GMM estimation of Spatial Panels with Fixed Effects

Francesco Moscone and Elisa Tosetti

Brunel University, University of Cambridge - Faculty of Economics

19. January 2010

GMM estimation of Spatial Panels with Fixed Effects*

F. Moscone
Brunel University

E. Tosetti
University of Cambridge

January 19, 2010

Abstract

In this paper we consider the estimation of a panel data regression model with spatial autoregressive disturbances, fixed effects and unknown heteroskedasticity. Following the work by Kelejian and Prucha (1999), Lee and Liu (2006a) and others, we adopt the Generalized Method of Moments (GMM) and consider as moments a set linear quadratic conditions in the disturbances. As in Lee and Liu (2006a), we assume that the inner matrices in the quadratic forms have zero diagonal elements to robustify moments against unknown heteroskedasticity. We derive the asymptotic distribution of the GMM estimator based on such conditions. Hence, we carry out some Monte Carlo experiment to investigate the small sample properties of GMM estimators based on various sets of moment conditions.

1 Introduction

GMM estimation of spatial regression models in a single cross sectional setting has been originally advanced by Kelejian and Prucha (1999). They focused on a regression equation with spatial autoregressive (SAR) disturbances, and suggested the use of three moment conditions that exploit the properties of disturbances implied by a standard set of assumptions. Estimation consists of solving a non-linear optimization problem, which yields a consistent estimator under a number of regularity conditions.

Recently, considerable work has been carried out to extend the procedure advanced by Kelejian and Prucha in various directions. Liu, Lee, and Bollinger (2006) and Lee and Liu (2006a) suggested a set of moments that encompass Kelejian and Prucha conditions as special cases. They considered a vector of linear and quadratic conditions in the error term, where the matrices appearing in the linear and quadratic forms have bounded row and column norms (see also Lee (2007)). Hence, they focused on the problem of selecting the matrices appearing in the vector of linear and quadratic moment conditions, in order to obtain the lowest variance for the GMM estimator. Lin and Lee (2005) also showed that these moments can be made robust against unknown heteroskedasticity by imposing that the diagonal elements of inner matrices are zero.

*The authors acknowledge financial support from ESRC (Ref. no. RES-061-25-0317).

Lee and Liu (2006b) have extended this framework to estimate the SAR model with higher-order spatial lags. Kelejian and Prucha (2008) have generalized their original work to include spatial lags in the dependent variable as well as allowing for heteroskedastic disturbances. This setting has been extended by Kapoor, Kelejian, and Prucha (2007) to estimate a spatial panel regression model with individual-specific error components. Druska and Horrace (2004) have introduced the Keleijan and Prucha GMM within the framework of a panel with SAR disturbances, time dummies and time-varying spatial weights, while Fingleton (2008b) and Fingleton (2008a) have extended it to the case of a regression model with spatial moving average disturbances.

In this paper, we focus on the GMM estimation of a panel data regression model with fixed effects, unknown heteroskedasticity, and spatial autoregressive (SAR) errors. Quasi-maximum likelihood (ML) estimation of a panel with fixed effects and spatial lags both in the dependent variable and in the disturbances, under homoskedastic errors, has been developed by Lee and Yu (2008). The authors propose a transformation approach to eliminate the fixed effects that yields consistent estimators for regression parameters when either N or T are large. Yu, de Jong, and Lee (2008) and Yu, de Jong, and Lee (2007) have investigated the properties of the quasi-ML estimator of a spatial dynamic panel with fixed effects, possibly non-stationarity. Mutl and Pfaffermayr (2008) consider GMM estimation of fixed effects vs random effects spatial panel specifications. Hence, they propose a spatial Hausman test that which compares the two models, accounting for spatial autocorrelation in the disturbances.

Following the work by Lee and Liu (2006a) and Kelejian and Prucha (1999), in this paper we adopt the GMM and consider as moments a set linear quadratic conditions in the disturbances. To eliminate the individual effects, we transform data by applying the demeaning operator. As in Lee and Liu (2006a), we assume that the inner matrices in the quadratic forms have zero diagonal elements to robustify moments against unknown heteroskedasticity. We show that consistency and asymptotic normality of the parameters of the SAR process is achieved for N and T going to infinity, with no restrictions on the relative rate at which N and T increase. We then perform a Monte Carlo exercise to compare the small sample properties of GMM estimators based on different sets of moment conditions.

In the following, Section 2 sets out the framework of a regression model with SAR disturbances; Section 3 introduces the GMM estimator; Section 4 carries a small Monte Carlo exercise; Section 5 concludes.

2 The framework

Consider the panel data regression model

$$y_{it} = \alpha_i + \boldsymbol{\beta}'\mathbf{x}_{it} + u_{it}, \quad i = 1, ..., N, t = 1, ..., T \tag{1}$$

where α_i are fixed parameters, and errors are assumed to follow the SAR process

$$u_{it} = \delta \sum_{j=1}^{N} s_{ij} u_{jt} + \varepsilon_{it} \tag{2}$$

and s_{ij} is the $(i,j)^{th}$ element of an $N \times N$ spatial weights matrix, $\mathbf{S}$. In matrix form,

$$\mathbf{y} = (\mathbf{1}_T \otimes \boldsymbol{\alpha}) + \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{3}$$

$$\mathbf{u} = \delta (\mathbf{I}_T \otimes \mathbf{S}) \mathbf{u} + \boldsymbol{\varepsilon}, \tag{4}$$

where $\mathbf{y} = (\mathbf{y}_1', ..., \mathbf{y}_T')'$, $\mathbf{X} = (\mathbf{X}_1', ..., \mathbf{X}_T')'$, $\mathbf{u} = (\mathbf{u}_1', ..., \mathbf{u}_T')'$, and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1', ..., \boldsymbol{\varepsilon}_T')'$ with $\mathbf{y}_t = (y_{1t}, ..., y_{Nt})'$,$\mathbf{X}_t = (\mathbf{x}_{1t}, ..., \mathbf{x}_{Nt})'$, $\mathbf{u}_t = (u_{1t}, ..., u_{Nt})'$, and $\boldsymbol{\varepsilon}_t = (\varepsilon_{1t}, ..., \varepsilon_{Nt})'$. $\mathbf{1}_T$ is a T-dimensional vector of ones and $\otimes$ is the Kronecker product. We make use of the following assumptions:

Assumption 1: ε_{it} are independently distributed random variables with zero mean, variance $0 < E(\varepsilon_{it}^2) = \sigma_i^2 \leq \sigma_{\max}^2 < \infty$, and such that $E|\varepsilon_{it}|^{4+\eta} \leq K < \infty$ for some $\eta > 0$ and for $i = 1, ..., N; t = 1, ..., T$.

Assumption 2: $\mathbf{X}_t$ and $\boldsymbol{\varepsilon}_{t'}$ are independently distributed for all t and t'. As N and/or T go to infinity, the matrix $\frac{1}{NT}\mathbf{X}'(\mathbf{M} \otimes \mathbf{I}_N)\mathbf{X} \to \mathbf{C}$, where $\mathbf{C}$ is a finite, nonsingular matrix.

Assumption 3: The main diagonal elements of $\mathbf{S}$ are zero. The row and column norms of the matrices $\mathbf{S}$ and $(\mathbf{I}_N - \delta\mathbf{S})^{-1}$ are uniformly bounded.

Assumption 4: $\delta_0 \in [c_l, c_u]$, with $-\infty < c_l, c_u < \infty$, and $(\mathbf{I}_N - \delta\mathbf{S})^{-1}$ is non-singular for all $\delta \in [c_l, c_u]$.

The existence of moments of order higher than four stated in Assumption 1 is needed for applicability of the central limit theorem by Kelejian and Prucha (2001). Assumption 2 implies strict exogeneity of regressors, while Assumption 4 allows rewriting equations (4) as:

$$\mathbf{u} = (\mathbf{I}_T \otimes \mathbf{R}) \boldsymbol{\varepsilon}, \tag{5}$$

where $\mathbf{R} = (\mathbf{I}_N - \delta\mathbf{S})^{-1}$. The OLS estimator applied to (1) yields the fixed effects (FE) estimator (or within estimator) of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{X}'(\mathbf{M} \otimes \mathbf{I}_N)\mathbf{X}\right]^{-1} \mathbf{X}'(\mathbf{M} \otimes \mathbf{I}_N)\mathbf{y} \tag{6}$$

where $\mathbf{M} = \mathbf{I}_T - \mathbf{1}_T'\mathbf{1}_T/T$ is the matrix that converts y_{it} and $\mathbf{x}_{it}$ in deviations from their individual-specific means. Under Assumptions 1-4 the above estimator is $\sqrt{NT}$-consistent. However, when $\delta \neq 0$ $\hat{\boldsymbol{\beta}}$ is in general not efficient since the covariance of errors (4) is non-diagonal and the elements along its main diagonal are not constant. Efficient estimation of the slope coefficients $\boldsymbol{\beta}$ can be achieved by estimating the parameters of equation (5) and then computing the feasible fixed-effects Generalized Least Squares (GLS) of the slope coefficients (see Qian and Schmidt (2003) on this). In this paper we are concerned with consistent estimation of δ_0 via GMM. In the following, in order to distinguish the true parameters from other possible values in the parameter space, we denote by $\boldsymbol{\beta}_0$, δ_0, and σ_{0i}^2 the true parameters, which generate an observed sample.

3 GMM estimation of SAR error models

3.1 Moment conditions

Following Kelejian and Prucha (1999), Lee and Liu (2006a), and others, for GMM estimation we consider a set of r linear quadratic conditions in the error term. In a panel data setting, the ℓth population moment condition is

$$\mathcal{M}_\ell(\delta) = \frac{1}{NT} E\left[\boldsymbol{\varepsilon}(\delta)' (\mathbf{I}_T \otimes \mathbf{A}_\ell) \boldsymbol{\varepsilon}(\delta)\right], \quad \ell = 1, 2, ..., r \tag{7}$$

where

$$\boldsymbol{\varepsilon}(\delta) = [\mathbf{I}_T \otimes (\mathbf{I}_N - \delta\mathbf{S})] \mathbf{u} = [\mathbf{I}_T \otimes (\mathbf{I}_N - \delta\mathbf{S})] [\mathbf{y} - (\mathbf{1}_T \otimes \boldsymbol{\alpha}) - \mathbf{X}\boldsymbol{\beta}]$$

for any possible value of δ. We adopt the convention that $\boldsymbol{\varepsilon}(\delta_0) = \boldsymbol{\varepsilon}$. In (7), $\mathbf{A}_\ell$ are non-stochastic matrices with generic elements $a_{ij,\ell}$, and having bounded row and column norms. Following the work by Lee and Liu (2006a), we assume that the matrices inside the quadratic form have zero diagonal elements, namely $a_{ii,\ell} = 0$, for $i = 1, ..., N$ and $\ell = 1, 2, ..., r$, so that $\mathcal{M}_\ell(\delta_0) = 0$[1]. As explained by Lee and Liu (2006a), this assumption makes the GMM procedure robust to unknown heteroskedasticity. Further, this assumption in our specific framework is needed for some of the theoretical results reported in the appendix. The empirical counterpart of (7) is

$$M_{NT,\ell}(\delta) = \frac{1}{NT} \hat{\boldsymbol{\varepsilon}}(\delta)' (\mathbf{I}_T \otimes \mathbf{A}_\ell) \hat{\boldsymbol{\varepsilon}}(\delta),$$

where

$$\hat{\boldsymbol{\varepsilon}}(\delta) = [\mathbf{I}_T \otimes (\mathbf{I}_N - \delta\mathbf{S})] \hat{\mathbf{u}} = [\mathbf{M} \otimes (\mathbf{I}_N - \delta\mathbf{S})] \left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)$$

$\hat{\boldsymbol{\beta}}$ being the FE estimator (6). The following propositions hold.

