Journal of Statistical Research *ISSN 0256 - 422 X*
2012, Vol. 46, No. 1, pp. 1-18
Bangladesh

MOMENT RESTRICTIONS FOR OPTIMUM GMM ESTIMATORS UNDER SPATIAL SIMULTANEOUS SYSTEMS

P. E. SIMLAI
College of Business and Public Administration
University of North Dakota, Grand Forks, ND 58202, USA
Email: psimlai@und.edu

SUMMARY

It is widely known that for large samples maximum-likelihood based estimators are not easy to implement for spatial models unless weights matrix satisfy certain basic conditions. As an alternative recently in a series of paper Kelejian and Prucha (1999, 2004) proposed a computationally feasible three step procedure for spatial models involving both lagged dependent variable and spatially correlated disturbance term. The idea of this paper is to use their set up for spatial simultaneous system, and construct a GMM type estimator based on spatial first difference. The over-identification of the moment equation comes to the picture by considering first two moments of a possibly heteroskedastic disturbance. Following Chamberlains (1987) idea, a popular issue of optimum GMM based on conditional spatial moment restrictions and asymptotic efficiency has been discussed.

Keywords: spatial dependence; first difference; conditional spatial moment; optimum GMM; asymptotic efficiency

1 Introduction

Throughout the paper we are interested in spatial moment restrictions of the form

$$E(\psi(z_n, \theta_0)) = 0, \tag{1.1}$$

where $\{z_n\}$'s are generated from a spatial simultaneous system and θ_0 constitutes a unique solution to equations $E(\psi(z_n, \theta)) = 0$. We keep the possibility of $dim(\psi) \geq dim(\theta)$ and want to determine the lower bound on asymptotic variance for a consistent estimator of θ_0. In order to estimate the model parameters, we utilize the generalized method of moments (GMM) methodology pioneered by Hansen et al. (1996).

GMM estimates the parameters by making the sample averages in moment conditions as close to each other as possible and gives a statistical test of the hypothesis that the underlying population means are in fact zero. The idea is to exploit the sample mean of (1.1) given by $g(\theta) = 1/T\{\sum_{t=1}^{T} \psi(z_t, \theta)\}$ where the parameter vector is θ in a sample of size T, and estimate θ by minimizing a quadratic form of the sample mean of moments $argmin_\theta\{g(\theta)'Wg(\theta)\}$ for some arbitrary weight matrix W. In the process it provides a consistent, asymptotically normal, and asymptotically efficient estimate of the parameter. "Efficient" simply means that it has the smallest variance-covariance matrix among all estimators that set different linear combinations of $g(\theta)$ to zero or all choices of weighting matrix W.

In this paper, we are interested in a consistent spatial GMM estimator $\widehat{\theta}_n$ in the context of spatial simultaneous system. We want to know if it is possible to determine $b_n \to \infty$ such that $b_n(\widehat{\theta}_n - \theta) \to N$, where N is a non-degenerate distribution. There are two reasons why we are interested in the moment conditions based on cross-sectional dependence. First, it helps us to illustrate the importance of one of the popular econometric debate about choosing correct orthogonality condition for making GMM based inference. Second, even when we find a set of orthogonality condition based on economic theory, we need a guarantee that it will produce most efficient estimate. In the spatial econometrics literature both of them are still partially unanswered.

2 Where the paper is nested in the current literature

It is widely known that in any regression model the choice of moment depends on the incorporation of information that comes through auxiliary distribution assumptions. In the cross-sectional data analysis there is no exception. Till now we see four avenues on this topic. In one of the first approach, Kelejian and Prucha (1998, 1999) proposed generalized (GM) spatial 2SLS and 3SLS based on exactly identified moments condition. In another approach, Conely (1999) showed consistency and asymptotic normality of over-identified GMM estimators under a set of general strong mixing conditions. There is a third approach by Lee (2001), which is very much similar to Kelejian and Prucha except that Lee's choice of moment condition is more general and includes the case of over identification. Recently for a dynamic spatial Probit model, Pinske, Shen and Slade (2003) made weaker than strong mixing conditions and proved the optimal asymptotic properties of one-step GMM estimators.

In this paper we use Kelejian and Prucha set up for spatial simultaneous system and construct a GMM type estimator based on spatial first difference. It is important to note that there is subtle difference between our GMM approach and the existing GM method Kelejian and Prucha. In the GM method the number of moment conditions is exactly equal to the number of parameters, therefore the model is exactly identified. In our GMM approach however, the number of moment conditions are greater than the number of parameters. Therefore, unlike the GM method, we entertain the possibility of over-identified system.

Intuitively, if there are as many moments as parameters, we set each moment to zero; when there are fewer parameters than moments, (1.1) captures the natural idea that we will set some moments, or some linear combinations of moments, to zero in order to estimate the parameters. Our general GMM procedure therefore allows us to pick arbitrary linear combinations of the moments to set to zero in parameter estimation. Following Hansen et al. (1996), one can show that the expression for asymptotic variance of a GMM estimator satisfying (1.1) becomes

$$\left(E\left[m_\theta\right]\right)^{-1} E\left[mm'\right]\left(E\left[m'_\theta\right]\right)^{-1}, \tag{2.1}$$

where $m_\theta = (E[\frac{\partial}{\partial\theta}\psi(z_n, \theta_0)])' A(\frac{\partial}{\partial\theta}\psi(z_n, \theta_0))$, $m = (E[\frac{\partial}{\partial\theta}\psi(z_n, \theta_0)])' A\psi(z_n, \theta_0)$ and $A = a'a$ is a positive definite matrix. Chamberlain (1987) argued that any distribution can be approximated arbitrarily well by a multinomial distribution and expression (2.1) will be the same asymptotic variance that can be attained by a semi-parametric model. He derived the bound on asymptotic efficiency using conditional moment restrictions $E[\rho(y, x, \theta_0)|x] = 0$, so we can construct a ψ function based upon $h_j(x)\rho(y, x, \theta_0)$ where $h_j(x) = 1$ if $x = \tau_j$ and 0 otherwise, and x can take finite sets of values $\tau_1, ..., \tau_l$. He showed that asymptotic variance bound of the estimator using optimal instruments becomes $\Lambda^* = \left(E\left[D_0(x)\Sigma_0^{-1}(x)D_0(x)\right]\right)^{-1}$, where $D_0(x) = E(\partial\rho(y, x, \theta_0)/\partial\theta|x)$ and $\Sigma_0(x) = E(\rho(y, x, \theta_0)\rho(y, x, \theta_0)'|x)$. Even though there exists an infinite number of conditional moment restrictions, this is the best one can get for information matrix bound on the asymptotic variance of a consistent estimator of θ_0.

In this paper we are addressing the same question but in a spatial setting. What we are trying to find is a finite set of optimal orthogonality condition for first order spatial simultaneous equations model to obtain the asymptotic Cramér-Rao information bound (inverse information matrix) for unknown parameter θ. They are optimal because the asymptotic variance of consistent estimator of θ can not be reduced adding extra ψ's. In other words, no efficiency gain is possible by exploiting same model information because we will attain semi-parametric efficiency bound in the sense of Chamberlain (1987).

We will also extend Chamberlain's (1987) argument one step further and show that in spatial setting this set of moment condition, indeed, satisfy the recently popular information theoretic approaches, namely Empirical Likelihood (EL) and Kullback-Leibler Information Criterion (KLIC) based estimation. For EL and KLIC will follow the discussion of Qin and Lawless (1993) and Imbens, Spady and Johnson (1998) to show their implementability for spatial models. Overall in our presentation we concentrate more on the distributional property of spatial simultaneous system's GMM estimator. Even though consistency is a desirable property, it is not useful in itself. Especially, being asymptotic in nature, for any finite sample size it is not applicable. This motivates us to concentrate on the distribution of $\widehat{\theta}_n$. From the perspective of spatial econometric inference this is much stronger and useful than mere consistency of $\widehat{\theta}_n$.

