

Ingmar R. Prucha

Contents

80.1	Introduction	1597
80.2	A Primer on GMM Estimation	1599
80.2.1	Model Specification and Moment Conditions	1599
80.2.2	One-Step GMM Estimation	1602
80.2.3	Two-Step GMM Estimation	1604
80.3	GMM Estimation of Models with Spatial Lags	1606
80.3.1	GMM Estimation of Spatial-Autoregressive Parameter	1608
80.3.2	GMM Estimation of Regression Parameters	1609
80.3.3	Guide to Literature	1611
80.3.4	Exemplary GMM Estimators	1612
80.4	GMM Estimation of Models with Spatial Mixing	1615
80.5	Conclusion	1615
	References	1616

Abstract

The chapter discusses generalized method of moments (GMM) estimation methods for spatial models. Much of the discussion is on GMM estimation of Cliff-Ord-type models where spatial interactions are modeled in terms of spatial lags. The chapter also discusses recent developments on GMM estimation from data processes which are spatially α -mixing.

I.R. Prucha
Department of Economics, University of Maryland, College Park, MD, USA
e-mail: prucha@econ.umd.edu

M.M. Fischer, P. Nijkamp (eds.), *Handbook of Regional Science*,
DOI 10.1007/978-3-642-23430-9_90, © Springer-Verlag Berlin Heidelberg 2014

1597

Generalized Method of Moments Estimation of Spatial Models

Ingmar R. Prucha ¹

April 15, 2012

¹Department of Economics, University of Maryland, College Park, MD 20742,
Tel.: +1-301-405-3499, e-mail: prucha@econ.umd.edu

Abstract

The paper discusses generalized method of moments (GMM) estimation methods for spatial models. Much of the discussion is on GMM estimation of Cliff-Ord type models where spatial interactions are modeled in terms of spatial lags. The paper also discusses recent developments on GMM estimation from data processes which are spatially α -mixing.

1 Introduction¹

Spatial econometric models have a long history. Paelink and Klaassen (1979) may arguably be viewed as the first comprehensive volume covering spatial econometrics. Anselin (2010) provides a recent review of the development of the field of spatial econometrics over the last thirty years. Important texts include Anselin (1988), Arbia (2006), Cliff and Ord (1973, 1981), Cressie (1993), Haining (2003), LeSage and Pace (2009).

Spatial models provide a formal expression of Tobler (1970)'s first law of geography stating that "Everything is related to everything else, but near things are more related to each other". An important aspect of spatial econometrics is the focus on the explicit modeling and empirical estimation of pathways of spatial interactions. That is, an important aspect is the focus on exploring the structure of spatial interactions, and not just on accounting for cross-sectional correlation in the computation of standard errors for parameter estimators.

Much of the spatial econometrics literature has focused on cross sectional data or panel data where the time dimension is small. A reason is that in situations where the time dimension, say T , is large relative to the cross-sectional dimension, say n , we can often simply employ classical methods for the estimation of simultaneous time series models to estimate general forms of spatial interactions. If the time dimension T is one or small, estimation will only be possible if we impose some parsimonious structure on the form of spatial interactions.

The development of a formal theory of estimation of spatial models has lagged behind corresponding developments for inference from time series data. A formal theory of inference requires the use of limit theorems, such as laws of large numbers and central limit theorems. In a time series setting there is a natural ordering of the data which can be exploited in deriving such limit theorems. In a spatial setting there is no natural ordering of the data, which made the development of such limit theorems more challenging.

Arguably the most widely used class of spatial models are variants of the ones considered in Cliff and Ord (1973,1981). In these models spatial interactions are modeled in terms of spatial lags, i.e., in terms of weighted averages of observations from neighboring units, where the weights are typically modeled

¹I would like to thank James LeSage and Pablo Salinas Macario for their helpful comments on this paper.

as inversely related to some measure of distance. Historically, Cliff-Ord-type models have been estimated by maximum likelihood (ML) methods. (See Pace chapter Maximum likelihood estimation, as well as Mills and Parent chapter on Bayesian MCMC estimation). However, one of the difficulties with ML is that the likelihood depends on the determinant of an $n \times n$ matrix, which limits its application to small and medium sample sizes due to the computational burden (unless the problem is sparse, etc.). Another issue was the lack of formal results concerning its asymptotic properties. In light of this, Kelejian and Prucha (1998,1999) suggested a generalized method of moments (GMM) estimator for a spatial autoregressive model with autoregressive disturbances, and established basic asymptotic properties for the estimator.² Conley (1999) considered GMM estimation within the context of α -mixing spatial processes and developed an asymptotic theory within this context.

Since those early contributions there has been a growing literature on GMM estimation for spatially dependent data. The aim of this paper is to provide some guidance through that literature, and to provide some insights into the subtle differences in asymptotic results. Basic reasons for these differences can be found in the moment conditions employed by respective GMM estimators, and whether or not an estimator is a one-step or a two-step estimator.

Owing to space limitations the literature cited in this paper is incomplete, and not all contributions and extensions of interest are covered. Also, the focus of this paper is solely on GMM estimation. It does not cover maximum likelihood estimation or testing procedures (apart from Wald tests that can be constructed in the usual way based on results for the asymptotic distribution of GMM estimators). Also, the paper does not cover inference for processes where cross-sectional dependence is implied by common factors.

Finally, while spatial models have a long history in geography and regional science, space is not limited to geographic space. Spatial models may more generally be viewed as a class of cross-sectional interaction models, with applications ranging from growth convergence among regions to social interactions between agents.

Section 2 of the paper contains a brief and intuitive primer on GMM

²Lee (2004) gives, to the best of our knowledge, first formal results for the maximum likelihood estimator of a spatial autoregressive model. The maintained assumptions are similar to those introduced in Kelejian and Prucha (1998,1999).

estimation to provide some background. Readers familiar with GMM estimation may wish to skip this section. Section 3 considers GMM estimation of models with spatial lags, and Section 4 considers GMM estimation for a general class of spatially mixing processes.

2 A Primer on GMM Estimation

2.1 Model Specification and Moment Conditions

Suppose the data are generated from a model

$$f(y_{in}, z_{in}, \theta_0) = u_{in}, \quad i = 1, \dots, n, \quad (1)$$

where y_{in} denotes the dependent variable corresponding to unit i , z_{in} is a vector of explanatory variables, u_{in} is a disturbance term, θ_0 is the $K \times 1$ unknown parameter vector, and $f(\cdot)$ is a known function. The above formulation is fairly general and contains typical Cliff-Ord (1973,1981) spatial models - possibly after some transformation to remove correlation in the disturbance term - as a special case. Additionally assume the availability of a $1 \times P$ vector of instruments h_{in} and let w_{in} be the vector of all observable variables, including instruments, pertaining to the i -th unit. For simplicity of presentation we assume in the following that the disturbances are i.i.d. $(0, \sigma^2)$ and that the instruments are non-stochastic, while noting that both assumptions can be relaxed.

We also note that in allowing for the variables to depend on the sample size we accommodate spatial lags. As an example, the explanatory variables could be of the form $z_{in} = [x_i, \bar{x}_{in}, \bar{y}_{in}]$ where x_i is some exogenous explanatory variable, and $\bar{x}_{in} = \sum_j m_{ij} x_j$ and $\bar{y}_{in} = \sum_j m_{ij} y_{jn}$ are spatial lags (where the m_{ij} denote spatial weights with $m_{ii} = 0$). However, to simplify notation for this primer we will suppress the index n in the following.

Now suppose that we have a $S \times 1$ vector of sample moments

$$\mathbf{q}_n(\theta) = \mathbf{q}_n(w_1, \dots, w_n, \theta) = \begin{bmatrix} q_{1,n}(w_1, \dots, w_n, \theta) \\ \vdots \\ q_{S,n}(w_1, \dots, w_n, \theta) \end{bmatrix} \quad (2)$$

with $S \geq K$, and suppose that

$$E\mathbf{q}_n(w_1, \dots, w_n, \theta) = 0 \quad \text{if and only if} \quad \theta = \theta_0. \quad (3)$$

The basic idea underlying the GMM methodology is to estimate θ_0 by, say, $\tilde{\theta}_n$ such that $\mathbf{q}_n(w_1, \dots, w_n, \tilde{\theta}_n)$ is “close to zero” in the sense that a quadratic form of the sample moment vector is close to zero. More specifically, let Υ_n be some $S \times S$ positive semi-definite weighting matrix, then the corresponding GMM estimator is defined as

$$\tilde{\theta}_n = \arg \min_{\theta} \{ \mathbf{q}_n(w_1, \dots, w_n, \theta)' \Upsilon_n \mathbf{q}_n(w_1, \dots, w_n, \theta) \}. \quad (4)$$

A special case arises if the number of moments equals the number of unknown parameters, i.e., if $S = K$. In this case $\tilde{\theta}_n$ can typically be found as a solution to the moment condition, i.e., $\mathbf{q}_n(w_1, \dots, w_n, \tilde{\theta}_n) = 0$. Of course, in this case the weighting matrix Υ_n becomes irrelevant.