Proposition 1 *Under Assumptions 1-4 we have, for all $\delta \in [c_l, c_u]$,*

$$\frac{1}{NT}\left(\hat{\boldsymbol{\varepsilon}}(\delta)' (\mathbf{I}_T \otimes \mathbf{A}_\ell) \hat{\boldsymbol{\varepsilon}}_t(\delta) - \boldsymbol{\varepsilon}(\delta)' (\mathbf{I}_T \otimes \mathbf{A}_\ell) \boldsymbol{\varepsilon}(\delta)\right) = O_p\left(\frac{1}{T\sqrt{N}}\right) + O_p\left(\frac{1}{T}\right) \tag{8}$$

Further, at δ_0 we have

$$\frac{1}{NT}\left[\hat{\boldsymbol{\varepsilon}}(\delta_0)' (\mathbf{I}_T \otimes \mathbf{A}_\ell) \hat{\boldsymbol{\varepsilon}}(\delta_0) - \boldsymbol{\varepsilon}(\delta_0)' (\mathbf{I}_T \otimes \mathbf{A}_\ell) \boldsymbol{\varepsilon}(\delta_0)\right] = O_p\left(\frac{1}{T\sqrt{N}}\right) \tag{9}$$

Proposition 2 *Under Assumptions 1-4 we have, for all $\delta \in [c_l, c_u]$,*

$$\frac{1}{NT}\left[\boldsymbol{\varepsilon}(\delta)' (\mathbf{I}_T \otimes \mathbf{A}_\ell) \boldsymbol{\varepsilon}(\delta) - E\left(\boldsymbol{\varepsilon}(\delta)' (\mathbf{I}_T \otimes \mathbf{A}_\ell) \boldsymbol{\varepsilon}(\delta)\right)\right] = O_p\left(\frac{1}{\sqrt{NT}}\right). \tag{10}$$

[1] See Lemma 4.

The proofs of the above propositions are reported in the Appendix. Result (8) states that, for any possible value of δ, as N and T tend to infinity, the ℓth empirical moment computed using the estimated regression coefficients converges in probability to the ℓth moment computed using the true regression coefficients. However, when $\delta = \delta_0$, the $O_p(1/T)$ rate of convergence disappears, as shown by result (9). We remark that one essential assumption to obtain (9) is that the the inner matrix, $\mathbf{A}_\ell$, has zero diagonal elements (see equation (29)). Under (10), as N and/or T go to infinity, the quadratic form in the errors converges in probability to its mean, for any possible value of δ. Similar results have been obtained by Lee and Liu (2006a) for the single cross section case, with no individual fixed effects.

3.2 Estimation

Let $\mathcal{M}(\delta) = [\mathcal{M}_1(\delta), ..., \mathcal{M}_r(\delta)]'$ be a vector containing r moments, and let $\mathbf{M}_{NT}(\delta) = [M_{NT,1}(\delta), ..., M_{NT,r}(\delta)]'$ be the vector of the corresponding sample moments. Let

$$\mathbf{V}(\boldsymbol{\theta}_0) = \lim_{N,T\to\infty} E\left[NT\mathbf{M}_{NT}(\delta_0)\mathbf{M}_{NT}(\delta_0)'\right] \tag{11}$$

where $\boldsymbol{\theta}_0 = \left(\delta_0; \sigma_{01}^2, ..., \sigma_{0N}^2\right)'$. Given result (9) and Lemma 4 (see the Appendix), the above matrix has generic $(\ell, h)^{th}$ element, $v_{\ell h}$, given by

$$v_{\ell h} = \frac{1}{N} Tr\left[\boldsymbol{\Sigma}\mathbf{A}_\ell\boldsymbol{\Sigma}\mathbf{A}_h + \boldsymbol{\Sigma}\mathbf{A}_\ell\mathbf{A}_h'\boldsymbol{\Sigma}\right]. \tag{12}$$

with $\boldsymbol{\Sigma}$ being a diagonal matrix with elements $\sigma_{01}^2, ..., \sigma_{0N}^2$ on the main diagonal. Under the assumption of bounded row and column norms of the matrices $\mathbf{A}_\ell$ and $\mathbf{A}_h$, it is easily shown that $v_{\ell h} = O(1)$. We take up the following assumptions needed for identificability of parameters:

Assumption 5: The matrix $\mathbf{V}(\boldsymbol{\theta}_0)$ is non-singular, i.e. we assume $\lambda_r(\mathbf{V}(\boldsymbol{\theta}_0)) \geq K > 0$.

Assumption 6: There exists at least one moment condition, the ℓ^{th}, for which we have either $Tr\left[\boldsymbol{\Sigma}\mathbf{A}_\ell\mathbf{S}(\mathbf{I}_N - \delta_0\mathbf{S})^{-1}\right] \neq 0$, or $Tr\left[\boldsymbol{\Sigma}\left((\mathbf{I}_N - \delta_0\mathbf{S}')^{-1}\mathbf{S}'\mathbf{A}_\ell\mathbf{S}(\mathbf{I}_N - \delta_0\mathbf{S})^{-1}\right)\right] \neq 0$.