In the next section we first describe our specification of the model and proper choice of spatial weights matrix. Then we briefly outline the basic steps of Kelejian-Prucha GM estimators. In section 4, we investigate the asymptotic properties of over-identified spatial simultaneous system based on first difference. This is followed by a implementation of optimum GMM based on conditional moments. The technical details and proof of theoretical results are given in appendix.

3 Specification of Model

We start with a simultaneous system of equation which has both spatially lagged dependent variable as well as error term. This model has been discussed by Kelejian and Prucha (2004) in their GM approach, so we borrow heavily their notation. The basic model is

$$\begin{aligned} Y_n &= Y_n B + X_n C + \overline{Y}_n \Lambda + U_n \\ U_n &= \overline{U}_n R + E_n, \end{aligned} \tag{3.1}$$

where $Y_n = (y_{1,n}, ..., y_{m,n})$, $X_n = (x_{1,n}, ..., x_{k,n})$,$U_n = (u_{1,n}, ..., u_{m,n})$, $\overline{y}_{j,n} = W_n y_{j,n}$, $j = 1, ..., m$, $\overline{Y}_n = (\overline{y}_{1,n}, ..., \overline{y}_{m,n})$, $\overline{U}_n = (\overline{u}_{1,n}, ..., \overline{u}_{m,n})$, $\overline{u}_{j,n} = W_n u_{j,n}, E_n = (\epsilon_{1,n}, ..., \epsilon_{m,n})$, $R = diag_{j=1}^{m}(\rho_j)$. Note that, $y_{j,n}, u_{j,n}$ and $\epsilon_{j,n}$ are all $n \times 1$ vectors in the jth equation, $x_{l,n}$ is $n \times 1$ vector on lth exogenous variable and the ith element of $\overline{y}_{j,n}$ is $\overline{y}_{ij,n} = \sum_{n}^{r=1} w_{ir,n} y_{rj,n}$. From this given set up we derive the next set of equation

$$\begin{aligned} y_n &= B_n^* y_n + C_n^* x_n + u_n \\ u_n &= R_n^* u_n + \epsilon_n, \end{aligned} \tag{3.2}$$

where $y_n = vec(Y_n), x_n = vec(X_n), u_n = vec(U_n), \epsilon_n = vec(E_n)$, $B_n^* = [(B' \otimes I_n) + (\Lambda \otimes W_n)]$, $C_n^* = (C' \otimes I_n)$, $R_n^* = (R \otimes W_n) = diag_{j=1}^{n}(\rho_j W_j)$.

Finally, for $j = 1, \dots, m$, we can express the entire system by the following common form

$$\begin{aligned} y_{j,n} &= z_{j,n} \delta_j + u_{j,n} \\ u_{j,n} &= \rho_j W_n u_{j,n} + \epsilon_{j,n}, \end{aligned} \tag{3.3}$$

where $z_{j,n} = (Y_{j,n}, X_{j,n}, \overline{Y}_{j,n})$, $\delta_j = (\beta_j', \gamma_j', \lambda_j')'$. Here ρ_j captures the extent of spatial error dependence that exists in the model. Following the common wisdom of the literature, we can term ρ_j as the spatial lag parameter. Since $I_{mn} - R_n^* = diag_{j=1}^{m}(I_n - \rho_j W_n)$, equation (4) implies the following

$$\begin{aligned} y_n &= (I_{mn} - B_n^*)^{-1}[C_n^* x_n + u_n] \\ u_n &= (I_{mn} - R_n^*)^{-1} \epsilon_n. \end{aligned} \tag{3.4}$$

Based on the above set up we define simultaneous structure accordingly.

Definition 3.1. The reduced form of the spatial simultaneous system (3.1) is given by (3.4).

Denote Ξ_n be the vector space over which W_n is defined and make the following assumptions.

- A1: *The diagonal elements of the spatial weights matrix $W_n = (w_{ij,n}) \in \Xi_n$ are zero and off-diagonal elements are bounded uniformly in absolute value.*
- A2: *The matrices $(I_{mn} - B_n^*)$ and $(I_n - \rho_j W_n)$ are non-singular with $|\rho_j| < 1, j = 1, \dots, m$.*
- A3: *The row and column sums of the matrices $\widetilde{B} = (I_{mn} - B_n^*)^{-1}$ and $\widetilde{W} = (I_n - \rho_j W_n), j = 1, \dots, m$ are bounded in absolute value, i.e., $\sum_{i=1}^{n} \widetilde{b}_{ij} \leq K_b$, $\sum_{j=1}^{n} \widetilde{b}_{ij} \leq K_b$, $\sum_{i=1}^{n} \widetilde{w}_{ij} \leq K_w$ and $\sum_{j=1}^{n} \widetilde{w}_{ij} \leq K_w, n \geq 1, k_b, k_w < \infty, \forall i = 1, \dots, n, j = 1, \dots, m$.*
- A4: *(a)$\epsilon_n = (\Sigma_*' \otimes I_n)\nu_n$, where Σ_*' is nonsingular $m \times m$ matrix. (b) $\{\nu_{ij,n} : i = 1, \dots, n, j = 1, \dots, m\}$ are $iid(0, 1)$ and $E|\nu_{ij,n}|^{2+s} \leq K_\nu < \infty, \forall s > 0$. (c) $\Sigma = \Sigma_*' \Sigma_*$ and diag of Σ are bounded by some constant.*

A5: *The predetermined variables are nonstochastic (exogenous) and for a positive constant* K_x, $|x_{ij,n}| < K_x, \forall i, j$.

Assumption 1 and 2 are standard in terms of spatial weights specification. Similar assumption has been made by Kelejian and Prucha (1999, 2004). In assumption 4, we expressed the error term as generated by some iid random variables which assumes certain basic properties. The second part of A4 are usual sufficient conditions restricting the tails of the error distribution. Later we will modify our set up by assuming a general form conditional error specification. Assumption 3 maintains some restriction on spatial interaction regardless of the sample size. This types of restriction dated back to Anselin and Kelejian (1997) and can be put forward by other type of mathematical arguments like mixing conditions (see, e.g., Anselin,1988, Ch.5 and Conely, 1999). The implication of A5 is that both $\lambda_{max}(X_n' X_n)$ and $\lambda_{min}(X_n' X_n)$ goes to ∞ in an order not faster than the sample size n. In fact the growth rate can be very slow.

We formalize the set up of Kelejian and Prucha (2004) in terms of the following result.

Proposition 3.2: *Under assumptions A1-A4, we have the following distribution for* u_n *and* y_n*:*

$$\begin{aligned} u_n &\sim N(0, \Omega_{u,n}) \\ y_n &\sim N(\mu_y, \Omega_{y,n}), \end{aligned} \tag{3.5}$$

where $\Omega_{u,n} = (I_{mn} - R_n^*)^{-1}(\Sigma \otimes I_n)(I_{mn} - R_n^{*'})^{-1}$, $\mu_y = (I_{mn} - B_n^*)^{-1} C_n^* x_n$ *and* $\Omega_{y,n} = (I_{mn} - B_n^*)^{-1} \Omega_{u,n} (I_{mn} - B_n^{*'})^{-1}$.

3.1 The Choice of Spatial Weights

The crucial point that distinguishes spatial models from time series counterpart is it's dependence structure. Unlike time domain, here the conditional and simultaneously specified models are not the same. So the equivalence of joint-probability and conditional-probability definitions does not hold in general. In practice, for Gaussian data, since the conditionally specified model has a particular simple joint distribution, it turns out to be more natural to consider. Another point of departure is the likelihood function for the joint normal density of the error term. Under spatial framework the likelihood function has an additional Jacobian of the transformation, which implies that OLS will no longer be equal to MLE.