The classical GMM literature exploits “linear” moment conditions of the form

$$E \left\{ n^{-1} \sum_{i=1}^n h'_{ip} u_i \right\} = 0, \quad (5)$$

which clearly holds since $E h'_{ip} u_i = h'_{ip} E u_i = 0$ under the maintained assumptions. The spatial literature frequently also considers “quadratic” moment conditions. Let $A_q = (a_{ijq})$ be some $n \times n$ matrix with $tr(A_q) = 0$, and assume for ease of exposition that A_q is non-stochastic. Then the quadratic moment conditions considered in the spatial literature are of the form

$$E \left\{ n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ijq} u_i u_j \right\} = 0, \quad (6)$$

which clearly holds under maintained assumptions.³

Depending on the functional form of $f(\cdot)$, the number of moment conditions, the number of parameters, etc., the computation of the GMM estimator $\tilde{\theta}_n$ defined by (4) may be numerically challenging. Now let $\theta_0 = [\rho'_0, \delta'_0]'$ and suppose the sample moment vector in (2) can be decomposed into

$$\mathbf{q}_n(w_1, \dots, w_n, \theta) = \begin{bmatrix} \mathbf{q}_n^\rho(w_1, \dots, w_n, \rho, \delta) \\ \mathbf{q}_n^\delta(w_1, \dots, w_n, \rho, \delta) \end{bmatrix} \quad (7)$$

³Let $u = [u_1, \dots, u_n]'$, then the above moment condition can be rewritten as $E [n^{-1} u' A_q u] = tr [n^{-1} A_q E u u'] = n^{-1} \sigma^2 tr(A_q) = 0$, since under the maintained assumptions $E u u' = \sigma^2 I_n$.

such that

$$E\mathbf{q}_n^\rho(w_1, \dots, w_n, \rho, \delta_0) = 0 \quad \text{if and only if} \quad \rho = \rho_0, \quad (8)$$

$$E\mathbf{q}_n^\delta(w_1, \dots, w_n, \rho_0, \delta) = 0 \quad \text{if and only if} \quad \delta = \delta_0, \quad (9)$$

and that some easily computable initial estimator, say $\check{\delta}_n$, for δ_0 is available. In this case we may consider the following GMM estimator for ρ_0 corresponding to some weighting matrix $\Upsilon_n^{\rho\rho}$:

$$\hat{\rho}_n = \arg \min_{\rho} \left\{ \mathbf{q}_n^\rho(w_1, \dots, w_n, \rho, \check{\delta}_n)' \Upsilon_n^{\rho\rho} \mathbf{q}_n^\rho(w_1, \dots, w_n, \rho, \check{\delta}_n) \right\}. \quad (10)$$

Of course, utilizing $\hat{\rho}_n$ we may further consider the following GMM estimator for δ_0 corresponding to some weight matrix $\Upsilon_n^{\delta\delta}$:

$$\hat{\delta}_n = \arg \min_{\delta} \left\{ \mathbf{q}_n^\delta(w_1, \dots, w_n, \hat{\rho}_n, \delta)' \Upsilon_n^{\delta\delta} \mathbf{q}_n^\delta(w_1, \dots, w_n, \hat{\rho}_n, \delta) \right\}. \quad (11)$$

GMM estimators like $\tilde{\theta}_n$ in (4) are often referred to as one-step estimators. Estimators like $\hat{\rho}_n$ and $\hat{\delta}_n$ in (10) and (11) above, where the sample moments depend on some initial estimator, are often referred to as two-step estimators.

Given the moment conditions are valid, we would expect the most efficient one-step estimator to be more efficient than the most efficient two-step estimators. However, as usual, there are trade-offs. One trade-off is in terms of computation. As remarked previously, for small sample sizes ML is available as an alternative to GMM. For large sample sizes, statistical efficiency may be less important than computational efficiency and feasibility, and thus the use of two-step GMM estimators may be attractive. Also, Monte Carlo studies suggest that in many situations the loss of efficiency may be relatively small. Another trade-off is that the misspecification of one moment condition will typically result in inconsistent estimates of all model parameters.

In the following we provide some basic results for the limiting distribution of one-step and two-step GMM estimators as background for our discussion of specific GMM estimators for respective spatial models.

2.2 One-Step GMM Estimation

The usual approach to deriving the limiting distribution of GMM estimators is to manipulate the score of the objective function by expanding the sample

moment vector around the true parameter, using a Taylor expansion. Applying this approach to (4), and assuming that typical regularity conditions hold, yields

$$n^{1/2}(\widehat{\theta}_n - \theta_0) = -[\mathbf{G}'\Upsilon\mathbf{G}]^{-1}\mathbf{G}'\Upsilon [n^{1/2}\mathbf{q}_n(\theta_0)] + o_p(1) \quad (12)$$

with $\mathbf{G} = p \lim_{n \rightarrow \infty} \partial \mathbf{q}_n(\theta_0) / \partial \theta$ and $\Upsilon = p \lim_{n \rightarrow \infty} \Upsilon_n$. Now suppose for a moment that it can be shown that

$$n^{1/2}\mathbf{q}_n(\theta_0) \xrightarrow{d} N(0, \Psi), \quad (13)$$

where Ψ is some positive definite matrix. Then

$$n^{1/2}(\widehat{\theta}_n - \theta_0) \xrightarrow{d} N[0, \Phi] \quad (14)$$

with

$$\Phi = [\mathbf{G}'\Upsilon\mathbf{G}]^{-1}\mathbf{G}'\Upsilon\Psi\Upsilon\mathbf{G}[\mathbf{G}'\Upsilon\mathbf{G}]^{-1}.$$

From this it is seen that if we choose $\Upsilon_n = \widehat{\Psi}_n^{-1}$ where $\Psi = p \lim_{n \rightarrow \infty} \widehat{\Psi}_n$, the variance-covariance simplifies to

$$\Phi = [\mathbf{G}'\Psi^{-1}\mathbf{G}]^{-1}.$$

Since $[\mathbf{G}'\Upsilon\mathbf{G}]^{-1}\mathbf{G}'\Upsilon\Psi\Upsilon\mathbf{G}[\mathbf{G}'\Upsilon\mathbf{G}]^{-1} - [\mathbf{G}'\Psi^{-1}\mathbf{G}]^{-1}$ is positive semidefinite it follows that using for the weighting matrix Υ_n a consistent estimator of the inverse of the limiting variance-covariance matrix Ψ of the sample moment vector yields the efficient GMM estimator.

As remarked above, for spatial estimators the sample moment vector will typically be composed of linear and quadratic moment conditions of the form given in (4) and (5). Thus in order to establish (13) we need a central limit theorem (CLT) for linear quadratic forms. Kelejian and Prucha (2001) introduced such a theorem for a single linear quadratic form under assumptions useful for spatial models. The generalization to vectors of linear quadratic forms is given in Kelejian and Prucha (2010). To provide some insight into the expressions for the asymptotic variance-covariance matrix Ψ associated with the sample moment vector underlying the spatial GMM estimators below we next give a version of that CLT.