The GMM estimator $\hat{\delta}$ of δ_0 is the solution to the following optimization problem

$$\hat{\delta} = \arg\min_{\delta\in[c_l,c_u]} \left\{\mathbf{M}_{NT}(\delta)'\mathbf{Q}_{NT}\mathbf{M}_{NT}(\delta)\right\} \tag{13}$$

where $[c_l, c_u]$ is the parameter space (see Assumption 4), and $\mathbf{Q}_{NT}$ is a $r \times r$, positive definite, weighting matrix, such that

$$\mathbf{Q}_{NT} \xrightarrow{p} \mathbf{Q}$$

The following theorem states that $\hat{\delta}$ is consistent for δ_0 and establishes its asymptotic distribution.

Theorem 3 *Under Assumptions 1-6, $\hat{\delta}$ in (13) is consistent for δ as N and T going to infinity. Further, we have*

$$\sqrt{NT}\left(\hat{\delta}-\delta_0\right) \overset{a}{\sim} N\left(0, \left(\mathbf{d}'\mathbf{Q}\mathbf{d}\right)^{-1}\mathbf{d}'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{d}\left(\mathbf{d}'\mathbf{Q}\mathbf{d}\right)^{-1}\right) \tag{14}$$

where $\mathbf{d}=\mathbf{d}\left(\boldsymbol{\theta}_0\right)=\underset{N,T\to\infty}{p\lim}\, E\left[\frac{\partial}{\partial\delta}\mathbf{M}_{NT}\left(\delta\right)\Big|_{\delta=\delta_0}\right]$.

The proof is reported in the Appendix (see Section 6.2). The efficient GMM estimator can be obtained by imposing, in (13), the optimal weighting matrix $\mathbf{Q}=\mathbf{Q}^*=\mathbf{V}^{-1}$ (see Greene (2008) on this). Notice that the ℓth element of $\mathbf{d}$ is (see Section 6.3)

$$d_\ell=\frac{1}{N}Tr\left(\boldsymbol{\Sigma}\left(\mathbf{A}_\ell+\mathbf{A}_\ell'\right)\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right) \tag{15}$$

Since both $\mathbf{Q}^*$ and $\mathbf{d}$ depend on δ_0 and σ_{0i}^2, in practise, $\mathbf{Q}$ and $\mathbf{d}$ are evaluated at point estimates, $\mathbf{Q}^*=\mathbf{Q}^*\left(\hat{\boldsymbol{\theta}}\right)$, and $\mathbf{d}=\mathbf{d}\left(\hat{\boldsymbol{\theta}}\right)$, where the elements in the matrix $\boldsymbol{\Sigma}$ may be consistently estimated by

$$\hat{\sigma}_i^2=\frac{1}{T}\sum_{t=1}^{T}\hat{\varepsilon}_{it}^2, \quad i=1,...,N. \tag{16}$$

If $\mathbf{A}_\ell$ does not depend on unknown parameters (for example, δ_0) we can compute $\hat{\delta}$ in a single step by minimizing (13). However, in general, $\mathbf{A}_\ell$ (and hence also the optimal weighting matrix, $\mathbf{Q}_{NT}^*$) may depend on unknown parameters, such as δ_0 or the measures of skewness and kurtosis of the distribution of ε_{it}. In this case it is possible to apply an iterative two-stages procedure described in the next section.

3.3 Choice of the inner matrices

We now consider the problem of selecting the inner matrices, $\mathbf{A}_\ell$ for $\ell=1,2,...,r$ for building the moment conditions. Kelejian and Prucha (1999) and Kelejian and Prucha (2008) suggest[2]

$$\mathbf{A}_{1,KP}=\mathbf{S}'\mathbf{S}-diag\left(\mathbf{S}'\mathbf{S}\right), \quad \mathbf{A}_{2,KP}=\mathbf{S} \tag{17}$$

When inner matrices $\mathbf{A}_{1,KP}$,$\mathbf{A}_{2,KP}$ are employed, the optimal weighting matrix in the minimization problem (13) is (see Lemma 4)

$$\mathbf{Q}^{*KP}=\frac{1}{N}\begin{pmatrix} 2Tr\left[\left(\boldsymbol{\Sigma}\mathbf{S}'\mathbf{S}\right)^2\right] & Tr\left(\boldsymbol{\Sigma}\mathbf{S}'\mathbf{S}\boldsymbol{\Sigma}\mathbf{S}+\boldsymbol{\Sigma}\mathbf{S}'\mathbf{S}\mathbf{S}'\boldsymbol{\Sigma}\right) \\ Tr\left(\boldsymbol{\Sigma}\mathbf{S}'\mathbf{S}\boldsymbol{\Sigma}\mathbf{S}+\boldsymbol{\Sigma}\mathbf{S}'\mathbf{S}\mathbf{S}'\boldsymbol{\Sigma}\right) & Tr\left(\boldsymbol{\Sigma}\mathbf{S}\boldsymbol{\Sigma}\mathbf{S}+\boldsymbol{\Sigma}\mathbf{S}\mathbf{S}'\boldsymbol{\Sigma}\right)\end{pmatrix} \tag{18}$$

[2] Notice that we consider Kelejian and Prucha (1999) inner matrices in the form of deviations from their diagonal elements.

Since $\mathbf{Q}^{*KP}$ depends on the elements of $\mathbf{\Sigma}$, estimation can proceed by adopting the following two-stages iterative procedure. First, minimize (13) using as weighting matrix the identity matrix $\mathbf{I}_r$, and the OLS residuals, $\hat{u}_{it}$ to obtain an initial estimate, say $\hat{\delta}^{(1)}$. In the second stage, employ $\hat{\delta}^{(1)}$ to estimate errors ε_{it}^2 and σ_{0i}^2 using (16), and hence $\mathbf{Q}^*$ and use this in the minimization problem (13). We can alternate back and forth between the estimation of δ conditional upon a weighting matrix $\mathbf{Q}^*$ and the estimation of $\mathbf{Q}^*$ conditional upon a value for δ, until convergence is obtained. Standard errors of the final estimate of the spatial parameter can be obtained by formula (14).