Typically the spatially weights matrix $W_n = (w_{ij,n})$ are a set of non-negative weights representing the "degree of possible interaction" between location j and location i. By convention we always set the diagonal term of the weights matrix as zero and each row sum to one. In the literature there is a debate why spatial weights are symmetric. Usually for any first order spatial model we assume that $(1/\omega_{min}) < \rho < (1/\omega_{max})$, where ω's are characteristics roots of W_n. Kelejian and Robinson (1995) noted that so long as the matrix $(I - \rho W_n)$ is non-singular any first order spatial model is defined and so the restriction may become not necessary.

Anselin and Bera (1998) pointed out that even though we assume W_n symmetric, row standardization makes it asymmetric; so the equality of determinant and product of eigenvalues relation may no longer hold and *even being asymmetric* makes eigenvalues complex. However, row standardization[1] constructed from symmetric contiguity matrix makes all eigenvalues real. The point is that we can show largest eigenvalues are one but for lowest eigenvalue the absolute value can be greater or less than one. So unlike time series, parameter space will be asymmetric around zero. One interesting implication is that for a model like (3.3) if we take conditional model, then the orthogonality condition between the error terms and $\sum_{j\neq i} w_{ij,n}y_{j,n}$ may not satisfy, so the OLS estimators will be inconsistent.

In an interesting paper, Lee (2002) proposed a consistent OLS estimator by considering special form of weights matrices; however we will not consider that approach. For the required uncorrelatedness what we need is to impose certain restrictions on the information set, which again unlike time dependent model is not one-sided. Another important features of spatial weights is that we should not restrict $\{w_{ij,n}\}$ to look for interaction between two geographically connected cross-sectional units. In other words the choice of spatial weights should not be exogenous to the model, it should follow certain underlying factors "which will depend upon the study in hand" (Cliff and Ord, 1973, p.12).

4 Kelejian-Prucha GM

By considering the simultaneous system (3.2) Kelejian and Prucha (2004) considered the issue of estimation based on limited and full information instrumental variables. The method is very similar to classical two and three stage least squares and computationally tractable. Based on Kelejian and Prucha (1999), the idea is to pick three moment condition to estimate spatial dependence parameter ρ. For spatially autoregressive error term of equations (3.3) these moment equations based on three moments corresponds to

$$\psi(u_{j,n},\theta) = \left(\; u'_{j,n}u_{j,n} - n\sigma^2, \overline{u}'_{j,n}\overline{u}_{j,n} - \sigma^2 Tr(W_nW_n), \overline{u}'_{j,n}u_{j,n} \; \right)', \tag{4.1}$$

where $\overline{u}_{j,n} = W_n u_{j,n}$. In there set up the number of moment condition becomes equal to the number of parameter to be estimated, so the question of optimal weights matrix does not arise. In summary, Kelejian-Prucha (2004) GM for spatial simultaneous system involve three easy steps:

- Estimate first part of (3.4) using 2SLS/IV with instruments Q where choice of instruments are linear columns of X_n and $W_n^s X_n$ for $s \geq 1$.
- Use the residual from first step and estimate ρ's and using (4.1) by $\min_\theta \psi(u_{j,n},\theta)'\psi(u_{j,n},\theta)$
- Obtain 2SLS estimates of δ's by Cochran-Orcutt type transformation

[1]Note that it may be the case $\sum_{j=1}^{n} w_{ij,n} \neq \sum_{i=1}^{n} w_{ij,n}$ but $\sum_{j=1}^{n} w^*_{ij,n} = \sum_{i=1}^{n} w^*_{ij,n}$, where $w^*_{ij,n} = w_{ij,n}/\sum_{j\neq i} w_{ij,n}$.

5 Spatial First Difference GMM

Ideally alternative specification of variance-covariance matrix associated with error vector should become the source of information. In our approach we utilize that by using first two moments based on the first difference of model (3.3). For regular panel data this approach has been analyzed by Holtz-Eakin, Newey and Rosen (1988), Arellano and Bond (1991) who applied this with respect to time variable. The way we select our moment conditions resembles that of Azomahou and Kuhry (2001) but our approach is more general. The idea is not new and the applicability of analogous method for a broad class of model has been introduced by McCurdy (1982), Chamberlain (1987), Newey (1993) and Newey and McFadden (1994) to name a few. Also similar procedure using instrumental variable procedure discussed by Amemiya and MaCurdy (1986) for an error component model. Even though the procedure varies, basically the sources of endogeneity and disturbance covariance properties provides the basis for asymptotically efficient estimators. We start by applying a Cochran-Orcutt type transformation to equation (3) and that yields the following

$$y_{j,n} = F(\overline{y}_{j,n}, \overline{x}_{j,n}, \gamma_0) + \epsilon_{j,n}, \tag{5.1}$$

where $F(\overline{y}_{j,n}, \overline{x}_{j,n}, \gamma_0) = \rho_j W_n Y_{j,n} + (z_{j,n} - \rho_j W_n Z_{j,n})\delta$, $\overline{x}_{j,n} = (W_n y_{j,n}, Z_{j,n}, W_n Z_{j,n})$ and $\gamma = (\delta', \rho)'$.

Now set $\overline{z}_n = (\overline{y}_{j,n}, \overline{x}_{j,n})$ and suppose that

$$E(\epsilon_{j,n}|\overline{x}_{j,n}) = 0 \text{ and } E(\epsilon_{j,n}\epsilon'_{j,n}|\overline{x}_{j,n}) = \Omega^*_{\epsilon,n}(\overline{x}_{j,n}, \gamma, \xi) = \Omega^*_{\epsilon,n},$$

i.e., the conditional covariance matrix has a known parametric form and it depends both on model parameter γ as well some unknown parameter ξ. We will discuss later the importance of this being homoskedastic, but for the time being assume that this known form is heteroskedastic. Note that, as a result of our assumption 4, the errors are assumed to be zero mean and finite unconditional variance Σ. Now to make error moments finite up to fourth order we make the following assumption

A6: *There exists two finite matrix Δ_3 and Δ_4, such that $E[\epsilon_{j,n}(\epsilon'_{j,n} \otimes \epsilon'_{j,n})]B'_1 = \Delta_3$, and $B_1 E(\epsilon_{j,n}\epsilon'_{j,n} \otimes \epsilon_{j,n}\epsilon'_{j,n})B'_1 = \Delta_4$, where B_1 is a selection matrix that eliminates from Δ_3 and Δ_4 some of the repeated cross-moments.*

Such assumption has been made by Arellano (1989) in calculating efficient GLS estimator for restricting error covariance matrix. In our context, later we will see that a similar general form of third and fourth order moments will play an important role in the asymptotic variance calculation.

By denoting $\theta = (\gamma', \xi')'$ and $\alpha = (\alpha'_1, \alpha'_2)'$, the first two moment restrictions can be expressed as

$$\psi(\overline{z}_{j,n}, \theta_0) = \begin{pmatrix} y_{j,n} - F(\overline{y}_{j,n}, \overline{x}_{j,n}, \gamma_0) \\ vech[(y_{j,n} - F(\overline{y}_{j,n}, \overline{x}_{j,n}, \gamma_0))(y_{j,n} - F(\overline{y}_{j,n}, \overline{x}_{j,n}, \gamma_0))' - \Omega^*_{\epsilon,n} \end{pmatrix} \tag{5.2}$$

Let there exists a random weights matrix $A_n = a_n' a_n$, where the sequence a_n has a full row rank that converges almost surely to a constant a and with H as $q \times s$ instruments matrix, ψ as $s \times 1$ vector, we consider

$$\varphi(\overline{z}_{j,n}, \theta) = H(\overline{x}_{j,n}) \otimes \psi(\overline{z}_{j,n}, \theta) \tag{5.3}$$

as a candidate that can be useful as population moment condition for a unique parameter θ_0. For the choice of instruments we make the following assumption.

