phi}', \lambda)$ and the likelihood function is as above but with $\boldsymbol{\Psi}$ being a function of $\boldsymbol{\Omega}_\varepsilon$ and $\boldsymbol{\Phi}$, as described in Section XX.

3.2.1 Computational Issues

The computation of the likelihood function should exploit the structure of the $[\mathbf{I}_T \otimes (\mathbf{I}_N - \lambda \mathbf{W})]$ and $\boldsymbol{\Sigma}$ matrices when evaluating their determinants and inverses. In particular, we can express $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Psi} & (\mathbf{A}_1 \otimes \boldsymbol{\Omega}_\varepsilon) \\ (\mathbf{A}'_1 \otimes \boldsymbol{\Omega}_\varepsilon) & (\mathbf{A}_2 \otimes \boldsymbol{\Omega}_\varepsilon) \end{pmatrix}, \quad (3.33)$$

where \mathbf{A}_1 and \mathbf{A}_2 are matrices of constants. The inverse of $\boldsymbol{\Sigma}_{\Delta\eta}$ is then

$$\boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \mathbf{D}^{-1} & -\mathbf{D}^{-1}(\mathbf{A}_1 \mathbf{A}_2^{-1} \otimes \boldsymbol{\Omega}_\varepsilon) \\ (\mathbf{A}_2^{-1} \mathbf{A}'_1 \otimes \boldsymbol{\Omega}_\varepsilon) \mathbf{D}^{-1} & \mathbf{D}^{-1} - (\mathbf{A}_2^{-1} \mathbf{A}'_1 \otimes \boldsymbol{\Omega}_\varepsilon) \mathbf{D}^{-1} (\mathbf{A}_1 \mathbf{A}_2^{-1} \otimes \boldsymbol{\Omega}_\varepsilon) \end{pmatrix}, \quad (3.34)$$

where $\mathbf{D} = \boldsymbol{\Psi} - (\mathbf{A}_1 \mathbf{A}_2^{-1} \mathbf{A}_1 \otimes \boldsymbol{\Omega}_\varepsilon)$.

3.3 Constrained QML Estimation

Although the QML estimation based on the likelihood function (3.31) is feasible,⁴ it might become extremely computationally intensive. In this paper, I propose an alternative approach that takes a consistent estimator of the spatial correlation parameter λ and maximizes a constrained likelihood function. That is, maximize

$$Q_N(\tilde{\boldsymbol{\theta}}) = -\frac{mNT}{2} \log(2\pi) - \frac{N}{2} \ln |\boldsymbol{\Sigma}| + \ln |\mathbf{I}_{mNT} - \hat{\lambda} \mathbf{W}| - \frac{1}{2} \text{tr} \left[\mathbf{R}' (\mathbf{I}_{mNT} - \hat{\lambda} \mathbf{W}) (\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1}) (\mathbf{I}_{mNT} - \hat{\lambda} \mathbf{W}') \mathbf{R} \mathbf{S}_N \right], \quad (3.35)$$

with respect to $\tilde{\boldsymbol{\theta}} = (\text{vech}\boldsymbol{\Psi}', \text{vech}\boldsymbol{\Omega}'_\varepsilon, \text{vec}\boldsymbol{\Phi}')'$, taking the consistent estimator $\hat{\lambda}$ of λ as given. The consistent estimator of the spatial correlation be based on the two-step procedure proposed above. Note that the constrained likelihood estimator is equivalent to using the spatial Cochrane-Orcutt transformation $(\mathbf{I}_{mN} - \lambda \mathbf{W})$ and then maximizing the QML function derived under the assumption that the disturbances are independent, i.e. the same as in Binder et al. (2005).

⁴The QML estimator is likely to be computationally expensive due to the necessity to calculate eigenvalues of a sparse matrix $(I - \lambda \mathbf{W}_t)$ which is of the dimension N . With large N this becomes a very demanding problem.

4 Monte Carlo Simulations

I now turn to the small sample performance of the different estimators. To this end I replicate the simulations in Binder et al. (2005) who consider the same PVAR model but with independent disturbances. I modify their generation of the disturbances and, in particular, consider the spatial autoregressive specification with a spatial weight matrix \mathbf{W} and parameter λ . The specification of the spatial weights corresponds to the designs used in Kapoor et al. (2007). In particular, I consider three specifications for \mathbf{W} which differ in their degree of sparseness. Each matrix uses a rook design with $J = 2, 6$ or 10 non-zero off-diagonal elements. The parameter λ takes values in the set $\{-9., -5., -.25, 0., .25, .5, .9\}$. I thus have the three different weights matrices and six different values of λ for each of the five simulation designs considered in Binder et al. (2005), i.e. 75 different simulation designs in total. For each of the parameter designs I consider four different sample sizes given by combinations of $T \in \{3, 10\}$ and $N \in \{50, 250\}$. In each simulation design and sample size, I use the same VC matrix for the innovations and the individual effects, i.e. set $\Omega_\varepsilon = \Omega_\mu$ and draw the random variables from a normal distribution. As a robustness check, results are available upon request that use different ratio of the two variances as well as alternative distributions (ch-square and student-t).

Each simulation design and sample size is replicated 1,000 times and the resulting estimators are saved. Table 1 provides the root-mean-square errors of the different estimators. The estimators considered in the experiments are the same estimators as considered in Binder et al. (2005), i.e. the four initial GMM estimators as well as the FE-QML derived under the assumption that the disturbances are independent. Additionally I report results for the three-step procedures described in this paper - the spatial GMM estimators using different sets of moment conditions (corresponding to the same set of moments as the initial GMM estimators) as well as the constrained likelihood approach. As a benchmark, also the results for the full FE-QMLE estimator (taking into account the spatial autocorrelation of the disturbances) are reported.

Since this paper extends the spatial GM procedure to the multivariate context, I also report in Table 2 the performance of the spatial GM estimation. The different spatial GM estimators correspond to whether these are based on the true or estimated disturbances (e.g. corresponding to the four initial estimators) and whether the first three moments or the full set of weighted moment conditions are utilized.