We notice that, if it is reasonable to assume homoskedasticity, i.e. $\sigma_{0i}^2 = \sigma_0^2$ for $i = 1, ..., N$, then (18) reduces to

$$\mathbf{Q}^{*KP} = \frac{\sigma_0^4}{N} \begin{pmatrix} 2Tr\left[(\mathbf{S}'\mathbf{S})^2\right] & 2Tr\left(\mathbf{S}'\mathbf{S}^2\right) \\ 2Tr\left(\mathbf{S}'\mathbf{S}^2\right) & Tr\left(\mathbf{S}^2 + \mathbf{S}\mathbf{S}'\right) \end{pmatrix}$$

and σ_0^4 enters in $\mathbf{Q}^{*KP}$ only as a scale factor. In this case, $\hat{\delta}$ can be computed in a single step, since $\mathbf{Q}^{*KP}$ does not involve estimation of unknown parameters.

In the context of a simple regression model with homoskedastic, spatial autoregressive errors, (Liu, Lee, and Bollinger 2006) suggests to base GMM estimation of δ on an empirical moment having the following inner matrix:

$$\mathbf{A}_L = \mathbf{H}_0 - \frac{1}{N} Tr\left(\mathbf{H}_0\right)\mathbf{I}_N - \frac{\eta_4 - 3}{\eta_4 - 1}\left(diag\left(\mathbf{H}_0\right) - \frac{1}{N} Tr\left(\mathbf{H}_0\right)\mathbf{I}_N\right), \tag{19}$$

with $\mathbf{H}_0 = \left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1}\mathbf{S}$, and $\eta_4 = \mu_{04}/\sigma_0^4$ being the kurtosis parameters of the distribution of ε_{it}. Under the homoskedasticity assumption, the author shows that employing the above matrix leads to a GMM estimator with minimal variance. Since using the above matrix requires estimating δ_0, $\mathbf{\Sigma}$, and η_4, estimation may proceed by adopting the iterative two-stages estimation procedure outlined above. In the first step an initial guess of δ_0 and η_4 need to be formulated and used to build $\mathbf{A}_L$.

Other moment conditions can be obtained by looking at the properties of (4). For example, the following inner matrices can be suggested:

$$\mathbf{A}_1 = \mathbf{R}_0'\mathbf{S}'\mathbf{S}\mathbf{R}_0 - diag\left(\mathbf{\Sigma}\mathbf{R}_0'\mathbf{S}'\mathbf{S}\mathbf{R}_0\right), \quad \mathbf{A}_2 = \mathbf{R}_0'\mathbf{S}\mathbf{R}_0 - diag\left(\mathbf{\Sigma}\mathbf{R}_0'\mathbf{S}\mathbf{R}_0\right), \tag{20}$$

$$\mathbf{A}_3 = \mathbf{R}_0'\mathbf{S}'\mathbf{S} - diag\left(\mathbf{\Sigma}\mathbf{R}_0'\mathbf{S}'\mathbf{S}\right), \quad \mathbf{A}_4 = \mathbf{R}_0'\mathbf{S} - diag\left(\mathbf{\Sigma}\mathbf{R}_0'\mathbf{S}\right). \tag{21}$$

where $\mathbf{R}_0 = \left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1}$. Moment based on $\mathbf{A}_1$ exploits the variance of the spatial lag $\left(\mathbf{I}_T \otimes \mathbf{S}\right)\mathbf{u}$. Moments $\mathbf{A}_2, \mathbf{A}_3$ are based on the covariance of $\mathbf{u}$ with $\left(\mathbf{I}_T \otimes \mathbf{S}\right)\mathbf{u}$ and $\left(\mathbf{I}_T \otimes \mathbf{S}\right)\boldsymbol{\varepsilon}$, respectively. Finally, $\mathbf{A}_4$ arises when looking at the covariance between the spatial lags $\left(\mathbf{I}_T \otimes \mathbf{S}\right)\mathbf{u}$ and $\left(\mathbf{I}_T \otimes \mathbf{S}\right)\boldsymbol{\varepsilon}$. The above inner matrices depend on δ_0, that needs to be estimated in a first step. We notice that, under $\delta_0 = 0$, $\mathbf{A}_1$ would be identical to $\mathbf{A}_3$ and $\mathbf{A}_{1,KP}$, while $\mathbf{A}_2$ would coincide with $\mathbf{A}_4$ and $\mathbf{A}_{2,KP}$.

Before concluding, we remark that in principle any matrix having bounded row and colum norms can be selected for building the moment conditions. However, as will we see in the next section, the choice of inner matrices do have an impact on the small sample properties of the GMM estimator.

In the following, we assess and compare the performance of GMM estimators based on conditions that use as inner matrices (17), (19) or (20)-(21) by the means of Monte Carlo experiments.