Theorem 1 For $r = 1, \dots, m$ let $A_{r,n} = (a_{ijr})_{i,j=1,\dots,n}$ be a $n \times n$ nonstochastic symmetric real matrix with $\sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^n |a_{ijr}| < \infty$, and let $a_r = (a_{1r}, \dots, a_{nr})'$ be a $n \times 1$ nonstochastic real vector with $\sup_n n^{-1} \sum_{i=1}^n |a_{ir}|^{\delta_1} < \infty$ for some $\delta_1 > 2$. Let $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$ be a $n \times 1$ random vector with the ε_i distributed totally independent with $E\varepsilon_i = 0$, $E\varepsilon_i^2 = \sigma_i^2$ and $\sup_{1 \leq i \leq n, n \geq 1} E|\varepsilon_i|^{\delta_2} < \infty$ for some $\delta_2 > 4$. Consider the $m \times 1$ vector of linear quadratic forms $V_n = [Q_{1n}, \dots, Q_{mn}]'$ with

$$Q_{rn} = \varepsilon' A_r \varepsilon + a_r' \varepsilon = \sum_{i=1}^n \sum_{j=1}^n a_{ijr} \varepsilon_i \varepsilon_j + \sum_{i=1}^n a_{ir} \varepsilon_i.$$

Let $\mu_{V_n} = EV_n = [\mu_{Q_1}, \dots, \mu_{Q_m}]'$ and $\Sigma_{V_n} = [\sigma_{Q_{rs}}]_{r,s=1,\dots,m}$ denote the mean and VC matrix of V_n , respectively, then

$$\begin{aligned} \mu_{Q_r} &= \sum_{i=1}^n a_{iir} \sigma_i^2, \\ \sigma_{Q_{rs}} &= 2 \sum_{i=1}^n \sum_{j=1}^n a_{ijr} a_{ijs} \sigma_i^2 \sigma_j^2 + \sum_{i=1}^n a_{ir} a_{is} \sigma_i^2 \\ &\quad + \sum_{i=1}^n a_{iir} a_{iis} \left[\mu_i^{(4)} - 3\sigma_i^4 \right] + \sum_{i=1}^n (a_{ir} a_{iis} + a_{is} a_{iir}) \mu_i^{(3)}, \end{aligned}$$

with $\mu_i^{(3)} = E\varepsilon_i^3$ and $\mu_i^{(4)} = E\varepsilon_i^4$. Furthermore, given that $n^{-1} \lambda_{\min}(\Sigma_{V_n}) \geq c$ for some $c > 0$, then

$$\Sigma_{V_n}^{-1/2} (V_n - \mu_{V_n}) \xrightarrow{d} N(0, I_m)$$

and thus

$$n^{-1/2} (V_n - \mu_{V_n}) \sim AN(0, n^{-1} \Sigma_{V_n}).$$

Remark: Note that the mean μ_{Q_r} of Q_{rn} is zero if $a_{iir} = 0$; if the ε_i are homoskedastic, i.e., $\sigma_i^2 = \sigma^2$, then $\text{tr}(A_r) = \sum_{i=1}^n a_{iir} = 0$ suffices for the mean to be zero. Next note that the first two terms in the expression for the covariance $\sigma_{Q_{rs}}$ between Q_{rn} and Q_{sn} can be written more compactly as $2\text{tr}(A_r \Sigma A_s \Sigma) + a_r' \Sigma a_s$ with $\Sigma = \text{diag}(\sigma_i^2)$. Also note that if $a_{iir} = a_{iis} = 0$, then the last two terms drop out from the expression for covariance. Observe further that under normality the last two terms are always equal to zero.

2.3 Two-Step GMM Estimation

The derivation of the limiting distribution two-step of GMM estimators is a bit more delicate. The usual approach to deriving the limiting distribution of two-step GMM estimators is to manipulate the score of the objective function by expanding the sample moment vector around the true parameter, using a Taylor expansion. Consider in particular the two-step GMM estimators for ρ_0 defined in (10). Applying this approach, and assuming typical regularity conditions, yields

$$\begin{aligned} & n^{1/2}(\widehat{\rho}_n - \rho_0) \\ &= -[(\mathbf{G}^{\rho\rho})'\mathbf{\Upsilon}^{\rho\rho}\mathbf{G}^{\rho\rho}]^{-1}(\mathbf{G}^{\rho\rho})'\mathbf{\Upsilon}^{\rho\rho} \left[n^{1/2}\mathbf{q}_n^\rho(\rho_0, \delta_0) + \mathbf{G}^{\rho\delta}n^{1/2}(\check{\delta}_n - \delta_0) \right] + o_p(1) \end{aligned} \quad (15)$$

where $\mathbf{G}^{\rho\rho} = p \lim_{n \rightarrow \infty} \partial \mathbf{q}_n^\rho(\rho_0, \delta_0) / \partial \rho$, $\mathbf{G}^{\rho\delta} = p \lim_{n \rightarrow \infty} \partial \mathbf{q}_n^\rho(\rho_0, \delta_0) / \partial \delta$, and $\mathbf{\Upsilon}^{\rho\rho} = p \lim_{n \rightarrow \infty} \mathbf{\Upsilon}_n^{\rho\rho}$. From (15) we see that in general the limiting distribution of $\widehat{\rho}_n$ will depend on the limiting distribution of $\check{\delta}_n$, unless $\mathbf{G}^{\rho\delta} = 0$, in which case we refer to $\check{\delta}_n$ as a nuisance parameter. It turns out that if ρ_0 denotes the spatial autoregressive parameters in the disturbance process and δ_0 the vector of regression parameters, then for typical estimators $\mathbf{G}^{\rho\delta} \neq 0$. In many cases the estimator $\check{\delta}_n$ will be asymptotically linear in the sense that

$$n^{1/2}(\check{\delta}_n - \delta_0) = n^{-1/2}\mathbf{T}'_n \mathbf{u}_n + o_p(1) \quad (16)$$

where \mathbf{T}_n is a nonstochastic $n \times k_\delta$ matrix, where k_δ is the dimension of δ_0 , and where $\mathbf{u}_n = (u_1, \dots, u_n)'$. Now define

$$\mathbf{q}_{*n}^\rho(\rho_0, \delta_0) = \mathbf{q}_n^\rho(\rho_0, \delta_0) + n^{-1}\mathbf{G}^{\rho\delta}\mathbf{T}'_n \mathbf{u}_n, \quad (17)$$

then (15) can be rewritten as

$$n^{1/2}(\widehat{\rho}_n - \rho_0) = -[(\mathbf{G}^{\rho\rho})'\mathbf{\Upsilon}^{\rho\rho}\mathbf{G}^{\rho\rho}]^{-1}(\mathbf{G}^{\rho\rho})'\mathbf{\Upsilon}^{\rho\rho} [n^{1/2}\mathbf{q}_{*n}^\rho(\rho_0, \delta_0)] + o_p(1). \quad (18)$$

Now suppose that

$$n^{1/2}\mathbf{q}_{*n}^\rho(\rho_0, \delta_0) \xrightarrow{d} N(0, \mathbf{\Psi}_*^{\rho\rho}), \quad (19)$$

where $\mathbf{\Psi}_*^{\rho\rho}$ is some positive definite matrix. Then

$$n^{1/2}(\widehat{\rho}_n - \rho_0) \xrightarrow{d} N[0, \mathbf{\Phi}_*^{\rho\rho}] \quad (20)$$

with

$$\mathbf{\Phi}_*^{\rho\rho} = [(\mathbf{G}^{\rho\rho})'\mathbf{\Upsilon}^{\rho\rho}\mathbf{G}^{\rho\rho}]^{-1}(\mathbf{G}^{\rho\rho})'\mathbf{\Upsilon}^{\rho\rho}\mathbf{\Psi}_*^{\rho\rho}\mathbf{\Upsilon}^{\rho\rho}\mathbf{G}^{\rho\rho}[(\mathbf{G}^{\rho\rho})'\mathbf{\Upsilon}^{\rho\rho}\mathbf{G}^{\rho\rho}]^{-1}.$$

From this it is seen that if we choose $\Upsilon_n^{\rho\rho} = \left(\tilde{\Psi}_{*n}^{\rho\rho}\right)^{-1}$ where $\Psi_*^{\rho\rho} = p \lim_{n \rightarrow \infty} \tilde{\Psi}_{*n}^{\rho\rho}$, then variance-covariance simplifies to

$$\Phi_*^{\rho\rho} = [(\mathbf{G}^{\rho\rho})'(\Psi_*^{\rho\rho})^{-1}\mathbf{G}^{\rho\rho}]^{-1}.$$

Therefore using for the weighting matrix $\Upsilon_n^{\rho\rho}$ a consistent estimator for the inverse of the limiting variance-covariance matrix $\Psi_*^{\rho\rho}$ yields the efficient two-step GMM estimator.

Suppose (13) holds and

$$\Psi = \begin{bmatrix} \Psi^{\rho\rho} & \Psi^{\rho\delta} \\ \Psi^{\delta\rho} & \Psi^{\delta\delta} \end{bmatrix},$$

then the limiting distribution of the sample moment vector $\mathbf{q}_n^\rho(\rho_0, \delta_0)$ evaluated at the true parameter values is given by

$$n^{1/2}\mathbf{q}_n^\rho(\rho_0, \delta_0) \xrightarrow{d} N(0, \Psi^{\rho\rho}). \quad (21)$$

It is important to note that in light of (17) in general $\Psi_*^{\rho\rho} \neq \Psi^{\rho\rho}$, unless $\mathbf{G}^{\rho\delta} = 0$, and that in general $\Psi_*^{\rho\rho}$ will depend on \mathbf{T}_n , which in turn will depend on the employed estimator $\check{\delta}_n$. In other words, unless $\mathbf{G}^{\rho\delta} = 0$, for a two-step GMM estimator we cannot simply use the variance-covariance matrix $\Psi^{\rho\rho}$ of the sample moment vector $\mathbf{q}_n^\rho(\rho_0, \delta_0)$, rather we need to work with the variance-covariance matrix $\Psi_*^{\rho\rho}$.