Results TBA.

5 Conclusion

This paper develops an estimation approach for a panel VAR model with spatial dependence. I suggest a three-step estimation procedure. In the first step, instrumental variables procedure is used to consistently estimate the spatially correlated disturbances. In the second step, a method of moments estimation is used to obtain a consistent estimate of the spatial parameter. The final step of the procedure could be either a constrained maximum likelihood procedure or moments estimation based on a model transformed by a spatial Cochrane-Orcutt transformation.

I introduce the constrained maximum likelihood estimator based on a consistent estimate of the spatial dependence parameter and sketch a proof of its consistency when the time dimension of the panel is fixed. In future versions of this paper, I plan to explore the small sample properties of the QML and constrained QML estimators with a Monte Carlo study. It would also be of interest to prove asymptotic normality of the proposed estimator as well as to derive the asymptotic properties of the QML estimator under some reasonable set of assumptions.

A Appendix - Derivatives of the QML Function

To speed up computation, I derive analytical expressions for the partial derivatives of the likelihood function $L_N(\boldsymbol{\theta})$. The first differential is

$$\begin{aligned}
dL_N &= -\frac{N}{2} \text{tr}(\boldsymbol{\Sigma}^{-1} d\boldsymbol{\Sigma}) - \text{tr}[(\mathbf{I}_{mNT} - \lambda \mathbf{W})^{-1} \cdot d\lambda \mathbf{W}] \tag{A.1} \\
&\quad + \frac{1}{2} \text{tr}[(\mathbf{I}_N \otimes \mathbf{R}')(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1} d\boldsymbol{\Sigma} \cdot \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}] \\
&\quad - \frac{1}{2} \text{tr}[(\mathbf{I}_N \otimes d\mathbf{R}')(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}] \\
&\quad - \frac{1}{2} \text{tr}[(\mathbf{I}_N \otimes \mathbf{R}')(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes d\mathbf{R}) \mathbf{S}] \\
&\quad - \frac{1}{2} \text{tr}[(\mathbf{I}_N \otimes \mathbf{R}')(-d\lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}] \\
&\quad - \frac{1}{2} \text{tr}[(\mathbf{I}_N \otimes \mathbf{R}')(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(-d\lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}] \\
&= -\frac{N}{2} \text{vec} \boldsymbol{\Sigma}^{-1} \mathbb{D}_{mT} d\text{vech} \boldsymbol{\Sigma} - \text{vec}(\mathbf{I}_{mNT} - \lambda \mathbf{W}')^{-1} \text{vec} \mathbf{W} \cdot d\lambda \\
&\quad + \frac{1}{2} (\text{vec}[(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}(\mathbf{I}_N \otimes \mathbf{R}')(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})])' \text{vec}(\mathbf{I}_N \otimes d\boldsymbol{\Sigma}) \\
&\quad - \frac{1}{2} (\text{vec}[\mathbf{S}(\mathbf{I}_N \otimes \mathbf{R}')(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')])' \cdot \text{vec}(\mathbf{I}_N \otimes d\mathbf{R}') \\
&\quad - \frac{1}{2} (\text{vec}[(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}])' \cdot \text{vec}(\mathbf{I}_N \otimes d\mathbf{R}) \\
&\quad + \frac{1}{2} \text{tr}[(\mathbf{I}_N \otimes \mathbf{R}') \mathbf{W}(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}] \cdot d\lambda \\
&\quad + \frac{1}{2} \text{tr}[(\mathbf{I}_N \otimes \mathbf{R}')(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{W}'(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}] \cdot d\lambda,
\end{aligned}$$

where \mathbb{D}_{mT} is a duplication matrix (such as that $\mathbb{D}_k \text{vech}(\mathbf{X}) = \text{vec}(\mathbf{X})$ for any $k \times k$ matrix \mathbf{X}), \mathbb{K}_{sq} is a commutation matrix (such that $\mathbb{K}_{sq} \text{vec}(\mathbf{X}) = \text{vec}(\mathbf{X}')$ for any $s \times q$ matrix \mathbf{X}). Hence

$$\begin{aligned}
dL_N &= -\frac{1}{2} \text{vec}(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1}) \mathbf{H} \mathbb{D}_{mT} d\text{vech} \boldsymbol{\Sigma} - \text{tr}[(\mathbf{I}_{mNT} - \lambda \mathbf{W})^{-1} \mathbf{W}] \cdot d\lambda \\
&\quad + \frac{1}{2} (\text{vec}[(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}(\mathbf{I}_N \otimes \mathbf{R}')(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})])' \mathbf{H} \mathbb{D}_{mT} \cdot \text{vec}(\mathbf{I}_N \otimes d\boldsymbol{\Sigma}) \\
&\quad - \frac{1}{2} (\text{vec}[\mathbf{S}(\mathbf{I}_N \otimes \mathbf{R}')(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')])' \mathbf{H} \mathbb{K}_{mT, mT} \cdot d\text{vec} \mathbf{R} \\
&\quad - \frac{1}{2} (\text{vec}[(\mathbf{I}_{mNT} - \lambda \mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}])' \mathbf{H} \cdot d\text{vec} \mathbf{R} \\
&\quad + \text{tr}[(\mathbf{I}_N \otimes \mathbf{R}') \mathbf{W}(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda \mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) \mathbf{S}] \cdot d\lambda,
\end{aligned}$$

where the matrix of constants \mathbf{H} is given by

$$\mathbf{H} = [(\mathbf{I}_N \otimes \mathbb{K}_{mT, N})(\text{vec} \mathbf{I}_N) \otimes \mathbf{I}_{mT}] \otimes \mathbf{I}_{mT}. \tag{A.2}$$