4 Monte Carlo experiments

4.1 The design

We consider the following data generating process:

$$\begin{aligned} y_{it} &= \alpha_i + \beta_1 x_{1,it} + \beta_2 x_{2,it} + u_{it}, i = 1, ..., N; T = 1, ..., T \\ u_{it} &= \delta \sum_{j=1}^{N} s_{ij} u_{jt} + \varepsilon_{it} \end{aligned}$$

where we assume $\alpha_i = IIDN(1,1), \beta_1 = \beta_2 = 1$ and

$$\begin{aligned} x_{\ell,it} &= \rho_i x_{\ell,it-1} + \upsilon_{\ell it}, i = 1, ..., N, t = -49, ..., 1, ..., T, \ell = 1, 2, x_{\ell,i-50} = 0, \\ \upsilon_{\ell,it} &\sim IIDN(0, 1-\rho_i^2), .\rho_i \sim U(0.5, 0.95). \end{aligned}$$

Errors ε_{it} are generated under two alternative schemes: (i) normal errors, $\varepsilon_{it} \sim IIDN(0, \sigma_i^2)$; (ii) chi-squared errors, $\varepsilon_{it} \sim IID\left(\chi_1^2 - 1\right)/\sqrt{2}$, with $\sigma_i^2 \sim \chi_2^2/2$. The values of $x_{\ell,it}$ and u_{it} are drawn for each i and t, and at each replication, while α_i and σ_i^2 are kept fixed across replications. The first 50 observations are discarded to avoid possible initial value effect. We carry out our experiments for $N, T = 20, 50, 100$, and $T = 20, 50, 100$. The matrix $\mathbf{S}$ has elements $s_{ii} = 0$, and $s_{ij} = 1$ if units i and j are adjacent and $s_{ij} = 0$ otherwise, for $i \neq j$. In a first set of experiments we assume $\mathbf{S}$ is a regular lattice, and cross section units are arranged so that the pth order neighbours of the ith cross section unit can be defined as the $(i-p)$th and $(i+p)$th units. In our experiments, we try with $p = 1$ and $p = 2$. We define $\mathbf{S}$ in a circular world, where the first observation is adjacent to the last observation, and express it in row-standardized form. In a second set of experiments, we use real-world spatial weights matrices that describe the spatial arrangement of three subsets of English local authorities[3]: the first subset contains 13 contiguous authorities[4]; the second subset is made of 33 contiguous authorities[5]; the third subset has all English local authorities, except the Isle of Scilly, for a total of 149 cross section units.

[3] See .http://geodacenter.asu.edu/. In constructing spatial weights matrices for the English local authorities, we used a rook contiguity criterion.

[4] This set includes all local authorities of inner London.

[5] This set includes the all authorities of greater London.

Table 1: Properties of spatial weights matrices adopted in Monte Carlo experiments

Spatial weights	N	% non-zero links	$\|\mathbf{S}_N\|_C$
Regular lattices; $p = 1$			
$\mathbf{S}_{20}$	20	10.00	2
$\mathbf{S}_{50}$	50	4.00	2
$\mathbf{S}_{100}$	100	2.00	2
Regular lattices; $p = 2$			
$\mathbf{S}_{20}$	20	20.00	4
$\mathbf{S}_{50}$	50	8.00	4
$\mathbf{S}_{100}$	100	4.00	4
Irregular lattices			
$\mathbf{S}_{13}$	13	25.00	13
$\mathbf{S}_{33}$	33	14.84	7
$\mathbf{S}_{149}$	149	3.41	8

The purpose of using such real-world spatial weights matrices is to investigate the properties of GMM estimators when there exists an irregular spatial arrangement of units, a situation often encountered in empirical work.

The connectedness characteristics of the spatial weights in unstandardized form used in our experiments, in terms of percent of non-zero elements and maximum number of neighbours, are given in Table 1. We notice that the regular lattices are always sparser than their irregular counterparts for similar sample size. The sparseness of the spatial weights matrix adopted in the analysis is an important factor in determining the extent to which the central limit theorems on dependent spatial processes hold, and hence is likely affect the performance of our estimators. In our simulation exercise all weights matrices are used in row-standardized form.

We experimented with $\delta = -0.8, -0.3, 0.0, 0.3, 0.8$, and provide results for the following estimators of δ:

1. GMM estimator using Kelejian and Prucha inner matrices (17), $\hat{\delta}^{KP}$.
2. GMM estimator using Lee inner matrix (19), $\hat{\delta}^{L}$.
3. GMM estimator based on inner matrices (20), $\hat{\delta}^{(1)}$.
4. GMM estimator based on inner matrices (21), $\hat{\delta}^{(2)}$.
5. GMM estimator based on an inner matrix with zero diagonal elements, and all remaining entries equal to $1/N$, $\hat{\delta}^{(3)}$.

6. Quasi-maximum likelihood (quasi-ML) estimator, in a model with fixed effects and unknown heteroskedasticity[6], $\hat{\delta}_{ML}$.

By the means of the above Monte Carlo experiments, we wish to investigate a number of issues. First, we wish to assess the small sample properties of $\hat{\delta}^{KP}$ and $\hat{\delta}^{L}$, also in comparison with the quasi-maximum likelihood estimators in a panel data context with fixed effects and unknown heteroskedasticity. Second, we want to investigate the performance of $\hat{\delta}^{(1)}$ and $\hat{\delta}^{(2)}$, also in comparison with $\hat{\delta}^{KP}$ and $\hat{\delta}^{L}$. A further aim of these experiments is to explore the properties of a GMM estimator based on a inner matrix with all entries equal to $1/N$ (namely, $\hat{\delta}^{(3)}$). Such matrix assigns equal weights to all cross products appearing in the quadratic form as opposed to other matrices that give more importance to cross products of close observations.

Estimation of δ is performed on residuals $\hat{\mathbf{u}} = (\mathbf{M} \otimes \mathbf{I}_N) \left(\mathbf{y} - \hat{\beta}_1 \mathbf{x}_1 - \hat{\beta}_2 \mathbf{x}_2\right)$, where $\hat{\beta}_1$ and $\hat{\beta}_2$ are FE estimates of β_1 and β_2. We assess the performance of estimators of δ by computing their bias, RMSE, size and power. In computing size and power, we adopt a significance level of 5 per cent. The number of replications is $1,000$.