We next illustrate the difference between $\Psi^{\rho\rho} = (\psi_{rs}^{\rho\rho})$ and $\Psi_*^{\rho\rho} = (\psi_{*rs}^{\rho\rho})$ for the important special case where the moment conditions are quadratic and u_i is i.i.d. $N(0, \sigma^2)$. For simplicity assume that

$$\mathbf{q}_n^\rho(\rho_0, \delta_0) = n^{-1} \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n a_{ij1} u_i u_j \\ \sum_{i=1}^n \sum_{j=1}^n a_{ij1} u_i u_j \end{bmatrix}.$$

Now, for $r = 1, 2$, let a_{ir} denote the (i, r) -th element of $\mathbf{G}^{\rho\delta}\mathbf{T}'_n$, then by (17)

$$\mathbf{q}_{*n}^\rho(\rho_0, \delta_0) = n^{-1} \begin{bmatrix} \sum_{i=1}^n \sum_{j=1}^n a_{ij1} u_i u_j + \sum_{i=1}^n a_{i1} u_i \\ \sum_{i=1}^n \sum_{j=1}^n a_{ij2} u_i u_j + \sum_{i=1}^n a_{i2} u_i \end{bmatrix}.$$

It then follows from Theorem 1 that

$$\psi_{rs}^{\rho\rho} = 2\sigma^4 \sum_{i=1}^n \sum_{j=1}^n a_{ijr} a_{ijs},$$

but

$$\psi_{*rs}^{\rho\rho} = 2\sigma^4 \sum_{i=1}^n \sum_{j=1}^n a_{ijr}a_{ijs} + \sigma^2 \sum_{i=1}^n a_{ir}a_{is}.$$

We emphasize that the a_{ir} and a_{is} in the last sum on the r.h.s. for the expression for $\psi_{*rs}^{\rho\rho}$ depend on what estimator $\check{\delta}_n$ is employed in the sample moment vector $\mathbf{q}_n^\rho(\rho, \check{\delta}_n)$ used to form the objective function for the two-step GMM estimator $\hat{\rho}_n$ defined in (10). It is for this reason that in the literature on two-step GMM estimation users are often advised to follow a specific sequence of steps, to ensure the proper estimation of respective variance-covariance matrices.

3 GMM Estimation of Models with Spatial Lags

As remarked in the Introduction, arguably the most widely used class of spatial models are variants of the ones considered in Cliff and Ord (1973,1981), which build on the fundamental contribution of Whittle (1954). In these models, spatial interactions are modeled in terms of spatial lags. In particular, consider the following Cliff-Ord-type model relating a cross section of n spatial units:

$$\begin{aligned} \mathbf{y}_n &= \mathbf{X}_n \boldsymbol{\beta}_{0n} + \lambda_{0n} \mathbf{W}_n \mathbf{y}_n + \mathbf{u}_n \\ &= \mathbf{Z}_n \boldsymbol{\delta}_{0n} + \mathbf{u}_n \\ \mathbf{u}_n &= \rho_{0n} \mathbf{M}_n \mathbf{u}_n + \boldsymbol{\varepsilon}_n. \end{aligned} \tag{22}$$

where $\mathbf{Z}_n = [\mathbf{X}_n, \mathbf{W}_n \mathbf{y}_n]$, and $\boldsymbol{\delta}_{0n} = [\boldsymbol{\beta}'_{0n}, \lambda_{0n}]'$. Here $\mathbf{y}_n = (y_{1,n}, \dots, y_{n,n})'$ is the $n \times 1$ vector of the dependent variable, $\mathbf{X}_n = (x_{ik,n})$ is the $n \times K$ matrix of the non-stochastic exogenous regressors, $\mathbf{W}_n = (w_{ij,n})$ and $\mathbf{M}_n = (m_{ij,n})$ are $n \times n$ observed non-stochastic weights matrices, $\mathbf{u}_n = (u_{1,n}, \dots, u_{n,n})'$ is the $n \times 1$ vector of regression disturbances, and $\boldsymbol{\varepsilon}_n = (\varepsilon_{1,n}, \dots, \varepsilon_{n,n})'$ is an $n \times 1$ vector of innovations. The vectors $\bar{\mathbf{y}}_n = (\bar{y}_{1,n}, \dots, \bar{y}_{n,n})' = \mathbf{W}_n \mathbf{y}_n$ and $\bar{\mathbf{u}}_n = (\bar{u}_{1,n}, \dots, \bar{u}_{n,n})' = \mathbf{M}_n \mathbf{u}_n$ represent spatial lags, the scalars λ_{0n} and ρ_{0n} denote the corresponding true parameters, typically referred to as spatial-autoregressive parameters, and $\boldsymbol{\beta}_{0n}$ is a $k \times 1$ true parameter vector. In analogy to the time series literature, the above model is often referred to as a spatial autoregressive autoregressive (1,1) model, for short an *SARAR*(1,1) model.

In the above formulation, all data vectors and matrices, as well as all parameters are allowed to depend on the sample size n , i.e., to form triangular arrays. To see why this is necessary consider, e.g., the i -th elements of the spatial lag $\bar{\mathbf{y}}_n = \mathbf{W}_n \mathbf{y}_n$, which is given by

$$\bar{y}_{i,n} = \sum_{j=1}^n w_{ij,n} y_{j,n}.$$

From this it is obvious that even if the weights $w_{ij,n}$ do not depend on n , the weighted average $\bar{y}_{i,n}$ and thus $y_{i,n}$ will depend on n . In allowing for the elements of \mathbf{X}_n to depend on n we allow implicitly for some of the regressors to be spatial lags, e.g., the regressor matrix could be of the form $\mathbf{X}_n =$

$[\mathbf{x}_{1,n}, \mathbf{W}_n \mathbf{x}_{1,n}, \dots]$. In allowing for the elements of the spatial weights matrices to depend on n we allow implicitly for normalized spatial weights matrices, as is frequently the case in applications. In allowing also for the parameters to depend on n allows us to assume a common parameter space for all sample sizes; see Kelejian and Prucha (2010) for a more detailed discussion. For simplicity of notation we will, for the most part, drop again subscripts n in the following.

The spatial model (22) represents a system of n simultaneous equations. The reduced form of the model is given by

$$\mathbf{y} = (\mathbf{I} - \lambda_0 \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta}_0 + (\mathbf{I} - \lambda_0 \mathbf{W})^{-1} (\mathbf{I} - \rho_0 \mathbf{M})^{-1} \mathbf{u}. \quad (23)$$

If $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, then clearly $\mathbf{u} \sim N(\boldsymbol{\mu}_y, \boldsymbol{\Omega}_y)$ with

$$\begin{aligned} \boldsymbol{\mu}_y &= (\mathbf{I} - \lambda_0 \mathbf{W})^{-1} \mathbf{X} \boldsymbol{\beta}_0, \\ \boldsymbol{\Omega}_y &= \sigma^2 (\mathbf{I} - \lambda_0 \mathbf{W})^{-1} (\mathbf{I} - \rho_0 \mathbf{M})^{-1} (\mathbf{I} - \rho_0 \mathbf{M}')^{-1} (\mathbf{I} - \lambda_0 \mathbf{W}')^{-1}. \end{aligned}$$

From this we see that while it is easy to write down the log-likelihood function for model (22), the computation of the ML estimator is challenging or non-feasible for larger sample sizes n . The reason is that it requires the computation of the determinant of the $n \times n$ matrices $\mathbf{I} - \lambda_0 \mathbf{W}$ and $\mathbf{I} - \rho_0 \mathbf{M}$, which is taxing for large n unless the spatial weights matrices have structure that can be exploited.

Our discussions will also utilize the following spatial Cochrane-Orcutt transformation of (22):

$$\mathbf{y}_*(\rho_0) = \mathbf{Z}_*(\rho_0) \boldsymbol{\delta}_0 + \boldsymbol{\varepsilon}, \quad (24)$$

where $\mathbf{y}_*(\rho_0) = \mathbf{y} - \rho_0 \mathbf{M} \mathbf{y}$ and $\mathbf{Z}_*(\rho_0) = \mathbf{Z} - \rho_0 \mathbf{M} \mathbf{Z}$. The transformed model is readily obtained by pre-multiplying (22) by $\mathbf{I} - \rho_0 \mathbf{M}$.