The differential of the inverse of the variance covariance matrix under condition (IOR) is

$$\begin{aligned}
d\text{vech}\boldsymbol{\Sigma} &= \text{vech}[(\mathbf{A}_1 \otimes d\boldsymbol{\Psi}) + (\mathbf{A}_2 \otimes d\boldsymbol{\Omega})] \\
&= \mathbb{D}_{mT}^{-1}(\mathbf{I}_T \otimes \mathbb{K}_{m,T} \otimes \mathbf{I}_m)(\text{vec}\mathbf{A}_1 \otimes \mathbf{I}_{m^2})\mathbb{D}_m d\text{vech}\boldsymbol{\Psi} + \\
&\quad + \mathbb{D}_{mT}^{-1}(\mathbf{I}_T \otimes \mathbb{K}_{m,T} \otimes \mathbf{I}_m)(\text{vec}\mathbf{A}_2 \otimes \mathbf{I}_{m^2})\mathbb{D}_m d\text{vech}\boldsymbol{\Omega} \\
&= \mathbb{D}_{mT}^{-1}\mathbf{B}_1\mathbb{D}_m d\text{vech}\boldsymbol{\Psi} + \mathbb{D}_{mT}^{-1}\mathbf{B}_2\mathbb{D}_m d\text{vech}\boldsymbol{\Omega}
\end{aligned} \tag{A.3}$$

and the differential of the matrix \mathbf{R} containing the slope coefficients is

$$\begin{aligned}
d\text{vec}\mathbf{R} &= \text{vec}(\mathbf{A}_3 \otimes d\boldsymbol{\Phi}) \\
&= (\mathbf{I}_T \otimes \mathbb{K}_{m,T} \otimes \mathbf{I}_m)(\text{vec}\mathbf{A}_3 \otimes \mathbf{I}_{m^2})d\text{vec}\boldsymbol{\Phi} \\
&= \mathbf{B}_3 d\text{vec}\boldsymbol{\Phi}
\end{aligned} \tag{A.4}$$

with \mathbf{A}_1 , \mathbf{A}_2 , \mathbf{A}_3 , \mathbf{B}_1 , \mathbf{B}_2 and \mathbf{B}_3 being matrices of constants reflecting the structure of $\boldsymbol{\Sigma}^{-1}$ and \mathbf{R} .

In particular,

$$\mathbf{A}_1 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & 0 \end{pmatrix} \tag{A.5}$$

$$\mathbf{A}_2 = \mathbf{I}_T - \mathbf{A}_1 \tag{A.6}$$

and

$$\mathbf{A}_3 = \begin{pmatrix} 0 & & \cdots & 0 \\ -1 & 0 & & \vdots \\ 0 & -1 & \ddots & \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -1 & 0 \end{pmatrix} \tag{A.7}$$

Defining

$$\begin{aligned}
\mathbf{M}_1 &= -(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1}) + [(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda\mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R})\mathbf{S}(\mathbf{I}_N \otimes \mathbf{R}')(\mathbf{I}_{mNT} - \lambda\mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})] \\
\mathbf{M}_2 &= -(\mathbf{I}_{mNT} - \lambda\mathbf{W})(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda\mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R})\mathbf{S}, \\
\mathbf{M}_3 &= (\mathbf{I}_N \otimes \mathbf{R}')\mathbf{W}(\mathbf{I}_N \otimes \boldsymbol{\Sigma}^{-1})(\mathbf{I}_{mNT} - \lambda\mathbf{W}')(\mathbf{I}_N \otimes \mathbf{R}) - (\mathbf{I}_{mNT} - \lambda\mathbf{W})^{-1}\mathbf{W},
\end{aligned}$$

we can write the Jacobian of $L_N(\boldsymbol{\theta})$ in a partitioned form as

$$DL_N(\boldsymbol{\theta}) = \frac{1}{2} \begin{bmatrix} [\text{vec}(\mathbf{M}_1)]' \mathbf{H}\mathbf{B}_1 \mathbb{D}_m : \\ [\text{vec}(\mathbf{M}_1)]' \mathbf{H}\mathbf{B}_2 \mathbb{D}_m : \\ [\text{vec}(\mathbf{M}_2)]' \mathbf{H}(\mathbf{I} + \mathbb{K}_{mT,mT}) \mathbf{B}_3 : \\ 2\text{tr}(\mathbf{M}_3) \end{bmatrix}, \tag{A.9}$$

where $:$ denotes horizontal stacking.

Under the (IOR) condition the differential of the variance covariance matrix of the initial observations becomes