4.2 Results

Monte Carlo results for estimators of δ are given in Table 2-5. For $\hat{\delta}^{L}$, $\hat{\delta}^{(1)}$, $\hat{\delta}^{(2)}$, and $\hat{\delta}^{(3)}$ we report the estimates obtained adopting as weighting matrix in (13) both the optimal weighting matrix and the identity matrix, $\mathbf{I}_r$. To save space, we only report results for for $\delta = 0.0, 0.3, 0.8$ and for $p = 1$. Further, in the case of non-normal errors and real-world matrices we only report the estimates obtained adopting the optimal weighting matrix. Finally, when using real-world matrices we only show results for $\delta = 0.3$. Other results are available upon request.

As expected, the bias and RMSE of $\hat{\delta}^{KP}$ decrease as N and/or T get large, for all values of δ. The empirical rejection rates corresponding to $\hat{\delta}^{KP}$ are close to the nominal 5 per cent level for $\delta = 0.0, 0.3$, and for all T greater than 10. Conversely, they slightly deviate from the theoretical 5 per cent level when $T = 10$ and N is equal or smaller than 50. When $\delta = 0.8$, $\hat{\delta}^{KP}$ has the correct size only for large N and T, while for other combinations of N and T the empirical rejection frequencies are slightly larger than the nominal 5 per cent level. We observe that, for a given pair of N and T, larger absolute values of δ are associated to a smaller RMSE and a higher power of $\hat{\delta}^{KP}$. According to Arraiz, Drukker, Kelejian, and Prucha (2008), an explanation for this result is that a larger δ in absolute value increases the variability of the term $(\mathbf{I}_N \otimes \mathbf{S})\,\mathbf{u}$ in (4), and hence increases the precision of the GMM estimator.

A similar pattern can be observed for the GMM estimator based on other sets of conditions (i.e., for $\hat{\delta}^{L}$, $\hat{\delta}^{(1)}$, $\hat{\delta}^{(2)}$, and $\hat{\delta}^{(3)}$). However, some differences in the performance of these estimators can be noted. First, we observe that the estimator $\hat{\delta}^{L}$ (i.e., the GMM estimator based on

[6] The derivation of the information matrix used for computing the standard errors of ML estimates is available upon request.

conditions with inner matrix (19)) performs overall better respect to other GMM estimators, using either $\mathbf{I}_3$ or $\mathbf{Q}^*$ as weighting matrix. In particular, the bias and RMSE of $\hat{\delta}_{2,GMM}$ are lower than those for $\hat{\delta}^{KP}$, $\hat{\delta}^{(1)}$, $\hat{\delta}^{(2)}$, and $\hat{\delta}^{(3)}$, for all values of δ, while the rejection rates are very close to 5 per cent. This finding corroborates the theoretical results obtained by Liu, Lee, and Bollinger (2006) on the best GMM estimator. The performance of GMM estimators $\hat{\delta}^{(1)}$ and $\hat{\delta}^{(2)}$ is similar to that of $\hat{\delta}_{KP}$, using either $\mathbf{I}_3$ or $\mathbf{Q}^*$ as weighting matrix, and for normal or non-normal errors. Finally, $\hat{\delta}^{(3)}$ performs quite well, its bias and RMSE are only slightly above that of $\hat{\delta}^{L}$. To conclude, the quasi-ML estimator shows little bias and RMSE, but it is characterized by high rejection rates when T is small, especially for high values of δ and for non-normal errors.

5 Conclusions

In this paper we have focused on GMM estimation of a spatial panel with fixed effects and unknown heteroskedasticity. We have considered as moments a set linear quadratic conditions applied to residuals transformed by the demeaning operator. To robustify moments against unknown heteroskedasticity we have set to zero the diagonal elements of inner matrices, as in Lee and Liu (2006a). We show that consistency and asymptotic normality of the parameters of the SAR process is achieved for N and T going to infinity, with no restrictions on the relative rate at which N and T rise. Our Monte Carlo exercise shows that the GMM estimator has good small sample properties, when compared to the performance of the quasi-ML, especially when T is relatively small, and when the spatial parameter is close to 1.

References

Arraiz, I., D. M. Drukker, H. H. Kelejian, and I. R. Prucha (2008). A spatial cliff-ord-type model with heteroskedastic innovations: Small and large sample results. Mimeo, University of Maryland.

Druska, V. and W. C. Horrace (2004). Generalized moments estimation for spatial panels: indonesian rice farming. *American Journal of Agricultural Economics 86*, 185–198.

Fingleton, B. (2008a). A generalized method of moments estimator for a spatial model with moving average errors, with application to real estate prices. *Empirical Economics 34*, 35Ű57.

Fingleton, B. (2008b). A generalized method of moments estimator for a spatial panel model with an endogenous spatial lag and spatial moving average errors. *Spatial Economic Analysis 3*, 27–44.

Greene, W. (2008). *Econometric Analysis* (Sixth ed.). Prentice Hall.

Kapoor, M., H. H. Kelejian, and I. Prucha (2007). Panel data models with spatially correlated error components. *Journal of Econometrics 140*, 97–130.

Kelejian, H. H. and I. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*, 509–533.

Kelejian, H. H. and I. Prucha (2001). On the asymptotic distribution of the moran i test with applications. *Journal of Econometrics 104*, 219–257.

Kelejian, H. H. and I. Prucha (2008). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Forthcoming, Journal of Econometrics.

Lee, L. F. (2007). Gmm and 2sls estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics 137*, 489–514.

Lee, L. F. and X. Liu (2006a). Efficient GMM estimation of a SAR model with autoregressive disturbances. Mimeo.

Lee, L. F. and X. Liu (2006b). Efficient GMM estimation of high order spatial autoregressive models. Mimeo.