3.1 GMM Estimation of Spatial Autoregressive Parameter

Motivated by the potential numerical problems in computing the ML estimator for larger sample sizes, Kelejian and Prucha (1998,1999) introduced an alternative GMM estimation approach which remains feasible even for large sample sizes and full spatial weights matrices.⁴ (See Pace chapter Maximum likelihood estimation). Another motivation was that at the time there

⁴Recall, e.g., that there are more than 33,000 zip codes in the U.S.

were no formal results available regarding the consistency and asymptotic normality of the ML estimator for the above model.

The GMM estimation approach put forward in Kelejian and Prucha (1998,1999) employs the following simple quadratic moment conditions, based on the assumption that the ε_i are i.i.d. $(0, \sigma^2)$:

$$En^{-1}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \sigma^2, \quad En^{-1}\bar{\boldsymbol{\varepsilon}}'\bar{\boldsymbol{\varepsilon}} = \sigma^2 n^{-1}tr(\mathbf{M}'\mathbf{M}), \quad En^{-1}\bar{\boldsymbol{\varepsilon}}'\boldsymbol{\varepsilon} = 0,$$

with $\bar{\boldsymbol{\varepsilon}} = \mathbf{M}\boldsymbol{\varepsilon}$. Substituting out σ^2 yields the following two quadratic moment conditions

$$En^{-1}\boldsymbol{\varepsilon}'\mathbf{A}_1\boldsymbol{\varepsilon} = 0, \quad En^{-1}\boldsymbol{\varepsilon}'\mathbf{A}_2\boldsymbol{\varepsilon} = 0, \quad (25)$$

with⁵

$$\mathbf{A}_1 = \mathbf{M}'\mathbf{M} - n^{-1}tr(\mathbf{M}'\mathbf{M})\mathbf{I}, \quad \mathbf{A}_2 = \mathbf{M}. \quad (26)$$

We note that for the weights matrices in (26) we have $tr(\mathbf{A}_q) = 0$ for $q = 1, 2$, but $diag(\mathbf{A}_1) \neq 0$. Kelejian and Prucha (2010) relax the assumption that the innovations are homoskedastic and allow for heteroskedasticity of unknown form. More specifically, they consider the case where the ε_i are i.i.d. $(0, \sigma_i^2)$ with σ_i^2 unknown.⁶ For this case they consider the following modified version of the above moment conditions where

$$\mathbf{A}_1 = \mathbf{M}'\mathbf{M} - n^{-1}diag(\mathbf{M}'\mathbf{M}), \quad \mathbf{A}_2 = \mathbf{M}. \quad (27)$$

Note that in this specification $diag(\mathbf{A}_q) = 0$ for $q = 1, 2$. Given this, the moment conditions in (25) continue to hold since $E\boldsymbol{\varepsilon}'\mathbf{A}_q\boldsymbol{\varepsilon} = \sum_{i=1}^n a_{q,ii}\sigma_i^2 = 0$. From this we see that, in general, moment conditions that employ weights matrices with $a_{q,ii} = 0$ and not just $tr(\mathbf{A}_q) = \sum_{i=1}^n a_{q,ii} = 0$ are robust against heteroskedasticity.

Of course, the above setup can be generalized to the case where we have S_ρ quadratic moment conditions ($q = 1, \dots, S_\rho$):

$$En^{-1}\boldsymbol{\varepsilon}'\mathbf{A}_q\boldsymbol{\varepsilon} = 0. \quad (28)$$

⁵To obtain the estimator of Kelejian and Prucha (1998,1999) the matrix \mathbf{A}_1 has to be scaled by $v = 1/[1 + [n^{-1}tr(\mathbf{M}'\mathbf{M})]^2]$. Of course, the scaling factor only comes into play if the moment conditions are not optimally weighted, as was the case in the early literature.

⁶Lin and Lee (2010) also allow for heteroskedastic innovations for model (22) with $\rho_0 = 0$.

In light of (22) those moment conditions can be written equivalently as ($q = 1, \dots, S_\rho$):

$$En^{-1}\mathbf{u}'(\mathbf{I} - \rho_0\mathbf{M}')\mathbf{A}_q(\mathbf{I} - \rho_0\mathbf{M})\mathbf{u} = 0. \quad (29)$$

Now let $\check{\boldsymbol{\delta}}$ be some initial estimator for $\boldsymbol{\delta}_0$ and let $\check{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\check{\boldsymbol{\delta}}$. Then we can formulate the following corresponding sample moment vector:

$$\mathbf{q}_n^\rho(\rho, \check{\boldsymbol{\delta}}) = n^{-1} \begin{bmatrix} \check{\mathbf{u}}'(\mathbf{I} - \rho\mathbf{M}')\mathbf{A}_1(\mathbf{I} - \rho\mathbf{M})\check{\mathbf{u}} \\ \vdots \\ \check{\mathbf{u}}'(\mathbf{I} - \rho\mathbf{M}')\mathbf{A}_{S_\rho}(\mathbf{I} - \rho\mathbf{M})\check{\mathbf{u}} \end{bmatrix}. \quad (30)$$

Furthermore, as in (10), the class of corresponding two-step GMM estimators is then given by

$$\hat{\rho} = \arg \min_{\rho} \{ \mathbf{q}_n^\rho(\rho, \check{\boldsymbol{\delta}})' \boldsymbol{\Upsilon}_n^{\rho\rho} \mathbf{q}_n^\rho(\rho, \check{\boldsymbol{\delta}}) \}. \quad (31)$$

where $\boldsymbol{\Upsilon}_n^{\rho\rho}$ is a weighting matrix. As discussed in Section 2.3, the efficient choice for $\boldsymbol{\Upsilon}_n^{\rho\rho}$ will generally depend on the estimator $\check{\boldsymbol{\delta}}$ employed in the estimation of the disturbances.

3.2 GMM Estimation of Regression Parameters

In order to motivate the GMM estimator for the regression parameters $\boldsymbol{\delta}_0$ we note that the best instruments for the r.h.s. variables of model (22) and (24) are the conditional means. Since \mathbf{X} and \mathbf{MX} are non-stochastic (and their own best instruments) we can focus on the spatial lags $\mathbf{W}\mathbf{y}$ and $\mathbf{MW}\mathbf{y}$. The best instruments are given $\mathbf{W}\mathbf{E}\mathbf{y}$ and $\mathbf{MW}\mathbf{E}\mathbf{y}$ with

$$\mathbf{E}\mathbf{y} = (\mathbf{I} - \lambda_0\mathbf{W})^{-1}\mathbf{X}\boldsymbol{\beta}_0 = \sum_{l=1}^{\infty} \lambda_0^l \mathbf{W}^l \mathbf{X}\boldsymbol{\beta}_0, \quad (32)$$

given that spectral radius of $\lambda_0\mathbf{W}$ is less than one. To avoid issues associated with the computation of the inverse of the $n \times n$ matrix of $\mathbf{I} - \lambda_0\mathbf{W}$, Kelejian and Prucha (1998,1999) suggest the use of an approximation of the best instruments. More specifically, in light of the last expression in (32) they suggest using a set of instruments \mathbf{H} which contains, say, \mathbf{X} , \mathbf{MX} , \mathbf{MWX} , \dots , $\mathbf{MW}^p\mathbf{X}$, and to compute approximators of the best instruments from a regression of the r.h.s. variables against \mathbf{H} .

For the untransformed model this is equivalent to considering the moment condition $En^{-1}\mathbf{H}'\mathbf{u} = 0$. Of course, the corresponding GMM estimator is just the two-stage least squares (2SLS) estimator. For the transformed model (24) the moment condition would be

$$En^{-1/2}\mathbf{H}'\boldsymbol{\varepsilon} = 0. \quad (33)$$

Now let $\check{\rho}$ be some estimator for ρ_0 , then we can formulate the following corresponding sample moment vector:

$$\mathbf{q}^\delta(\check{\rho}, \boldsymbol{\delta}) = n^{-1/2}\mathbf{H}'[\mathbf{y}_*(\check{\rho}) - \mathbf{Z}_*(\check{\rho})\boldsymbol{\delta}]. \quad (34)$$