$$\begin{aligned}
d\text{vech}\Psi &= d\text{vech}\Omega_\varepsilon \\
&+ \mathbb{D}_m^{-1} [(\mathbf{I}_m - \Phi) \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \mathbb{D}_m d\text{vech}\Omega_\varepsilon \\
&- \mathbb{D}_m^{-1} [d\Phi \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \text{vec}\Omega_\varepsilon \\
&- \mathbb{D}_m^{-1} [(\mathbf{I}_m - \Phi) \otimes d\Phi] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \text{vec}\Omega_\varepsilon \\
&- \mathbb{D}_m^{-1} [(\mathbf{I}_m - \Phi) \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \\
&\cdot [(d\Phi \otimes \Phi) + (\Phi \otimes d\Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \text{vec}\Omega_\varepsilon \\
\\
&= d\text{vech}\Omega_\varepsilon + \mathbb{D}_m^{-1} [(\mathbf{I}_m - \Phi) \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \mathbb{D}_m d\text{vech}\Omega_\varepsilon \\
&- \left[(\text{vec}\Omega_\varepsilon)' [\mathbf{I}_{m^2} - (\Phi' \otimes \Phi')]^{-1} \otimes \mathbb{D}_m^{-1} \right] \cdot \text{vec} [d\Phi \otimes (\mathbf{I}_m - \Phi)] \\
&- \left[(\text{vec}\Omega_\varepsilon)' [\mathbf{I}_{m^2} - (\Phi' \otimes \Phi')]^{-1} \otimes \mathbb{D}_m^{-1} \right] \cdot \text{vec} [(\mathbf{I}_m - \Phi) \otimes d\Phi] \\
&- \left[(\text{vec}\Omega_\varepsilon)' [\mathbf{I}_{m^2} - (\Phi' \otimes \Phi')]^{-1} \otimes \mathbb{D}_m^{-1} [(\mathbf{I}_m - \Phi) \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \right] \\
&\cdot \text{vec} [(d\Phi \otimes \Phi) + (\Phi \otimes d\Phi)] \\
\\
&= d\text{vech}\Omega_\varepsilon + \mathbb{D}_m^{-1} [(\mathbf{I}_m - \Phi) \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \mathbb{D}_m d\text{vech}\Omega_\varepsilon \\
&- \left[(\text{vec}\Omega_\varepsilon)' [\mathbf{I}_{m^2} - (\Phi' \otimes \Phi')]^{-1} \otimes \mathbb{D}_m^{-1} \right] (\mathbf{I}_m \otimes \mathbb{K}_m \otimes \mathbf{I}_m) [\mathbf{I}_{m^2} \otimes \text{vec}(\mathbf{I}_m - \Phi)] \cdot d\text{vec}\Phi \\
&- \left[(\text{vec}\Omega_\varepsilon)' [\mathbf{I}_{m^2} - (\Phi' \otimes \Phi')]^{-1} \otimes \mathbb{D}_m^{-1} \right] (\mathbf{I}_m \otimes \mathbb{K}_m \otimes \mathbf{I}_m) [\text{vec}(\mathbf{I}_m - \Phi) \otimes \mathbf{I}_{m^2}] d\text{vec}\Phi \\
&- \left[(\text{vec}\Omega_\varepsilon)' [\mathbf{I}_{m^2} - (\Phi' \otimes \Phi')]^{-1} \otimes \mathbb{D}_m^{-1} [(\mathbf{I}_m - \Phi) \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \right] \\
&\cdot (\mathbf{I}_m \otimes \mathbb{K}_m \otimes \mathbf{I}_m) [(\mathbf{I}_{m^2} \otimes \Phi) + (\Phi \otimes \mathbf{I}_{m^2})] \cdot d\text{vec}\Phi
\end{aligned}$$

Jacobian of then becomes

$$DL_N(\vartheta) = \frac{1}{2} \begin{bmatrix} [\text{vec}(\mathbf{M}_1)]' \mathbf{H}(\mathbf{B}_1 \mathbf{C}_1 + \mathbf{B}_2) \mathbb{D}_m : \\ [\text{vec}(\mathbf{M}_1)]' \mathbf{H} \mathbf{B}_1 \mathbf{C}_2 + [\text{vec}(\mathbf{M}_2)]' \mathbf{H}(\mathbf{I} + \mathbb{K}_{mT, mT}) \mathbf{B}_3 : \\ 2\text{tr}(\mathbf{M}_3) \end{bmatrix}, \quad (\text{A.10})$$

where

$$\begin{aligned}
\mathbf{C}_1 &= \mathbf{I}_{m^2} + [(\mathbf{I}_m - \Phi) \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \quad (\text{A.11}) \\
\mathbf{C}_2 &= - \left[(\text{vec}\Omega_\varepsilon)' [\mathbf{I}_{m^2} - (\Phi' \otimes \Phi')]^{-1} \otimes \mathbb{D}_m^{-1} \right] (\mathbf{I}_m \otimes \mathbb{K}_m \otimes \mathbf{I}_m) [\mathbf{I}_{m^2} \otimes \text{vec}(\mathbf{I}_m - \Phi)] \\
&- \left[(\text{vec}\Omega_\varepsilon)' [\mathbf{I}_{m^2} - (\Phi' \otimes \Phi')]^{-1} \otimes \mathbb{D}_m^{-1} \right] (\mathbf{I}_m \otimes \mathbb{K}_m \otimes \mathbf{I}_m) [\text{vec}(\mathbf{I}_m - \Phi) \otimes \mathbf{I}_{m^2}] \\
&- \left[(\text{vec}\Omega_\varepsilon)' [\mathbf{I}_{m^2} - (\Phi' \otimes \Phi')]^{-1} \otimes \mathbb{D}_m^{-1} [(\mathbf{I}_m - \Phi) \otimes (\mathbf{I}_m - \Phi)] [\mathbf{I}_{m^2} - (\Phi \otimes \Phi)]^{-1} \right] \\
&\cdot (\mathbf{I}_m \otimes \mathbb{K}_m \otimes \mathbf{I}_m) [(\mathbf{I}_{m^2} \otimes \Phi) + (\Phi \otimes \mathbf{I}_{m^2})]
\end{aligned}$$

B Appendix - Derivation of the Moment Conditions

Observe first that

$$\begin{aligned}
\mathbf{Q}_0 \tilde{\nu} &= [(\mathbf{I}_T - \frac{1}{T} \mathbf{J}_T) \otimes \mathbf{I}_N] [(\boldsymbol{\nu}_T \otimes \tilde{\boldsymbol{\mu}}) + \tilde{\boldsymbol{\varepsilon}}] \\
&= \mathbf{Q}_0 \tilde{\boldsymbol{\varepsilon}} + [(\mathbf{I}_T - \frac{1}{T} \mathbf{J}_T) \otimes \mathbf{I}_N] (\boldsymbol{\nu}_T \otimes \tilde{\boldsymbol{\mu}}) \\
&= \mathbf{Q}_0 \tilde{\boldsymbol{\varepsilon}} + [(\mathbf{I}_T - \frac{1}{T} \mathbf{J}_T) \boldsymbol{\nu}_T \otimes \mathbf{I}_N \tilde{\boldsymbol{\mu}}] \\
&= \mathbf{Q}_0 \tilde{\boldsymbol{\varepsilon}} + [(\mathbf{I}_T \boldsymbol{\nu}_T - \frac{1}{T} \mathbf{J}_T \boldsymbol{\nu}_T) \otimes \tilde{\boldsymbol{\mu}}] \\
&= \mathbf{Q}_0 \tilde{\boldsymbol{\varepsilon}} + [(\boldsymbol{\nu}_T - \frac{T}{T} \boldsymbol{\nu}_T) \otimes \tilde{\boldsymbol{\mu}}] = \mathbf{Q}_0 \tilde{\boldsymbol{\varepsilon}}.
\end{aligned} \tag{B.1}$$