A7: *Consider the class of spatial instruments $H(\overline{x}_{j,n})$ such that*

- **(a)** $E[H(\overline{x}_{j,n}) \frac{\partial \psi(\overline{z}_{j,n},\theta)}{\partial \theta}] \leq \Delta_1 < \infty$
- **(b)** $E[\|H(\overline{x}_{j,n})\|^2 \|\psi(\overline{z}_{j,n}, \theta)\|^2] \leq \Delta_2 < \infty$

The implication is that we are restricting our attention only to set of linear functions $H(\overline{x}_{j,n})$. If we follow Kelejian and Prucha (1998, 2004) tradition and choose a $H(\overline{x}_{j,n})$ as a subset of the of the linear independent columns of $(X_n, W_n X_n, \dots, W_n^s X_n)$, $1 \leq s \leq 2$, the above restriction seems innocuous. Also due to assumptions 3 and 5, the elements of $H(\overline{x}_{j,n})$ are bounded in absolute value. Since our interest is with moment conditions (5.2) we claim their orthogonality condition in the following definition.

Definition 5.1. The model (3.4) is said to be correctly specified if there exists a unique solution $\theta_0 \in \Theta \subset \mathbb{R}^p$ such that $E\left[\varphi(\overline{z}_{j,n}, \theta_0)\right] = 0$.

The characterization of $\varphi(\overline{z}_{j,n}, \theta_0)$ will be useful for finding asymptotic efficiency of our spatial simultaneous systems GMM estimator. Although different weighting scheme may produce a different consistent estimator, we will choose weights to achieve lower bound on asymptotic efficiency within a broad class of consistent estimators. Also instead of (5.3) we will use $E[\varphi(\overline{z}_{j,n}, \theta_0)] = E[H(\overline{x}_{j,n})\psi(\overline{z}_{j,n}, \theta_0)] = 0$ to classify sample moment condition. In addition we can think the instrumental variable function to depend on some nuisance parameters[2], so $\varphi(\overline{z}_{j,n}, \gamma) = H(\gamma)\psi(\overline{z}_{j,n}, \theta)$ where $\gamma \in \Gamma \subset \mathbb{R}^H$ and may include θ as well as $\overline{x}_{j,n}$.

5.1 Asymptotic properties

By applying GMM we based the estimation on a linear combination of empirical counterpart of moment condition using (5.2) and the random weights matrix $a_n' a_n$, so one choose $\widehat{\theta}_n$ to

$$\text{argmin}_\theta \varphi(\overline{z}_{j,n}, \theta)' a_n' a_n \varphi(\overline{z}_{j,n}, \theta)$$

and the minimizer of the quadratic form satisfies the first order condition

$$\left[\frac{\partial \varphi(\widehat{\theta}_n)}{\partial \theta'}\right]' a_n' a_n \varphi(\widehat{\theta}_n) = 0, \tag{5.4}$$

[2]In that case we need to assume that we have an estimator $\widehat{\gamma}_n$ such that $\sqrt{n}(\widehat{\gamma}_n - \gamma_0) = O_p(1)$, where γ_0 could contain θ_0.

where for simplicity of notation we used $\varphi_{j,n}(\widehat{\theta}_n) = \varphi(\overline{z}_{j,n}, \widehat{\theta}_n)$ and $\varphi_n(\widehat{\theta}_n)$ denotes the corresponding vector. Same is true for $\varphi_n(\theta_0)$. Now we define $Q_{1n} = \frac{1}{n}\sum_j \left[H(\overline{x}_{j,n})\frac{\partial\psi(\theta)}{\partial\theta'}\right]$ and assume the following.

A8: The instruments has the following properties

- **(a)** $lim_{n\to\infty}\left[H(\overline{x}_{j,n})E\left\{\frac{\partial\psi(\theta)}{\partial\theta'}\right\}\right]$ is a finite matrix with full column rank.
- **(b)** $Sup_\theta\|Q_{1n}(\theta) - E(Q_{1n}(\theta))\| \to^P 0.$

The first part of the above assumption has some flavor of rank condition made by Amemiya (1985, p.246) in the context of nonlinear 2SLS estimators consistency proof. Also Kelejian and Prucha (1997) mentioned that failure of this may lead to a fundamental identification problem in the sense that objective function becomes flat in the direction of spatial lag parameter ρ as $n \to \infty$. By taking the usual mean value expansion of $\varphi_n(\widehat{\theta_n})$ around θ_0 and multiplying through $\sqrt{n}$ we have

$$\sqrt{n}\varphi_n(\widehat{\theta}_n) = \sqrt{n}\varphi_n(\theta_0) + Q_{1n}(\widetilde{\theta}_n)\sqrt{n}(\widehat{\theta}_n - \theta_0), \tag{5.5}$$

where $\widetilde{\theta}$ has elements between $\widehat{\theta}$ and θ_0. Note that based on above the GMM problem in the close neighborhood of θ_0 can be written as

$$\|a_n n^{1/2}\varphi_n(\widehat{\theta}_n)\|^2 = \|a_n n^{1/2}\varphi_n(\widehat{\theta}_n) + Q_{2n}(\theta_0)n^{1/2}(\theta_n - \theta_0)\|^2,$$

where $Q_{2n}(\theta) = a_n Q_{1n}$. Combining (5.4) and (5.5) yields

$$\begin{aligned}\sqrt{n}(\widehat{\theta}_n - \theta_0) &= -\left\{\left[\frac{1}{n}\sum_i H(\overline{x}_{j,n})\frac{\partial\psi(\widehat{\theta})}{\partial\theta'}\right]' a_n' a_n \left[\frac{1}{n}\sum_i H(\overline{x}_{j,n})\frac{\partial\psi(\widehat{\theta})}{\partial\theta'}\right]\right\}^{-1} \\ &= \left[\frac{1}{n}\sum_i H(\overline{x}_{j,n})\frac{\partial\psi(\widehat{\theta})}{\partial\theta'}\right]' a_n' a_n n^{-1/2}\varphi_n(\theta_0) \\ &= -Q_{3n}n^{-1/2}\varphi_n(\theta_0)\end{aligned} \tag{5.6}$$

where $Q_{3n} = -[Q_{1n}'a_n'a_nQ_{1n}]'Q_{1n}'a_n'a_n = -[Q_{2n}'Q_{2n}]'Q_{2n}'a_n$. Also note that

$$\frac{1}{n}\sum_i\left[H(\overline{x}_{j,n})\frac{\partial\psi(\widehat{\theta})}{\partial\theta'}\right] \to E\left[H(\overline{x}_{j,n})\frac{\partial\psi(\theta)}{\partial\theta'}\right]$$

i.e., $Q_{1n}(\widehat{\theta}) \to Q_1$ in probability and

$$\begin{aligned}&\left\{\left[\frac{1}{n}\sum_i H(\overline{x}_{j,n})\frac{\partial\psi(\widehat{\theta})}{\partial\theta'}\right]' a_n' a_n \left[\frac{1}{n}\sum_i H(\overline{x}_{j,n})\frac{\partial\psi(\widehat{\theta})}{\partial\theta'}\right]\right\}^{-1}\left[\frac{1}{n}\sum_i H(\overline{x}_{j,n})\frac{\partial\psi(\widehat{\theta})}{\partial\theta'}\right]' a_n' a_n \\ &\to \left\{E\left[H(\overline{x}_{j,n})\frac{\partial\psi(\theta)}{\partial\theta'}\right]' a'aE\left[H(\overline{x}_{j,n})\frac{\partial\psi(\theta)}{\partial\theta'}\right]\right\}^{-1} E\left[H(\overline{x}_{j,n})\frac{\partial\psi(\theta)}{\partial\theta'}\right]' a'a\end{aligned}$$

in probability. However, for the asymptotic distribution of $n^{-1/2}\varphi_n(\theta_0)$ we need to use the following lemma.