Under homoskedasticity the variance-covariance matrix of the moment vector $\mathbf{q}^\delta(\rho_0, \boldsymbol{\delta}_0) = n^{-1/2}\mathbf{H}'\boldsymbol{\varepsilon}$ is given by $\sigma^2 n^{-1}\mathbf{H}'\mathbf{H}$, which motivates the following two-step GMM estimators for $\boldsymbol{\delta}_0$:

$$\hat{\boldsymbol{\delta}} = \arg \min_{\boldsymbol{\delta}} \{ \mathbf{q}_n^\delta(\check{\rho}, \boldsymbol{\delta})' \boldsymbol{\Upsilon}_n^{\delta\delta} \mathbf{q}_n^\delta(\check{\rho}, \boldsymbol{\delta}) \} \quad (35)$$

with $\boldsymbol{\Upsilon}_n^{\delta\delta} = [n^{-1}\mathbf{H}'\mathbf{H}]^{-1}$. By observing that the quadratic form on the r.h.s. of (35) is just $[\mathbf{y}_*(\check{\rho}) - \mathbf{Z}_*(\check{\rho})\boldsymbol{\delta}]' \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'[\mathbf{y}_*(\check{\rho}) - \mathbf{Z}_*(\check{\rho})\boldsymbol{\delta}]$, apart from some scaling factors, we see that the estimator defined by (35) is just the 2SLS estimator applied to the transformed model (24) with ρ_0 replaced by $\check{\rho}$, i.e.

$$\hat{\boldsymbol{\delta}} = [\hat{\mathbf{Z}}_*(\check{\rho})'\hat{\mathbf{Z}}_*(\check{\rho})]^{-1}\hat{\mathbf{Z}}_*(\check{\rho})'\mathbf{y}_*(\check{\rho}) \quad (36)$$

where $\hat{\mathbf{Z}}_*(\check{\rho}) = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{Z}_*(\check{\rho})$. This estimator has been called the feasible generalized spatial two-stage least squares (FGS2SLS) estimator.

3.3 Guide to Literature

The above sections discussed basic ideas concerning moment conditions that can be exploited by GMM estimators for spatial Cliff-Ord-type models. Since the late 1990s, a considerable body of literature has developed regarding the GMM estimation of Cliff-Ord-type models. In the following we provide references to some of that literature. Naturally, given space limitations, the list of references is incomplete. Also, the list will focus on theoretical contributions and will not cover corresponding empirical work.⁷

⁷For an incomplete list of empirical work see, e.g., Kelejian and Prucha (2010).

By employing an approximation of the best instruments, the FGS2SLS estimator of Kelejian and Prucha (1998,1999) has the advantage of remaining computational feasible even for very large sample sizes since its formulation does not involve the computation of the inverse of the $n \times n$ matrix $\mathbf{I} - \lambda_0 \mathbf{W}$. However, as a result it is not fully efficient. Lee (2003) introduces a Best 2SLS estimator. This estimator uses the first expression for $E\mathbf{y}$ in (32) in forming best instruments for $\mathbf{W}\mathbf{y}$. It is best in the sense that its asymptotic variance-covariance matrix is smallest among the class of GMM estimators based on linear moment conditions. Kelejian, Prucha, and Yuzefovich (2004) introduce an alternative Best 2SLS estimator (with identical asymptotic properties). For computational ease this estimator uses a series approximation for the second expression for $E\mathbf{y}$ in (32) when forming best instruments for $\mathbf{W}\mathbf{y}$. All of the above S2SLS estimators break down if $\beta_0 = 0$, i.e., if there are no exogenous variables in the model. This is not the case with the ML estimator. As a consequence, one would expect the ML estimator, given it is computable, to increasingly outperform the above S2SLS estimators as the variation in the disturbances increases relative to the variation in the regressors. However, Das, Kelejian and Prucha (2003) provide Monte Carlo results which suggest that the loss of efficiency of 2SLS type estimators relative to ML estimation is modest for a wide range of specifications.

The above papers establish consistency of the GMM estimator for ρ_0 , but do not derive its limiting distribution. Drukker, Egger and Prucha (2011) derive the joint limiting distribution for two-step GMM estimators for δ_0 and ρ_0 .

Fingleton (2008) formulates moment conditions and GMM estimators for the case where the disturbances process is an MA rather than an AR process.

Lee (2007) considers an SARAR(1,0) model, i.e., model (22) with $\rho_0 = 0$. He suggests augmenting the usual linear moment conditions by quadratic moment conditions, and derives the best quadratic moment condition. This best quadratic moment condition involves the inverse of $\mathbf{I} - \lambda_0 \mathbf{W}$. Lee shows that the corresponding Best GMM estimator may have the same asymptotic distribution as the ML estimator under normality. Also, the estimator does not break down if there are no explanatory exogenous variables. Liu, Lee and Bollinger (2010) and Lee and Liu (2010) extend the results to one-step GMM estimators of an SARAR(1,1) and SARAR(p,q) model, respectively.

All of the above literature assumes that the basic innovations are homoskedastic. Kelejian and Prucha (2010) consider two-step GMM estimation of an SARAR(1,1) model under the assumption that the innovations

are heteroskedastic of unknown form. Badinger and Egger (2011) extend the approach the case of an SARAR(p,q) model. Lin and Lee (2010) consider one-step GMM estimation of an SARAR(1,0) with unknown heteroskedasticity, employing both linear and quadratic moment conditions.

Extensions of Cliff-Ord-type models to random and fixed effects panel data have been an important focus of recent research. Considered estimation methodologies have been GMM, quasi-ML and Bayesian Markov Chain Monte Carlo methods.⁸ The literature on GMM estimation for panel data includes Kapoor, Kelejian and Prucha (2007), Mutl and Pfaffermayr (2011), and Yu, de Jong and Lee (2012). Liu and Lee (2010) discuss GMM estimation (as well as other approaches) of a Cliff-Ord-type social interaction model. (See Elhorst chapter on Panel data models).

Kelejian and Prucha (2007), and Drukker, Egger and Prucha (2011) discuss GMM estimation for Cliff-Ord-type single equation models with additional outside endogenous variables. Kelejian and Prucha (2004) consider a Cliff-Ord-type simultaneous equation system and discuss both limited and full information GMM estimators.

Pinkse, Slade and Brett (2002) consider a semiparametric GMM approach, which allows for the spatial weights to be modeled as unknown functions of some distance measure. We note that if we are willing to assume that the weights can be expressed as, say, a finite polynomial in distance, then the substituted model will be of the form of an SARAR(p,q) model.

3.4 Exemplary GMM Estimators

In the following we give an illustrative result for the limiting distribution of GMM estimators for the SARAR(1,1) model (22). As remarked, for two-step GMM estimation the limiting distribution of the GMM estimator for ρ_0 will depend on the estimator for δ_0 used in constructing estimated residuals. Our illustrative example will focus on the two-step GMM estimators considered in Kelejian and Prucha (1998,1999), which can be viewed as a special case of the GMM estimators considered in Sections 3.1 and 3.2, with $S_\rho = 2$ and $\mathbf{A}_1 = v [\mathbf{M}'\mathbf{M} - n^{-1}tr(\mathbf{M}'\mathbf{M})\mathbf{I}]$ with $v = 1 / \left[1 + [n^{-1}tr(\mathbf{M}'\mathbf{M})]^2 \right]$, and

⁸Quasi-ML and Bayesian MCMC methods are not covered by this review. For recent papers employing those methods within the context of dynamic panel data models, see, e.g., Yu, de Jong and Lee (2008) and Parent and LeSage (2012), respectively. There is also an important literature on testing for spatial dependence in a panel context, which is not part of this review. For a partial review of this literature see, e.g., Baltagi (2011).

$\mathbf{A}_2 = \mathbf{M}$. The discussion below assumes that the assumptions maintained in that paper hold, including that the innovations ε_i are i.i.d. $(0, \sigma^2)$.

We next describe specific steps in computing the GMM estimators.

Step 1a: 2SLS Estimator

In the first step, estimate $\boldsymbol{\delta}$ by 2SLS from the untransformed model (22), using the instrument matrix \mathbf{H} as discussed in Section 3.2. The 2SLS estimator, say $\tilde{\boldsymbol{\delta}}$, is then given by $\tilde{\boldsymbol{\delta}} = (\tilde{\mathbf{Z}}'\tilde{\mathbf{Z}})^{-1}\tilde{\mathbf{Z}}'\mathbf{y}$, where $\tilde{\mathbf{Z}} = \mathbf{P}_\mathbf{H}\mathbf{Z}$ with $\mathbf{P}_\mathbf{H} = \mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$.