The first moment condition is based on the following observation:

$$\begin{aligned}
E \tilde{\nu}' \mathbf{Q}_0 \tilde{\nu} &= E \tilde{\nu}' \mathbf{Q}_0' \mathbf{Q}_0 \tilde{\nu} = E \tilde{\boldsymbol{\varepsilon}}' \mathbf{Q}_0 \tilde{\boldsymbol{\varepsilon}} = E \tilde{\boldsymbol{\varepsilon}}' [(\mathbf{I}_T - \frac{1}{T} \mathbf{J}_T) \otimes \mathbf{I}_N] \tilde{\boldsymbol{\varepsilon}} \\
&= E \tilde{\boldsymbol{\varepsilon}}' \tilde{\boldsymbol{\varepsilon}} - \frac{1}{T} E (\tilde{\boldsymbol{\varepsilon}}_1', \dots, \tilde{\boldsymbol{\varepsilon}}_T') (\mathbf{J}_T \otimes \mathbf{I}_N) (\tilde{\boldsymbol{\varepsilon}}_1', \dots, \tilde{\boldsymbol{\varepsilon}}_T')' \\
&= \sum_{t=1}^T E \tilde{\boldsymbol{\varepsilon}}_t' \tilde{\boldsymbol{\varepsilon}}_t - \frac{1}{T} E \left(\sum_{t=1}^T \tilde{\boldsymbol{\varepsilon}}_t', \dots, \sum_{t=1}^T \tilde{\boldsymbol{\varepsilon}}_t' \right) (\tilde{\boldsymbol{\varepsilon}}_1', \dots, \tilde{\boldsymbol{\varepsilon}}_T')' \\
&= T \sum_{i=1}^N E \tilde{\boldsymbol{\varepsilon}}_{it}' \tilde{\boldsymbol{\varepsilon}}_{it} - \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E \tilde{\boldsymbol{\varepsilon}}_t' \tilde{\boldsymbol{\varepsilon}}_s \\
&= NT \cdot \boldsymbol{\Omega}_\varepsilon - \frac{1}{T} E \sum_{t=1}^T \sum_{s=1}^T \tilde{\boldsymbol{\varepsilon}}_t' \tilde{\boldsymbol{\varepsilon}}_s = \left(NT - \frac{TN}{T} \right) \boldsymbol{\Omega}_\varepsilon = N(T-1) \cdot \boldsymbol{\Omega}_\varepsilon
\end{aligned} \tag{B.2}$$

The second moment conditions follows from

$$\begin{aligned}
E \tilde{\nu}' \mathbf{Q}_0 \tilde{\nu} &= E \tilde{\boldsymbol{\varepsilon}}' \mathbf{Q}_0 \tilde{\boldsymbol{\varepsilon}} = E \tilde{\boldsymbol{\varepsilon}}' [(\mathbf{I}_T - \frac{1}{T} \mathbf{J}_T) \otimes \mathbf{I}_N] \tilde{\boldsymbol{\varepsilon}} \\
&= E \tilde{\boldsymbol{\varepsilon}}' \tilde{\boldsymbol{\varepsilon}} - \frac{1}{T} E (\tilde{\boldsymbol{\varepsilon}}_1', \dots, \tilde{\boldsymbol{\varepsilon}}_T') (\mathbf{J}_T \otimes \mathbf{I}_N) (\tilde{\boldsymbol{\varepsilon}}_1', \dots, \tilde{\boldsymbol{\varepsilon}}_T')' \\
&= \sum_{t=1}^T E \tilde{\boldsymbol{\varepsilon}}_t' \tilde{\boldsymbol{\varepsilon}}_t - \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E \tilde{\boldsymbol{\varepsilon}}_t' \tilde{\boldsymbol{\varepsilon}}_s \\
&= \sum_{t=1}^T \sum_{i=1}^N E \tilde{\boldsymbol{\varepsilon}}_{it}' \tilde{\boldsymbol{\varepsilon}}_{it} - \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N E \tilde{\boldsymbol{\varepsilon}}_{it}' \tilde{\boldsymbol{\varepsilon}}_{is} = (T-1) \sum_{i=1}^N E \tilde{\boldsymbol{\varepsilon}}_{it}' \tilde{\boldsymbol{\varepsilon}}_{it} \\
&= (T-1) \sum_{i=1}^N (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN}) E (\boldsymbol{\varepsilon}'_{1t}, \dots, \boldsymbol{\varepsilon}'_{Nt})' (\boldsymbol{\varepsilon}'_{1t}, \dots, \boldsymbol{\varepsilon}'_{Nt}) (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN})' \\
&= (T-1) \sum_{i=1}^N (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN}) (\mathbf{I}_N \otimes \boldsymbol{\Omega}_\varepsilon) (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN})' \\
&= (T-1) \sum_{i=1}^N \sum_{j=1}^N \mathbf{w}_{ij} \boldsymbol{\Omega}_\varepsilon \mathbf{w}'_{ij}.
\end{aligned} \tag{B.3}$$