Lemma 5.1. *Given assumptions A1-A5, we have*

$$n^{-1/2}a_n\varphi_n(\theta_0) \to^D N(0, a'Va)$$

where $V = E\left[H(\overline{x}_{j,n})\frac{\partial}{\partial\theta'}\psi(\overline{z}_{j,n},\theta)\frac{\partial}{\partial\theta'}\psi(\overline{z}_{j,n},\theta)'H(\overline{x}_{j,n})'\right]$.

By premultiplying (5.5) by a_n we have

$$a_n\sqrt{n}\varphi_n(\widehat{\theta}_n) = a_n\sqrt{n}\varphi_n(\theta_0) + a_nQ_{1n}(\widetilde{\theta}_n)\sqrt{n}(\widehat{\theta}_n - \theta_0) \tag{5.7}$$

Note that

$$\begin{aligned} a_n\sqrt{n}\varphi_n(\widehat{\theta}_n) &= a_n\sqrt{n}\varphi_n(\theta_0) + a_nQ_{1n}(\widetilde{\theta}_n)[-Q_{1n}(\widehat{\theta}_n)'a_n'a_nQ_{1n}(\widetilde{\theta}_n)]Q_{1n}(\widehat{\theta}_n)a_n'a_nn^{-1/2}\varphi_n(\theta_0) \\ &= [I_{mn} - a_nQ_{1n}(\widetilde{\theta}_n)[Q_{1n}(\widehat{\theta}_n)'a_nQ_{1n}(\widetilde{\theta}_n)]'Q_{1n}(\widehat{\theta}_n)]a_nn^{-1/2}\varphi_n(\theta_0) \\ &= [I_{mn} - a_n[Q_{2n}(\widehat{\theta}_n)'a_nQ_{1n}(\widetilde{\theta}_n)]'Q_{1n}(\widehat{\theta}_n)]a_nn^{-1/2}\varphi_n(\theta_0) \end{aligned} \tag{5.8}$$

exploiting which we have

$$a_n\sqrt{n}\varphi_n(\widehat{\theta}_n) = [I_{mn} - a_n[Q_{2n}(\widehat{\theta}_n)'a_nQ_{1n}(\widetilde{\theta}_n)]'Q_{1n}(\widehat{\theta}_n)]a_nn^{1/2}\varphi_n(\theta_0) + o_p(1) \tag{5.9}$$

Continuing from Lemma 4.2, take the usual mean value expansion of $\varphi(\overline{z}_{j,n}, \widehat{\theta}_n)$ around θ_0. Then combining this with (5.4) produces the familiar expression for $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ given by (5.10), the generic structure of which is similar to $\sqrt{n}a_n\varphi_n(\widehat{\theta})$.

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = -[Q_{2n}(\widehat{\theta}_n)'Q_{2n}(\widetilde{\theta}_n)]^{-1}Q_{2n}(\widehat{\theta}_n)a_nn^{1/2}\varphi_n(\theta_0) + o_p(1) \tag{5.10}$$

Now we are in a position to state the following result by utilizing the above structure.

Proposition 5.3: *Suppose that assumption A1-A5 holds,* $\theta_0 \in$ *interior of* Θ, $\varphi(\overline{z}_{j,n}, \theta)$ *is continuously differentiable in a neighborhood of* θ_0 $w.p$ 1, $a_nE(\varphi_n(\theta_n)) = 0$ *has unique root at* θ_0 *in* Θ *and* Q_1 *is nonsingular, we have*

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) \to^D N[0, Cov(\widehat{\theta}_n^H)],$$

where

$$Cov(\widehat{\theta}_n^H) = (Q_1'a'aQ_1)^{-1}Q_1'a'VaQ_1(Q_1'a'aQ_1)^{-1}$$

By exploiting some intermediate lemmas and a set of conditions analogous to Newey (1993) we have the above result on asymptotic normality. Not to our surprise this is the same type of bound on asymptotic variance derived by Conely (1999, p.8) for a set of general over-identified moment condition; however, that set up was based on random field structure and asymptotic result was derived by applying a central limit theorem due to Bolthausen (1982) for stationary, mixing random fields on regular lattices. In our case we utilize the triangular structure of the model which has the spirit of Kelejian and Prucha (1999). In some sense this avoids making some hard to check first moment continuity condition required for uniform convergence of the objective function.

5.2 Optimum GMM

In search for optimum GMM we basically seek to exploit the conditional moment restrictions of the form $E[\psi(\overline{z}_{j,n},\theta_0)|\overline{x}_{j,n}]=0$ to derive an unconditional moment restriction that

$$E[H(\overline{x}_{j,n})\psi(\overline{z}_{j,n},\theta_0)]=0$$

for some suitable $q\times s$ matrix of functions $H(\overline{x}_{j,n})$. Some additional conditions will be helpful.

A9: $E[H(\overline{x}_{j,n})D(\overline{x}_{j,n})]$ *and* $E[H(\overline{x}_{j,n})H(\overline{x}_{j,n})']$ *are non-singular matrix and*

$$E\left[H(\overline{x}_{j,n})\Omega(\overline{x}_{j,n})H(\overline{x}_{j,n})'\right]$$

is finite.

Now we can define a GMM estimator $\widehat{\theta}_n$ based on any moment functions of the form $\varphi(z_{j,n},\theta)=H(\overline{x}_{j,n})\psi(\overline{z}_{j,n},\theta)$ and show that the asymptotic variance of $\widehat{\theta}_n$ will be

$$\left\{E\left[H(\overline{x}_{j,n})\frac{\partial\psi(\overline{z}_{j,n},\theta_0)}{\partial\theta}\right]\right\}^{-1}E\left[H(\overline{x}_{j,n})\psi(\overline{z}_{j,n},\theta_0)\psi(\overline{z}_{j,n},\theta_0)'H(\overline{x}_{j,n})'\right]$$
$$\left\{E\left[H(\overline{x}_{j,n})\frac{\partial\psi(\overline{z}_{j,n},\theta_0)}{\partial\theta}\right]\right\}^{-1'}$$

Note that, in this asymptotic variance expression no weighting scheme is needed because in $\varphi(z_{j,n},\theta)$ the number of components is same as θ. For example, based on conditional moment restrictions using (5.2) the variance will be

$$Cov(\widehat{\theta})=\left[H(\overline{x}_{j,n})D(\overline{z}_{j,n})\right]^{-1}E\left[H(\overline{x}_{j,n})\Omega(\overline{x}_{j,n})H(\overline{x}_{j,n})'\right]\left[D(\overline{z}_{j,n})'H(\overline{x}_{j,n})'\right]^{-1}$$

where

$$\Omega(\overline{x}_{j,n},\theta_0)=E[\psi(\overline{z}_{j,n},\theta_0)\psi(\overline{z}_{j,n},\theta_0)'|\overline{x}_{j,n}]$$

and $D(\overline{z}_{j,n},\theta_0)=E\left[\frac{\partial}{\partial\theta}\psi(\overline{z}_{j,n},\theta_0)|\overline{x}_{j,n}\right]$. By the same argument of Chamberlain (1987) and Newey (1993) we observe that an efficient choice of instruments exists which suggests the following choice for optimal IV:

$$H^*(\overline{x}_{j,n})=D(\overline{x}_{j,n})'\Omega(\overline{x}_{j,n})^{-1}$$

and this function minimizes the asymptotic variance. Note here we are using $\overline{x}_{j,n}$ in the arguments of D to reflect the conditionality. So for spatial simultaneous system (3.4) the bound on asymptotic variance becomes

$$Cov(\widehat{\theta}_n^{H^*})=E\left[D(\overline{x}_{j,n})'\Omega(\overline{x}_{j,n})^{-1}D(\overline{x}_{j,n})\right]^{-1} \tag{5.11}$$