Step 1b: Initial GMM Estimator of ρ Based on 2SLS Residuals

Let $\tilde{\mathbf{u}} = \mathbf{u}(\tilde{\boldsymbol{\delta}}) = \mathbf{y} - \mathbf{Z}\tilde{\boldsymbol{\delta}}$ denote the 2SLS residuals. Consider the following sample moments based on estimated 2SLS residuals:

$$\mathbf{q}_n^\rho(\rho, \tilde{\boldsymbol{\delta}}) = n^{-1} \begin{bmatrix} \tilde{\mathbf{u}}'(\mathbf{I} - \rho\mathbf{M}')\mathbf{A}_1(\mathbf{I} - \rho\mathbf{M})\tilde{\mathbf{u}} \\ \tilde{\mathbf{u}}'(\mathbf{I} - \rho\mathbf{M}')\mathbf{A}_2(\mathbf{I} - \rho\mathbf{M})\tilde{\mathbf{u}} \end{bmatrix}. \quad (37)$$

The initial GMM estimator for ρ is then defined as

$$\tilde{\rho} = \arg \min_{\rho} \left\{ \mathbf{q}_n^\rho(\rho, \tilde{\boldsymbol{\delta}})' \mathbf{q}_n^\rho(\rho, \tilde{\boldsymbol{\delta}}) \right\}.$$

Clearly $\tilde{\rho}$ is a special case of the class of estimators considered in (31) with $\Upsilon_n^{\rho\rho} = \mathbf{I}$.

Step 2a: FGS2SLS Estimator

In the second step re-estimate $\boldsymbol{\delta}$ by FGS2SLS, as discussed in Section 3.2. The FGS2SLS estimator is defined as the 2SLS estimator of the Cochrane-Orcutt transformed model (24) with the parameter ρ_0 replaced by $\tilde{\rho}$ computed in Step 1b. The FGS2SLS estimator is given by $\hat{\boldsymbol{\delta}} = [\hat{\mathbf{Z}}_*(\tilde{\rho})'\hat{\mathbf{Z}}_*(\tilde{\rho})]^{-1}\hat{\mathbf{Z}}_*(\tilde{\rho})'\mathbf{y}_*(\tilde{\rho})$ where $\hat{\mathbf{Z}}_*(\tilde{\rho}) = \mathbf{P}_\mathbf{H}\mathbf{Z}_*(\tilde{\rho})$.

Step 2b: Efficient GMM Estimator of ρ Based on FGS2SLS Residuals

Let $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\hat{\boldsymbol{\delta}}$ denote the FGS2SLS residuals, and let $\mathbf{q}_n^\rho(\rho, \hat{\boldsymbol{\delta}})$ be defined as in (37) with $\tilde{\mathbf{u}}$ replaced by $\hat{\mathbf{u}}$. By Drukker, Egger and Prucha (2011) the corresponding efficient GMM estimator for ρ_0 based on FGS2SLS residuals is then given by

$$\hat{\rho} = \arg \min_{\rho} \left[\mathbf{q}_n^\rho(\rho, \hat{\boldsymbol{\delta}})' \left(\hat{\Psi}_n^{\rho\rho} \right)^{-1} \mathbf{q}_n^\rho(\rho, \hat{\boldsymbol{\delta}}) \right],$$

where $\widehat{\Psi}_n^{\rho\rho} = (\widehat{\psi}_{rs}^{\rho\rho})_{r,s=1,2}$ is an estimator of the variance-covariance matrix of the limiting distribution of the normalized sample moments $n^{1/2}\mathbf{q}_n^\rho(\rho, \widehat{\boldsymbol{\delta}})$. In particular we have⁹

$$\begin{aligned}\widehat{\psi}_{rs}^{\rho\rho} &= \widehat{\sigma}^4(2n)^{-1}tr[(\mathbf{A}_r + \mathbf{A}'_r)(\mathbf{A}_s + \mathbf{A}'_s)] \\ &\quad + \widehat{\sigma}^2 n^{-1} \widehat{\mathbf{a}}'_r \widehat{\mathbf{a}}_s \\ &\quad + n^{-1}(\widehat{\mu}^{(4)} - 3\widehat{\sigma}^4)vec_D(\mathbf{A}_r)'vec_D(\mathbf{A}_s) \\ &\quad + n^{-1}\widehat{\mu}^{(3)}[\widehat{\mathbf{a}}'_r vec_D(\mathbf{A}_s) + \widehat{\mathbf{a}}'_s vec_D(\mathbf{A}_r)],\end{aligned}$$

where $\widehat{\mathbf{a}}_r = \mathbf{H}\widehat{\mathbf{P}}^*\widehat{\boldsymbol{\alpha}}_r$ and

$$\begin{aligned}\widehat{\mathbf{P}}^* &= (n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z}'_*(\widetilde{\rho})) \times \\ &\quad [(n^{-1}\mathbf{Z}'_*(\widetilde{\rho})\mathbf{H})(n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z}'_*(\widetilde{\rho}))]^{-1}, \\ \widehat{\boldsymbol{\alpha}}_r &= -n^{-1}\mathbf{Z}'_*(\widetilde{\rho})(\mathbf{A}_r + \mathbf{A}'_r)\widehat{\boldsymbol{\varepsilon}},\end{aligned}$$

and $\widehat{\sigma}_n^2$, $\widehat{\mu}_n^{(3)}$ and $\widehat{\mu}_n^{(4)}$ are standard sample estimators of σ^2 , $\mu^{(3)} = E\varepsilon_i^3$, $\mu^{(4)} = E\varepsilon_i^4$ based on $\widehat{\boldsymbol{\varepsilon}} = (\mathbf{I} - \widetilde{\rho}\mathbf{M})\widehat{\mathbf{u}}$.

The derivation of the limiting distribution of $n^{1/2}\mathbf{q}^\rho(\rho, \widehat{\boldsymbol{\delta}})$ used the CLT for linear quadratic forms given as Theorem 1. Observe that $\widehat{\boldsymbol{\alpha}}_r$ is an estimator for $\boldsymbol{\alpha}_r = -n^{-1}E\mathbf{Z}'_*(\rho)(\mathbf{A}_r + \mathbf{A}'_r)\boldsymbol{\varepsilon}$. If the model does not contain a spatial lag in \mathbf{y} , i.e., if $\mathbf{Z} = \mathbf{X}$, then $\boldsymbol{\alpha}_r = 0$ and we can then take $\widehat{\boldsymbol{\alpha}}_r = 0$.

Based on Drukker, Egger and Prucha (2011) we now have the following result for the joint asymptotic distribution of the final stage estimators $\widehat{\boldsymbol{\delta}}$ and $\widehat{\rho}$:

$$\begin{bmatrix} \widehat{\boldsymbol{\delta}} \\ \widehat{\rho} \end{bmatrix} \sim N \left(\begin{bmatrix} \boldsymbol{\delta}_0 \\ \rho_0 \end{bmatrix}, n^{-1} \begin{bmatrix} \widehat{\Omega}^{\delta\delta} & \widehat{\Omega}^{\delta\rho} \\ \widehat{\Omega}^{\delta\rho'} & \widehat{\Omega}^{\rho\rho} \end{bmatrix} \right),$$

⁹In the following $vec_D(\mathbf{A})$ refers to the column vector containing the diagonal elements of the matrix \mathbf{A} .

where

$$\begin{aligned}
\widehat{\Omega}^{\delta\delta} &= \widehat{\mathbf{P}}^{*'} \widehat{\Psi}^{\delta\delta} \widehat{\mathbf{P}}^*, \\
\widehat{\Omega}^{\delta\rho} &= \widehat{\mathbf{P}}^{*'} \widehat{\Psi}^{\delta\rho} \left(\widehat{\Psi}^{\rho\rho} \right)^{-1} \widehat{\mathbf{J}} \left[\widehat{\mathbf{J}}' \left(\widehat{\Psi}^{\rho\rho} \right)^{-1} \widehat{\mathbf{J}} \right]^{-1}, \\
\widehat{\Omega}^{\rho\rho} &= \left[\widehat{\mathbf{J}}' \left(\widehat{\Psi}^{\rho\rho} \right)^{-1} \widehat{\mathbf{J}} \right]^{-1}, \\
\widehat{\Psi}^{\delta\delta} &= \widehat{\sigma}^2 n^{-1} \mathbf{H}' \mathbf{H}, \\
\widehat{\Psi}^{\delta\rho} &= \widehat{\sigma}^2 n^{-1} \mathbf{H}' [\widehat{\mathbf{a}}_1, \widehat{\mathbf{a}}_2] + \widehat{\mu}^{(3)} n^{-1} \mathbf{H}' [vec_D(\mathbf{A}_1), vec_D(\mathbf{A}_{2,n})],
\end{aligned}$$

where $\widehat{\mathbf{P}}^*$, $\widehat{\mathbf{a}}_r$, $\widehat{\Psi}^{\rho\rho}$ are as defined above, and

$$\widehat{\mathbf{J}} = n^{-1} \begin{bmatrix} 2\widehat{\mathbf{u}}' \mathbf{M}' \mathbf{A}_1 \widehat{\mathbf{u}} & -\widehat{\mathbf{u}}' \mathbf{M}' \mathbf{A}_1 \mathbf{M} \widehat{\mathbf{u}} \\ 2\widehat{\mathbf{u}}' \mathbf{M}' \mathbf{A}_2 \widehat{\mathbf{u}} & -\widehat{\mathbf{u}}' \mathbf{M}' \mathbf{A}_2 \mathbf{M} \widehat{\mathbf{u}} \end{bmatrix} \begin{bmatrix} 1 \\ 2\widehat{\rho} \end{bmatrix}.$$

For interpretation, observe that $\widehat{\Omega}^{\delta\delta} = \widehat{\sigma}^2 [\widehat{\mathbf{Z}}_*(\widehat{\rho})' \widehat{\mathbf{Z}}_*(\widehat{\rho})]^{-1}$, i.e., the above expression for the estimator of variance-covariance matrix of the joint distribution of $\widehat{\boldsymbol{\delta}}$ and $\widehat{\rho}$ delivers the usual estimator for the variance-covariance matrix of the FGS2SLS estimator as a special case.