The third moment condition is based on

$$\begin{aligned}
E\tilde{\nu}'\mathbf{Q}_0\tilde{\nu} &= E\tilde{\varepsilon}'\mathbf{Q}_0\tilde{\varepsilon} = E\tilde{\varepsilon}'[(\mathbf{I}_T - \frac{1}{T}\mathbf{J}_T) \otimes \mathbf{I}_N]\tilde{\varepsilon} \\
&= E\tilde{\varepsilon}'\tilde{\varepsilon} - \frac{1}{T}E\left(\tilde{\varepsilon}'_1, \dots, \tilde{\varepsilon}'_T\right)(\mathbf{J}_T \otimes \mathbf{I}_N)\left(\tilde{\varepsilon}'_1, \dots, \tilde{\varepsilon}'_T\right)' \\
&= \sum_{t=1}^T E\tilde{\varepsilon}'_t\tilde{\varepsilon}_t - \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T E\tilde{\varepsilon}'_t\tilde{\varepsilon}_s \\
&= \sum_{t=1}^T\sum_{i=1}^N E\bar{\varepsilon}_{it}\varepsilon'_{it} - \frac{1}{T}\sum_{t=1}^T\sum_{s=1}^T\sum_{i=1}^N E\bar{\varepsilon}_{it}\varepsilon'_{is} = (T-1)\sum_{i=1}^N E\bar{\varepsilon}_{it}\varepsilon'_{it} \\
&= (T-1)\sum_{i=1}^N(\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN})E(\varepsilon'_{1t}, \dots, \varepsilon'_{Nt})'\varepsilon'_{it} \\
&= (T-1)\sum_{i=1}^N(\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN})(\mathbf{e}_{i,N} \otimes \boldsymbol{\Omega}_\varepsilon) = (T-1)\sum_{i=1}^N \mathbf{w}_{ii}\boldsymbol{\Omega}_\varepsilon,
\end{aligned} \tag{B.4}$$

where we use $\mathbf{e}_{i,N}$ to denote an $N \times 1$ vector of zeros with an entry of one on the i -th position.

To derive the next set of moment conditions involving \mathbf{Q}_1 , we note that

$$\begin{aligned}
\mathbf{Q}_1\tilde{\nu} &= \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right)[(\boldsymbol{\nu}_T \otimes \tilde{\boldsymbol{\mu}}) + \tilde{\varepsilon}] \\
&= \mathbf{Q}_1\tilde{\varepsilon} + \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right)(\boldsymbol{\nu}_T \otimes \tilde{\boldsymbol{\mu}}) \\
&= \mathbf{Q}_1\tilde{\varepsilon} + \left(\frac{1}{T}\mathbf{J}_T\boldsymbol{\nu}_T \otimes \tilde{\boldsymbol{\mu}}\right) \\
&= \mathbf{Q}_1\tilde{\varepsilon} + (\boldsymbol{\nu}_T \otimes \mathbf{I}_N)\tilde{\boldsymbol{\mu}}.
\end{aligned} \tag{B.5}$$

Furthermore, denoting

$$\tilde{\mathbf{W}} = [(\mathbf{w}_{11}, \dots, \mathbf{w}_{1N}), \dots, (\mathbf{w}_{1N}, \dots, \mathbf{w}_{NN})]', \tag{B.6}$$

we have that

$$\begin{aligned}
\mathbf{Q}_1\tilde{\nu} &= \mathbf{Q}_1\tilde{\varepsilon} + \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right)\begin{bmatrix} (\mathbf{I}_N \otimes \boldsymbol{\mu}'_1), \dots, (\mathbf{I}_N \otimes \boldsymbol{\mu}'_N) \\ \vdots \\ (\mathbf{I}_N \otimes \boldsymbol{\mu}'_1), \dots, (\mathbf{I}_N \otimes \boldsymbol{\mu}'_N) \end{bmatrix}\tilde{\mathbf{W}} \\
&= \mathbf{Q}_1\tilde{\varepsilon} + \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right)\left(\boldsymbol{\nu}_T \otimes [(\mathbf{I}_N \otimes \boldsymbol{\mu}'_1), \dots, (\mathbf{I}_N \otimes \boldsymbol{\mu}'_N)]\tilde{\mathbf{W}}\right) \\
&= \mathbf{Q}_1\tilde{\varepsilon} + \left(\frac{1}{T}\mathbf{J}_T\boldsymbol{\nu}_T \otimes [(\mathbf{I}_N \otimes \boldsymbol{\mu}'_1), \dots, (\mathbf{I}_N \otimes \boldsymbol{\mu}'_N)]\tilde{\mathbf{W}}\right) \\
&= \mathbf{Q}_1\tilde{\varepsilon} + \left(\boldsymbol{\nu}_T \otimes [(\mathbf{I}_N \otimes \boldsymbol{\mu}'_1), \dots, (\mathbf{I}_N \otimes \boldsymbol{\mu}'_N)]\tilde{\mathbf{W}}\right) \\
&= \mathbf{Q}_1\tilde{\varepsilon} + \left(\boldsymbol{\nu}_T \otimes \tilde{\boldsymbol{\mu}}\right).
\end{aligned} \tag{B.7}$$

Thus we have

$$\begin{aligned}
E\tilde{\nu}'\mathbf{Q}_1\tilde{\nu} &= E[\tilde{\varepsilon} + (\boldsymbol{\nu}_T \otimes \mathbf{I}_N)\tilde{\boldsymbol{\mu}}]' \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right) [\tilde{\varepsilon} + (\boldsymbol{\nu}_T \otimes \mathbf{I}_N)\tilde{\boldsymbol{\mu}}] \\
&= E\tilde{\boldsymbol{\varepsilon}}' \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right) \tilde{\boldsymbol{\varepsilon}} + E\tilde{\boldsymbol{\mu}}' (\boldsymbol{\nu}_T \otimes \mathbf{I}_N) \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right) (\boldsymbol{\nu}_T \otimes \mathbf{I}_N) \tilde{\boldsymbol{\mu}} \\
&= \frac{1}{T}E(\tilde{\boldsymbol{\varepsilon}}'_1, \dots, \tilde{\boldsymbol{\varepsilon}}'_T) (\mathbf{J}_T \otimes \mathbf{I}_N) (\tilde{\boldsymbol{\varepsilon}}'_1, \dots, \tilde{\boldsymbol{\varepsilon}}'_T)' + E\tilde{\boldsymbol{\mu}}' \left(\frac{1}{T}\boldsymbol{\nu}'_T \mathbf{J}_T \boldsymbol{\nu}_T \otimes \mathbf{I}_N\right) \tilde{\boldsymbol{\mu}} \\
&= \frac{1}{T}E\left(\sum_{t=1}^T \tilde{\boldsymbol{\varepsilon}}'_t, \dots, \sum_{t=1}^T \tilde{\boldsymbol{\varepsilon}}'_t\right) (\tilde{\boldsymbol{\varepsilon}}'_1, \dots, \tilde{\boldsymbol{\varepsilon}}'_T)' + T \cdot E\tilde{\boldsymbol{\mu}}'\tilde{\boldsymbol{\mu}} \\
&= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E\tilde{\boldsymbol{\varepsilon}}'_t \tilde{\boldsymbol{\varepsilon}}'_s + T \cdot E\tilde{\boldsymbol{\mu}}'\tilde{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N E\tilde{\boldsymbol{\varepsilon}}_{it} \tilde{\boldsymbol{\varepsilon}}'_{is} + T \sum_{i=1}^N E\tilde{\boldsymbol{\mu}}_i \tilde{\boldsymbol{\mu}}'_i \\
&= \left(\frac{TN}{T}\right) \boldsymbol{\Omega}_\varepsilon + TN \cdot \boldsymbol{\Omega}_\mu = N \cdot \boldsymbol{\Omega}_1.
\end{aligned} \tag{B.8}$$