Interestingly, this asymptotic variance is invariant to non-singular linear transformation, so that for any non-singular constant matrix Υ, $H^{**}(\overline{x}_{j,n})=\Upsilon H^*(\overline{x}_{j,n})$ will also minimize the asymptotic variance. Note that, in our spatial simultaneous system we can express each components separately as

$$\frac{\partial}{\partial\theta}\varphi(\overline{z}_{j,n},\theta_0)=\begin{pmatrix}A_{11} & A_{12}\\ A_{21} & A_{22}\end{pmatrix},$$

where $A_{11} = (\partial/\partial\alpha_1')\left(y_{j,n} - F(\overline{y}_{j,n}, \overline{x}_{j,n}, \gamma_0)\right)$, $A_{12} = (\partial/\partial\alpha_1') vech\Omega^*_{\epsilon,n}$, $A_{21} = B_1 B_2[\epsilon_{j,n} \otimes A_{11}]$ and $A_{22} = (\partial/\partial\alpha_2'),\ vech\Omega^*_{\epsilon,n}$. Also, B_1 and B_2 are matrices of constants such that $vech(\epsilon_{j,n}\epsilon_{j,n}') = B_1 vech(\epsilon_n \epsilon_n')$ and $(\epsilon_{j,n} \otimes A_{11}) + (A_{11} \otimes \epsilon_n) = B_2(\epsilon_n \otimes A_{11})$ respectively. So

$$D(\overline{z}_{j,n}, \theta_0) = \begin{pmatrix} E(A_{11}|\overline{x}_{j,n}) & A_{12} \\ B_1 B_2 E[(\epsilon_{j,n} \otimes A_{11})|\overline{x}_{j,n}] & A_{22} \end{pmatrix}$$

In general $E[(\epsilon_{j,n} \otimes A_{11})|\overline{x}_{j,n}] \neq 0$, since A_{11} depends on $y_{j,n}$ unlike single equation case. Also, using assumption 6, and triangular structure of error covariance matrix we can derive

$$\Omega(\overline{x}_{j,n}, \theta_0) = \begin{pmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{pmatrix}$$

where $C_{11} = \Omega^*_{\epsilon,n}, C_{12} = E(\epsilon_{j,n}\epsilon_{j,n}' \otimes \epsilon_{j,n}'|\overline{x}_{j,n})B_1, C_{21} = B_1 E(\epsilon_{j,n}\epsilon_{j,n}' \otimes \epsilon_{j,n}|\overline{x}_{j,n})$ and $C_{22} = B_1 E(\epsilon_{j,n}\epsilon_{j,n}' \otimes \epsilon_{j,n}\epsilon_{j,n}'|\overline{x}_{j,n})B_1$.

Note that, this is block diagonal even when $E(\epsilon_{j,n}\epsilon_{j,n}' \otimes \epsilon_{j,n}|\overline{x}_{j,n}) = 0$. But $Cov(\widehat{\theta})$ is not block diagonal even if C_{12} or $C_{21} = 0$, since $D(\overline{z}_{j,n}, \theta_0)$ is not block diagonal. Now we are in a position to state the main result of this section.

Proposition 5.4: *For a conditional moment using (5.2)the optimal instrument is* $H^*(\overline{x}_{j,n}) = D(\overline{x}_{j,n})'\Omega(\overline{x}_{j,n})^{-1}$ *and so* $Cov(\widehat{\theta}_n^H) - Cov(\widehat{\theta}_n^{H^*})$ *is negative semi definite.*

So we observe that even in our more general case the optimal GMM estimators attains the efficiency bound. For the components of the efficient variance bound we can utilize the expression

$$D(\overline{x}_{j,n})'\Omega(\overline{x}_{j,n})^{-1}D(\overline{x}_{j,n}) = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \tag{5.12}$$

the components of which depends on the conditional variance matrix of the error term. However, the existence of unknown heteroskedasticity or nuisance parameter or both, will complicate the form of the covariance matrix form.

For example, when $\Omega^*_{\epsilon,n} = \Omega^*_{\epsilon,n}(\overline{x}_{j,n}, \gamma, \xi)$,

$$\Omega_{11} = E(A_{11}|\overline{x}_{j,n})'C_{11}E(A_{11}|\overline{x}_{j,n}) + B_1'B_2'E[(\epsilon_{j,n} \otimes A_{11})|\overline{x}_{j,n}]'C_{22}A_{12},$$
$$\Omega_{12} = E(A_{11}|\overline{x}_{j,n})'C_{11}B_1B_2E[(\epsilon_{j,n} \otimes A_{11})|\overline{x}_{j,n}] + B_1'B_2'E[(\epsilon_{j,n} \otimes A_{11})|\overline{x}_{j,n}]'C_{22}A_{22},$$
$$\Omega_{21} = A_{12}'C_{11}E(A_{11}|\overline{x}_{j,n}) + A_{22}C_{22}A_{12}, and$$
$$\Omega_{22} = A_{12}'C_{11}B_1B_2E[(\epsilon_{j,n} \otimes A_{11})|\overline{x}_{j,n}] + A_{22}'C_{22}A_{22}.$$

Similarly, when $\Omega^*_{\epsilon,n} = \Omega^*_{\epsilon,n}(\overline{x}_{j,n}, \xi)$,

$$\Omega_{11} = E(A_{11}|\overline{x}_{j,n})'C_{11}E(A_{11}|\overline{x}_{j,n}),$$
$$\Omega_{12} = E(A_{11}|\overline{x}_{j,n})'C_{11}B_1B_2E[(\epsilon_{j,n} \otimes A_{11})|\overline{x}_{j,n}] + B_1'B_2'E[(\epsilon_{j,n} \otimes A_{11})|\overline{x}_{j,n}]'C_{22}A_{22},$$
$$\Omega_{21} = 0, \text{ and } \Omega_{22} = A_{22}'C_{22}A_{22}.$$

Also if $\Omega^*_{\epsilon,n} = \Omega^*_{\epsilon,n}(\xi)$, i.e., we assume homoskedasticity, the Ω_{12} term of the covariance matrix will be zero only when there is no endogenous variable on the right hand side of (5.1). Otherwise there is no gain in efficiency by including non heteroskedasticity assumption in the model even when $E(\epsilon_{j,n}\epsilon'_{j,n} \otimes \epsilon_{j,n}|\overline{x}_{j,n}) = 0$. This illustration of potential efficiency comparison through non-zero third moment information has been discussed by MaCurdy (2001) in the context of simple heteskedasticity corrected least squares estimator.

6 Conclusion

This paper proposes an instrumental variable procedures for use with spatial simultaneous system that produces asymptotically efficient estimators for both spatial autoregressive parameters as well as other regression parameters. Since the underlying model is more general one can always construct different subcases (e.g., a general taxonomy is given in Table 5.1 of Rey and Boarnet, 2004) and investigate estimators optimal properties. We strongly believe that a general framework to exploit over-identifying information contained in the higher order moments of the error term for multi equation spatial system will give researchers more freedom asymptotically. However, as reported in Blundell and Bond (1998) for simple panel data, GMM in the equation in first differences of Arellano and Bond (1993) type sometimes provides very small and imprecise estimates. Also it may suffer weak instrument problem where regressors in first differences are weakly autocorrelated. In spatial domain it would be interesting to explore such possibilities, specially whether over-identified restrictions are typically rejected too often using moment test statistic based on our set up.