The above joint-asymptotic-normality result allows a joint Wald test for the absence of spatial dependencies, i.e., a joint test of $H_0 : \lambda_0 = 0, \rho_0 = 0$.

4 GMM Estimation of Models with Spatial Mixing

Cliff-Ord-type models are linear simultaneous equation models where mixing of the data process is achieved through the assumption that the basic innovations are independently distributed combined with assumptions on the spatial weights matrices, such as that the row and column sums of their absolute elements are bounded.

In the time series literature a widely used notion of dependence is α -mixing. This concept has been generalized to spatial processes (or random fields). In an important paper Conley (1999) considered GMM estimators for stationary α -mixing spatial processes. Stationarity implies that the process has constant mean and variance, and that the covariances only depend on

distance (in a particular direction). Many economic processes are likely to exhibit some form of non-stationarity - e.g., housing prices may increase very much as we move towards the center of a city. Thus relaxing the stationarity assumption seemed important.

One difficulty in developing a generalized theory of inference for spatial processes was a paucity of limit theorems (laws of large numbers, uniform laws of large numbers and central limit theorems) which are sufficiently general. In light of this, Jenish and Prucha (2009) developed limit theorems for non-stationary α -mixing spatial processes, allowing also for the locations of observations to form a non-regular grid. Still, since α -mixing is not necessarily preserved under infinite lag formations, a further expansion of the theory to a class of spatial processes, which is closed under infinite lag formations, seemed desirable. To that effect Jenish and Prucha (2012) extended the notion of near-epoch dependence from the time series literature to spatial processes. They then developed limit theorems for possibly non-stationary spatial processes which are near-epoch dependent on an α -mixing process, and gave results concerning the consistency and asymptotic normality of GMM estimators for this generalized class of processes.

In a recent publication, Robinson and Thawornkaiwong (2012) consider a partially linear regression model. They define a semiparametric instrumental variable estimator and give results on its asymptotic properties, allowing for spatial dependence in the regressors and disturbances.

5 Conclusion

Over the last two decades significant strides have been made towards developing a formal methodology of inference for spatial models, or more generally, for cross sectional interaction models. GMM estimation has been an important part of this literature. As usual, empirical work often confronts us with more challenging realities than what can be handled by existing methodologies of inference, and much more work is needed.

References

- [1] Anselin L (1988) Spatial econometrics: Methods and models. Kluwer Academic Publishers, Boston
- [2] Anselin L (2010) Thirty years of spatial econometrics. *Papers in Regional Science* 89(1): 3-25
- [3] Arbia G (2006) Spatial econometrics, Statistical foundations and applications to regional convergence. Springer, New York
- [4] Arraiz I, Drukker DM, Kelejian HH, Prucha IR (2010) A spatial Cliff-Ord-type model with heteroskedastic innovations: Small and large sample results. *Journal of Regional Science* 50(2): 592-614
- [5] Badinger H, Egger P (2011) Estimation of higher-order spatial autoregressive cross-section models with heteroscedastic disturbances. *Papers in Regional Science* 90(1), 213-235
- [6] Baltagi BH (2011) Spatial panels. In: Ullah A, Giles DEA (eds) *The handbook of empirical economics and finance*, Chapman and Hall, Boca Raton, pp 435-454
- [7] Cliff A, Ord J (1973) Spatial autocorrelation. Pion, London
- [8] Cliff A, Ord J (1981) Spatial processes, Models and applications. Pion, London
- [9] Conley T (1999) GMM estimation with cross sectional dependence, *Journal of Econometrics* 92, 1-45
- [10] Cressie N (1993) *Statistics of spatial data*. Wiley, New York
- [11] Das D, Kelejian HH, Prucha IR (2003) Small sample properties of estimators of spatial autoregressive models with autoregressive disturbances. *Papers in Regional Science* 82(1): 1-26
- [12] Drukker DM, Egger P, Prucha IR (2011) On two-step estimation of a spatial autoregressive model with autoregressive disturbances and endogenous regressors. Forthcoming in *Econometric Reviews*

- [13] Haining R (2003) Spatial data analysis, Theory and practice. Cambridge University Press, Cambridge
- [14] Jenish N, Prucha IR (2009) Central limit theorems and uniform laws of large numbers for arrays of random fields. *Journal of Econometrics* 150(1): 86-98
- [15] Jenish, N, Prucha IR (2012) On spatial processes and asymptotic inference under near-epoch dependence. Department of Economics, University of Maryland, mimeo
- [16] Kapoor M, Kelejian HH, Prucha IR (2007) Panel data models with spatially correlated error components. *Journal of Econometrics* 140(1): 97-130
- [17] Kelejian HH, Prucha IR (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(1): 99-121
- [18] Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40(2): 509-533
- [19] Kelejian HH, Prucha IR (2001) On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104(2): 219-257
- [20] Kelejian HH, Prucha IR (2004) Estimation of simultaneous systems of spatially interrelated cross sectional equations. *Journal of Econometrics* 118(1-2): 27-50
- [21] Kelejian HH, Prucha IR (2007) HAC estimation in a spatial framework. *Journal of Econometrics* 140(1): 131-154
- [22] Kelejian HH, Prucha IR (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* 157(1): 53-67

- [23] Kelejian HH, Prucha IR, Yuzevovich E (2004) Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. In: LeSage JP, Pace PR (eds) *Advances in econometrics: Spatial and spatiotemporal econometrics*, Elsevier, New York, pp. 163-198
- [24] Lee L-F (2003) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. *Econometric Reviews* 22(4): 307-335
- [25] Lee L-F (2004) Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72(6): 1899-1925
- [26] Lee L-F (2007) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137(2): 489-514
- [27] Lee L-F, Liu X (2010) Efficient GMM estimation of higher order spatial autoregressive models with autoregressive disturbances. *Econometric Theory* 26(1): 187-230
- [28] LeSage JP, Pace RK (2009) *Introduction to spatial econometrics*. CRC Press Taylor and Francis Group, Boca Raton [FL]
- [29] Lin X, Lee L-F (2010) GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *Journal of Econometrics* 157(1): 34-52
- [30] Liu X, Lee L-F (2010) GMM estimation of social interaction models with centrality. *Journal of Econometrics* 159(1), 99-115
- [31] Liu X, Lee L-F, Bollinger CR (2010) An efficient GMM estimator of spatial autoregressive models. *Journal of Econometrics* 159(2), 303-319
- [32] Mutl J, Pfaffermayr M (2011) The Hausman test in a Cliff and Ord panel model. *Econometrics Journal* 14(1): 48-76
- [33] Paelinck JHP, Klaassen LH (1979) *Spatial econometrics*. Saxon House, Farnborough
- [34] Parent O, LeSage JP (2012) Spatial dynamic panel data models with random effects. Forthcoming in *Regional Science and Urban Economics*

- [35] Pinkse J, Slade ME, Brett C (2002) Spatial price competition: A semi-parametric approach. *Econometrica* 70(3), 1111-1153
- [36] Robinson PM, Thawornkaiwong S (2012) Statistical inference on regression with spatial dependence. *Journal of Econometrics* 167(2): 521-542
- [37] Tobler W (1970) A computer movie simulating urban growth in the Detroit region. *Economic Geography* 46(2): 234-240
- [38] Whittle, P (1954) On stationary processes in the plane. *Biometrika* 41(3.4): 434-449
- [39] Yu J, de Jong R, Lee L-F (2008) Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* 146(1), 118-134
- [40] Yu J, de Jong R, Lee L-F (2012) Estimation for spatial dynamic panel data with fixed effects: The case of spatial cointegration. *Journal of Econometrics* 167(1), 16-37