Furthermore,

$$\begin{aligned}
E\bar{\nu}'\mathbf{Q}_1\bar{\nu} &= E\left[\bar{\boldsymbol{\varepsilon}} + (\boldsymbol{\nu}_T \otimes \mathbf{I}_N)\bar{\boldsymbol{\mu}}\right]' \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right) \left[\bar{\boldsymbol{\varepsilon}} + (\boldsymbol{\nu}_T \otimes \mathbf{I}_N)\bar{\boldsymbol{\mu}}\right] \\
&= E\bar{\boldsymbol{\varepsilon}}' \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right) \bar{\boldsymbol{\varepsilon}} + E\bar{\boldsymbol{\mu}}' (\boldsymbol{\nu}_T \otimes \mathbf{I}_N) \left(\frac{1}{T}\mathbf{J}_T \otimes \mathbf{I}_N\right) (\boldsymbol{\nu}_T \otimes \mathbf{I}_N) \bar{\boldsymbol{\mu}} \\
&= \frac{1}{T}E(\bar{\boldsymbol{\varepsilon}}'_1, \dots, \bar{\boldsymbol{\varepsilon}}'_T) (\mathbf{J}_T \otimes \mathbf{I}_N) (\bar{\boldsymbol{\varepsilon}}'_1, \dots, \bar{\boldsymbol{\varepsilon}}'_T)' + E\bar{\boldsymbol{\mu}}' \left(\frac{1}{T}\boldsymbol{\nu}'_T \mathbf{J}_T \boldsymbol{\nu}_T \otimes \mathbf{I}_N\right) \bar{\boldsymbol{\mu}} \\
&= \frac{1}{T}E\left(\sum_{t=1}^T \bar{\boldsymbol{\varepsilon}}'_t, \dots, \sum_{t=1}^T \bar{\boldsymbol{\varepsilon}}'_t\right) (\bar{\boldsymbol{\varepsilon}}'_1, \dots, \bar{\boldsymbol{\varepsilon}}'_T)' + T \cdot E\bar{\boldsymbol{\mu}}'\bar{\boldsymbol{\mu}} \\
&= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E\bar{\boldsymbol{\varepsilon}}'_t \bar{\boldsymbol{\varepsilon}}'_s + T \cdot E\bar{\boldsymbol{\mu}}'\bar{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N E\bar{\boldsymbol{\varepsilon}}_{it} \bar{\boldsymbol{\varepsilon}}'_{is} + T \sum_{i=1}^N E\bar{\boldsymbol{\mu}}_i \bar{\boldsymbol{\mu}}'_i \\
&= \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN}) E(\boldsymbol{\varepsilon}'_{1t}, \dots, \boldsymbol{\varepsilon}'_{Nt})' (\boldsymbol{\varepsilon}'_{1t}, \dots, \boldsymbol{\varepsilon}'_{Nt}) (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN})' \\
&\quad + T \sum_{i=1}^N (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN}) E(\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_N)' (\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_N) (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN})' \\
&= \sum_{i=1}^N (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN}) (\mathbf{I}_N \otimes \boldsymbol{\Omega}_\varepsilon) (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN})' \\
&\quad + T \sum_{i=1}^N (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN}) (\mathbf{I}_N \otimes \boldsymbol{\Omega}_\mu) (\mathbf{w}_{i1}, \dots, \mathbf{w}_{iN})' \\
&= \sum_{i=1}^N \sum_{j=1}^N \mathbf{w}_{ij} \boldsymbol{\Omega}_\varepsilon \mathbf{w}'_{ij} + T \sum_{i=1}^N \sum_{j=1}^N \mathbf{w}_{ij} \boldsymbol{\Omega}_\mu \mathbf{w}'_{ij} = \sum_{i=1}^N \sum_{j=1}^N \mathbf{w}_{ij} \boldsymbol{\Omega}_1 \mathbf{w}'_{ij},
\end{aligned} \tag{B.9}$$