References

[1] Altonji, J. an Segal, L.(1996). Small sample bias in GMM estimation of covariance structures, *Journal of Business and Economic Statistics,* **14**, 353-366.

[2] Amemiya, T.(1985). *Advanced Econometrics*, Cambridge: Harvard University Press.

[3] Amemiya, T and MaCurdy, T.E.(1986). Instrumental variable estimation of an error components model, *Econometrica*, **54**, 886-881.

[4] Andrews, D.W.K.(1991). An empirical process central limit theorem for dependent nonidentically distributed random variables, *Journal of Multivariate Analysis*, **38**, 187-203.

[5] Anselin, L.(1988). *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, The Netherlands.

[6] Anselin, L. and Kelejian, H. H (1999). Testing for Spatial error autocorrelation in the presence of endogenous regressors, *International Regional Science Review*, **20**, 153-182.

[7] Anselin, L. and Bera, A. K.(1998). Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics, *Handbook of Applied Economic Statistics*, 237-289.

[8] Anselin, L., Bera. A. K, Florax, R. and Yoon, M.(1996). Simple Diagnostic Tests for Spatial Dependence, *Regional Science and Urban Economics*, **26**, 77-104.

[9] Arellano, M.(1989). On the efficient estimation of simultaneous equations with covariance restrictions, *Journal of Econometrics*, **42**, 247-265.

[10] Arellano, M. and Bond, S.R.(1991). Some tests of specification for panel data: monte carlo evidence and an application to employment equations, *Review of Economic Studies,* **58**, 277-297.

[11] Azomahou, T and Kuhry, Y.(2001). Efficient estimation of spatial autoregressive models, *Working Paper*, Department of Economics, Universite Louis Pasteur, France.

[12] Blundell, R and Bond, S.(1998). Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics*, **87**, 115-143.

[13] Bolthausen, E., (1982). On the central limit theorem for stationary mixing random fields, *The Annals of Probability*, **10**, 1047-1050.

[14] Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics,* **34**, 305-334.

[15] Cliff, A.D. and Ord, K. (1973). *Spatial Autocorrelation*, Pion, London.

[16] Conely, T.G.(1999). GMM Estimation with Cross Sectional Dependence, *Journal of Econometrics*, **92**, 1-45.

[17] Golan, A, Judge, G and Miller, D.(1998). Information recovery in simultaneous equations statistical models, In A.Ullah and D.Giles eds., *Handbook of Applied Economic Statistics*, 365-81, Marcel Dekker, New York.

[18] Hall, A, R. (2003). *Generalized Method of Moments,* Manuscript for "Advanced Text in Econometrics" series, Oxford University Press, Oxford, U.K.

[19] Hansen, L.P, Heaton, J and Yaron, Y.(1996). Large sample properties of GMM estimators, *Journal of Business and Economic Statistics,* **14**, 262-280.

[20] Holtz-Eakin, D., Newey, W and Rosen,H.(1988). Estimating vector autoregressions with panel data, *Econometrica*, **56**, 1371-1395.

[21] Imbens, G.W.(1997). One step estimators for over-identified GMM models, *Review of Economic Studies*, **64**, 359-383.

[22] Imbens, G.W, Spady, R.H and Johnson, P(1997). Information theoretic approaches to inference in moment condition models, *Econometrica*, **66**, 333-357.

[23] Kelejian, H. H and Prucha, I. R. (1997). Estimation of spatial regression models with autoregressive errors by two stage least squares procedures: a serious problem, *International Regional Science Review*, **20**, 103-112.

[24] Kelejian, H. H and Prucha, I. R. (1998). A Generalized Spatial Two-Stage Least Square Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances, *Journal of Real Estate Finance and Economics*, **17**, 99-121.

[25] Kelejian, H. H and Prucha, I. R. (1999). A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model, *International Economic Review*, **40**, 509- 532.

[26] Kelejian, H. H and Prucha, I. R. (2004). Estimation of simultaneous system of spatially interrelated cross sectional equations, *Journal of Econometrics*, **118**, 27-50.

[27] Kelejian, H. H and Robinson D.P.(1995). Spatial Correlation: A Suggested Alternative to the Autoregressive Model, In: Anselin R.,Florax R.J.G.M.(eds.) *New Directions in Spatial Econometrics*, Springer, Berlin.

[28] Kitamura, Y.(1997). Empirical likelihood methods for weakly dependent processes, *Annals of Statistics*, **25**, 2084-2102.

[29] Kitamura, Y. and Stutzer, M.(1997). An information theoretic alternative to GMM estimation, *Econometrica*, **65**, 861-874.

[30] Lee, L.F.(2001). GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models, *Working Paper*, Ohio State University, OH.

[31] Lee, L.F.(2002). Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models, *Econometric Theory*, **18**, 252-277.

[32] MaCurdy, T.E.(1982). Using information on the moments of the disturbance to increase the efficiency of estimation. *NBER Working Paper*#22, NBER, Cambridge, MA.

[33] MaCurdy, T.E.(2001). Using information on the moments of the disturbance to increase the efficiency of estimation. In Hsiao, Morimune and Powell eds. *Nonlinear Statistical Modelling*.

[34] Magnus, J.R and Neudecker, H.(1999). *Matrix differential calculus with applications in statistics and econometrics*, Revised edition, Wiley, New York.

[35] Newey, W.K. (1993). Efficient estimation of models with conditional moment restrictions, G.S.Maddala, C.R.Rao and H.D.Vinod (eds.)*Handbook of Statistics*, **11**, 419-454, Elsevier Science Publishers B.V.

[36] Newey, W.K., McFadden, D., (1994). Large Sample Estimation and Hypothesis Testing, *Handbook of Econometrics*, Ch.36, **4**, Noth-Holland, New York.

[37] Ord, K. (1975). Estimation Methods for Models of Spatial Interaction, *Journal of the American Statistical Association*, **70**, 120-126.

[38] Pinske, J, Shen, L and Slade (2003). Dynamic spatial probit with fixed effects using one step GMM: an application to mine operating decisions, *Working Paper*, Department of Economics, The Pennsylvania State University, PA.

[39] Qin, J and Lawless, J.(1994). Empirical likelihood and general estimating equations, *Annals of Statistics*, **22**, 300-325.

[40] Rey, S.J and Boarnet, M.G.(2004). A taxonomy of spatial econometric models for simultaneous systems, In L. Anselin, R. Florax and S.J.Rey (eds.) *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer, New York.

[41] Rao, C.R and Rao, M.B (1998). *Matrix algebra and its applications in statistics and econometrics*, World Scientific, London.

[42] Robertson, J and Pagan, A.(1997). Exploring some studies of finite sample performance of GMM, *Mimio*, Australian National University.

[43] Robinson, P.(1987). Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form, *Econometrica*, **55**, 875-891.

[44] Smirnov, O and Anselin, L. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. *Computational Statistics and Data Analysis*, **35**, 301-319.

A Appendix

In this section we provide the technical details of our presentation and proof of theoretical results.