and

$$\begin{aligned}
E\tilde{\boldsymbol{\nu}}' \mathbf{Q}_1 \tilde{\boldsymbol{\nu}} &= E \left[\tilde{\boldsymbol{\varepsilon}} + (\boldsymbol{\iota}_T \otimes \mathbf{I}_N) \tilde{\boldsymbol{\mu}} \right]' \left(\frac{1}{T} \mathbf{J}_T \otimes \mathbf{I}_N \right) \left[\tilde{\boldsymbol{\varepsilon}} + (\boldsymbol{\iota}_T \otimes \mathbf{I}_N) \tilde{\boldsymbol{\mu}} \right] & (B.10) \\
&= E\tilde{\boldsymbol{\varepsilon}}' \left(\frac{1}{T} \mathbf{J}_T \otimes \mathbf{I}_N \right) \tilde{\boldsymbol{\varepsilon}} + E\tilde{\boldsymbol{\mu}}' (\boldsymbol{\iota}'_T \otimes \mathbf{I}_N) \left(\frac{1}{T} \mathbf{J}_T \otimes \mathbf{I}_N \right) (\boldsymbol{\iota}_T \otimes \mathbf{I}_N) \tilde{\boldsymbol{\mu}} \\
&= \frac{1}{T} E \left(\tilde{\boldsymbol{\varepsilon}}'_1, \dots, \tilde{\boldsymbol{\varepsilon}}'_T \right) (\mathbf{J}_T \otimes \mathbf{I}_N) \left(\tilde{\boldsymbol{\varepsilon}}'_1, \dots, \tilde{\boldsymbol{\varepsilon}}'_T \right)' + E\tilde{\boldsymbol{\mu}}' \left(\frac{1}{T} \boldsymbol{\iota}'_T \mathbf{J}_T \boldsymbol{\iota}_T \otimes \mathbf{I}_N \right) \tilde{\boldsymbol{\mu}} \\
&= \frac{1}{T} E \left(\sum_{t=1}^T \tilde{\boldsymbol{\varepsilon}}'_t, \dots, \sum_{t=1}^T \tilde{\boldsymbol{\varepsilon}}'_t \right) \left(\tilde{\boldsymbol{\varepsilon}}'_1, \dots, \tilde{\boldsymbol{\varepsilon}}'_T \right)' + T \cdot E\tilde{\boldsymbol{\mu}}' \tilde{\boldsymbol{\mu}} \\
&= \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T E\tilde{\boldsymbol{\varepsilon}}'_t \tilde{\boldsymbol{\varepsilon}}'_s + T \cdot E\tilde{\boldsymbol{\mu}}' \tilde{\boldsymbol{\mu}} = \frac{1}{T} \sum_{t=1}^T \sum_{s=1}^T \sum_{i=1}^N E\tilde{\boldsymbol{\varepsilon}}_{it} \tilde{\boldsymbol{\varepsilon}}'_{is} + T \sum_{i=1}^N E\tilde{\boldsymbol{\mu}}_i \tilde{\boldsymbol{\mu}}'_i \\
&= \frac{1}{T} \sum_{t=1}^T \sum_{i=1}^N \sum_{i=1}^N E(\boldsymbol{\varepsilon}'_{1t}, \dots, \boldsymbol{\varepsilon}'_{Nt})' \boldsymbol{\varepsilon}'_{it} \\
&\quad + \sum_{i=1}^N \sum_{i=1}^N E(\boldsymbol{\mu}'_1, \dots, \boldsymbol{\mu}'_N)' \tilde{\boldsymbol{\mu}}'_i \\
&= \sum_{i=1}^N \mathbf{w}_{ii} \boldsymbol{\Omega}_\varepsilon + T \sum_{i=1}^N \mathbf{w}_{ii} \boldsymbol{\Omega}_\mu = \sum_{i=1}^N \mathbf{w}_{ii} \boldsymbol{\Omega}_1.
\end{aligned}$$

References

- [1] Ahn, S.C. and P. Schmidt (1995): Efficient Estimation of Models for Dynamic Panel Data, *Journal of Econometrics*, 68, 5-27.
- [2] Anselin, L. (1988): *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Boston.
- [3] Arellano, M. and S.R. Bond (1991): Some Tests of Specification for Panel Data: Monte Carlo Evidence and Application to Employment Equations, *Review of Economic Studies*, 58, 277-297.
- [4] Arellano, M. and O. Bover (1995): Another Look at the Instrumental Variable Estimation if Error-Component Models, *Journal of Econometrics*, 68, 28-51.
- [5] Binder, Michael, Cheng Hsiao and M. Hashem Pesaran (2005): Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration, *Econometric Theory* 21(4), 795-837.
- [6] Binder, Michael, Cheng Hsiao, Jan Mutl and M. Hashem Pesaran (2002): Computational Issues in Estimation of Higher-Order Panel Vector Autoregressions, *mimeo*.
- [7] Chang, Yoosoon (2002): Nonlinear IV Unit Root Tests in Panels with Cross-Sectional Dependency, *Journal of Econometrics* 110, 261-292.
- [8] Chen, Xiaohong and Timothy Conley (2000): A New Semiparametric Spatial Model for Panel Time Series, *Journal of Econometrics* 105, 59-83.
- [9] Cliff, A. and J. Ord (1973): *Spatial Autocorrelation*, Pion, London.
- [10] Cliff, A. and J. Ord (1981): *Spatial Processes, Models and Applications*, Pion, London.
- [11] Hsiao, C., M. H. Pesaran and A. K. Tahmiscioglu (2002): Maximum Likelihood Estimation of Fixed Effects Dynamic Panel Data Models Covering Short Time Periods, *Journal of Econometrics*, 109, 107-150.
- [12] Kapoor M., H. Kelejian and I.P. Prucha (2007): Panel Data Models with Spatially Correlated Error Components, *Journal of Econometrics* 140, 97-130.
- [13] Kelejian, Harry H. and Ingmar 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, Harry H. and Ingmar R. Prucha (1999): A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, *International Economic Review*, 40, 509-533.
- [15] Kelejian, Harry H. and Ingmar R. Prucha (2004): Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations, *Journal of Econometrics* 118, 27-50.

- [16] Kelejian, Harry H. and Ingmar R. Prucha (2001): On the Asymptotic Distribution of the Moran I Test Statistic with Applications, *Journal of Econometrics*, 104, 219-257..
- [17] O'Connell, paul G. J. (1998): The Overvaluation of Purchasing Power Parity, *Journal of International Economics*, 44, 1-19.
- [18] Pinkse, Joris (1999): Asymptotic Properties of Moran and Related Tests and Testing for Spatial Correlation in Probit Models, *mimeo*.
- [19] Whittle, P. (1954): On Stationary Processes in the Plane, *Biometrika*, 41, 434-449.