A.1 Matrix notations

Let m and n be two fixed positive integers. We denote a matrix of dimension $m \times n$ with entries a_{ij} as $A_{m,n} = (a_{ij})$ (for $m = n$ by A_n etc.) and collection of such matrices as $M_{m,n}$. So $M_{m,n}$ will be a vector space with the usual operations of addition and scalar multiplication of matrices over the field $\mathbb{R}$ of real numbers. For any $A_{m,n} \in M_{m,n}$, we denote $\|.\|_A$ as a matrix norm on the vector space $M_{m,n}$. We use the convention that a matrix $A_n = (a_{ij})$ of order $n \times n$ is diagonally dominant if $|a_{ii}| > \sum_{j \neq i}^{n} |a_{ij}|, i = 1, \ldots, n$, and the following easily verified fact: if A is a matrix and a is a vector, then $\|Aa\| \leq \|A\| \|a\|$.

For a $m \times n$ matrix $A_{m,n}$ by $vecA$ we mean a $m \times 1$ vector whose j-th column is a_i. For any two column vector a and b, the notation that we use to connect vec operator and the Kronecker product is $vec(ab') = b \otimes a$. If A and B are both upper (lower) triangular matrix, then $A \otimes B$ is a upper (lower) triangular matrix.

For any two matrix $A_{m,n}$ and $B_{p,q}$, we use the result $vec(A \otimes B) = (I_n \otimes K_{q,m} \otimes I_p)(vecA \otimes vecB)$ where $K_{q,m}$ is the commutation matrix. Also for vec derivative we use the following: for a vector of $c = (c_1, \ldots, c_s)$ of s variables, $(\partial vec(aa')/\partial c') = ((\partial a/\partial c') \otimes a) + (a \otimes (\partial a/\partial c'))$ and $(\partial vec(A \otimes B)/\partial c') = (I_n \otimes K_{q,m} \otimes I_p)((vecA) \otimes (\partial vecB/\partial c') + (\partial vecA/\partial c') \otimes (vecB))$.

For more see Magnus and Neudecker (1999,p.30,47,184), Rao and Rao (1998,p.368).

A.2 Proofs

Proposition 3.2: This result follows directly from Kelejian and Prucha (2004). Limiting distribution of ϵ_n is equivalent to the limiting distribution of y_n and u_n. The exact functional form determines the corresponding mean and variance.

Lemma 5.2: Let $m_{j,n} = a_n\varphi_{j,n}$, where again for simplicity denote $\varphi_{j,n} = \varphi(\overline{z}_{j,n}, \theta)$. We observe $\{m_{j,n} : 1 \le j \le n, n \ge 1\}$ is a triangular array of independent random variables. Note that $Em_{j,n} = 0$ and

$$
\begin{aligned}
E(m_{j,n})^2 &= E(a_n\varphi_{j,n})^2 \\
&\le E(\|a_n\|^2\|\varphi_{j,n}\|^2) \\
&= \|a_n\|^2 E\|\varphi_{j,n}\|^2 < \infty, n \ge 1
\end{aligned}
$$

The proof of the result lies in the fact that by Cramér-Wold device we can show that $a_n \sum_j \varphi_{j,n}$ converge in distribution to $M_n N(0, I)$ where $M_n = \|a_n\|_\nabla$ and $\nabla = \sum_j (E\varphi_{j,n}\varphi'_{j,n})$. Note also that

$$
\begin{aligned}
M_n^2 &= \sum_{j=1}^{n} E(m_{j,n})^2 \\
&= \sum_j a'_n E\varphi_{j,n}\varphi'_{j,n} a_n \\
&= a'_n(\sum_j E\varphi_{j,n}\varphi'_{j,n})a_n \\
&= \|a_n\|_\nabla^2, n \ge 1
\end{aligned}
$$

Next we verify that the triangular array $\{m_{j,n} : 1 \le j \le n, n \ge 1\}$ satisfy the Lindberg condition. Fix $\varrho > 0$,

$$
\begin{aligned}
M_n^{-2} \sum_{j=1}^{n} E(m_{j,n}^2) I(|m_{j,n}| > \varrho M_n) &= \|a_n\|_\nabla^2 \sum_j E(a_n\varphi_{j,n}) I(|a_n\varphi_{j,n}| > \varrho\|a_n\|_\nabla) \\
&\le \|a_n\|_\nabla^2 \sum_j E(\|a_n\|^2\|\varphi_{j,n}\|^2) I(\|a_n\|\|\varphi_{j,n}\| > \varrho\|a_n\|_\nabla) \\
&= \sum_j E(\|\varphi_{j,n}\|^2) I(\|\varphi_{j,n}\| > \varrho) \to 0
\end{aligned}
$$

as $n \to \infty$ by the multivariate version of Lindberg condition and so the limit condition of the triangular array central limit theorem holds. So we can apply CLT for triangular arrays to demonstrate that $M_n^{-1}\sum_j m_{j,n} \to N(0,I)$, from which the result follows.

Proposition 5.3: Proof of this kind of results are widely known, here we provide a brief outline. The first order conditions for minimization problem are (5.4). By applying the usual mean value expansion of (5.4) and using the fact that there exists a positive definite matrix a such that $a_n \to a$, use a pointwise law of large numbers so that $\frac{\partial \varphi(\overline{\theta})}{\partial \theta} \to^p E[H(\overline{x}_{j,n})\frac{\partial}{\partial \theta}\psi(\overline{z}_{j,n},\theta_0)]$ for any $\overline{\theta} \to^p \theta_0$.

Note that under the maintained assumption the elements of Q_2 are bounded in absolute value, so a multivariate version of the Lindberg condition can be easily verified. Now by replacing estimated averages by their probability limits and applying the Slutzky theorem on (5.10) yields the asymptotic distribution of right hand side. Then by Cramér-Wold device and another application of Stutzky's theorem results in the statement of the proposition. In plain words, we analyze the limiting behavior of a random matrix and a random vector separately and at the end combine them to derive our GMM estimators asymptotic distribution.

Proposition 5.4: A general form of the result is given in theorem 7.2.1 of Hall (2003), we basically extend that in our spatial set up. Let us denote $\widehat{\theta}(H)$ and $\widehat{\theta}(H^*)$ to be the estimates using any H and optimal H^* and $N(\theta^H)$ be the first quantity on the right hand side of (5.10). Then we can write

$$\varphi(\theta) = \varphi(\overline{z}_{j,n},\theta) = \frac{1}{n}\sum_i H(\overline{x}_{j,n})\psi(\overline{z}_{j,n},\theta)$$

Let

$$\begin{aligned}
\varphi_n^H &= \varphi_n^H(\theta) = \frac{1}{n}\sum_i H(\overline{x}_{j,n})\psi(\overline{z}_{j,n},\theta)\\
\varphi_n^{H^*} &= \varphi_n^{H^*}(\theta) = \frac{1}{n}\sum_i H^*(\overline{x}_{j,n})\psi(\overline{z}_{j,n},\theta)\\
\widehat{\theta}(H) &= \hat{\theta}(H^*) + [\hat{\theta}(H) - \hat{\theta}(H^*)]
\end{aligned}$$

or,

$$\sqrt{n}[\widehat{\theta}(H) - \theta_0] = \sqrt{n}[\widehat{\theta}(H^*) - \theta_0] + \sqrt{n}[\widehat{\theta}(H) - \widehat{\theta}(H^*)]$$

where each term follows (5.10). So in terms of asymptotic variance this can be written as

$$Asy.Var\widehat{\theta}(H) = Asy.Var\widehat{\theta}(H^*) + Asy.Var[\widehat{\theta}(H) - \widehat{\theta}(H^*)]$$

or,

$$Cov(\widehat{\theta}(H)) = Cov(\widehat{\theta}(H^*)) + \Phi + \Phi'$$

Now $\Phi = 0$ if $Asy.Var[\widehat{\theta}(H), \widehat{\theta}(H^*)] = Asy.Var(\widehat{\theta}(H^*))$, which we is not true since under our maintained assumptions $E[N(\theta^H)N(\theta^{H^*})]$ is positive semi definite. Hence the result follows.