Lee, L. F. and J. Yu (2008). Estimation of spatial autoregressive panel data models with fixed effects. Mimeo, Ohio State University.

Lin, X. and L. F. Lee (2005). Gmm estimation of spatial autoregressive models with unknown heteroskedasticity. Mimeo, Ohio State University.

Liu, X., L. F. Lee, and C. R. Bollinger (2006). Improved efficient quasi maximum likelihood estimator of spatial autoregressive models. Mimeo.

Mutl, J. and M. Pfaffermayr (2008). The spatial random effects and the spatial fixed effects model: The hausman test in a cliff and ord panel model. Mimeo, Institute for Advanced Studies, Vienna.

Qian, H. and P. Schmidt (2003). Partial gls regression. *Economics Letters 79*, 385–392.

Ullah, A. (2004). *Finite Sample Econometrics*. Oxford: Oxford University Press.

Yu, J., R. de Jong, and L. F. Lee (2007). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large: A nonstationary case. Mimeo, Ohio State University.

Yu, J., R. de Jong, and L. F. Lee (2008). Quasi-maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and t are large. *Journal of Econometrics 146*, 118–137. Mimeo, The Ohio State University.

6 Appendix

Lemma 4 *Let* $\mathbf{A}, \mathbf{B}$ *be two non-stochastic matrices with bounded row and column norms. We have:*

$$E\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{A}\right)\boldsymbol{\varepsilon}\right] = \frac{1}{N}Tr\left(\boldsymbol{\Sigma}\mathbf{A}\right), \tag{22}$$

$$Var\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{A}\right)\boldsymbol{\varepsilon}\right] = \frac{1}{N^2T}\sum_{i=1}^{N} a_{ii}^2\left[E\left(\varepsilon_{it}^4\right) - 3\sigma_i^4\right] + \frac{1}{N^2T}Tr\left[\left(\boldsymbol{\Sigma}\mathbf{A}\right)^2 + \boldsymbol{\Sigma}\mathbf{A}\mathbf{A}'\boldsymbol{\Sigma}\right], \tag{23}$$

$$\begin{aligned} Cov\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{A}\right)\boldsymbol{\varepsilon}, \frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{B}\right)\boldsymbol{\varepsilon}\right] &= \frac{1}{N^2T}\sum_{i=1}^{N} b_{ii}a_{ii}\left[E\left(\varepsilon_{it}^4\right) - 3\sigma_i^4\right] \\ &+ \frac{1}{N^2T}Tr\left[\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} + \boldsymbol{\Sigma}\mathbf{A}\mathbf{B}'\boldsymbol{\Sigma}\right]. \end{aligned} \tag{24}$$

If $\mathbf{A}, \mathbf{B}$ *have zero diagonal elements then*

$$\begin{aligned} E\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{A}\right)\boldsymbol{\varepsilon}\right] &= 0, \\ Var\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{A}\right)\boldsymbol{\varepsilon}\right] &= \frac{1}{N^2T}Tr\left[\left(\boldsymbol{\Sigma}\mathbf{A}\right)^2 + \boldsymbol{\Sigma}\mathbf{A}\mathbf{A}'\boldsymbol{\Sigma}\right], \\ Cov\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{A}\right)\boldsymbol{\varepsilon}, \frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{B}\right)\boldsymbol{\varepsilon}\right] &= \frac{1}{N^2T}Tr\left[\boldsymbol{\Sigma}\mathbf{A}\boldsymbol{\Sigma}\mathbf{B} + \boldsymbol{\Sigma}\mathbf{A}\mathbf{B}'\boldsymbol{\Sigma}\right]. \end{aligned}$$

Proof. See Ullah (2004). ■

6.0.1 Proof of Proposition 1

We now sketch the proof of proposition 1, and refer to Kelejian and Prucha (2008), Lee and Liu (2006a), Lee (2007), and Kelejian and Prucha (1999) for further details on the convergence of quadratic forms. First, consider

$$\begin{aligned} \hat{\boldsymbol{\varepsilon}}\left(\delta\right) &= \left[\mathbf{M} \otimes \left(\mathbf{I}_N - \delta\mathbf{S}\right)\right]\left(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \left[\mathbf{M} \otimes \left(\mathbf{I}_N - \delta\mathbf{S}\right)\right]\left[\mathbf{X}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1}\boldsymbol{\varepsilon}\right] \\ &= \left[\mathbf{M} \otimes \left(\mathbf{I}_N - \delta\mathbf{S}\right)\right]\mathbf{X}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \left[\mathbf{M} \otimes \left(\mathbf{I}_N - \delta\mathbf{S}\right)\left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1}\right]\boldsymbol{\varepsilon}. \end{aligned}$$

Noting that $\left(\mathbf{I}_N - \delta\mathbf{S}\right)\left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1}$ can be also written as

$$\begin{aligned} \left(\mathbf{I}_N - \delta\mathbf{S}\right)\left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1} &= \left(\mathbf{I}_N - \delta_0\mathbf{S} + \delta_0\mathbf{S} - \delta\mathbf{S}\right)\left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1} \\ &= \left(\mathbf{I}_N - \delta_0\mathbf{S}\right)\left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1} + \left(\delta_0 - \delta\right)\mathbf{S}\left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1} \\ &= \mathbf{I}_N + \left(\delta_0 - \delta\right)\mathbf{S}\left(\mathbf{I}_N - \delta_0\mathbf{S}\right)^{-1}, \end{aligned}$$

we can rewrite $\boldsymbol{\varepsilon}(\delta)$ and $\hat{\boldsymbol{\varepsilon}}(\delta)$ as follows

$$\begin{aligned}\boldsymbol{\varepsilon}(\delta) &= \left[\mathbf{I}_T \otimes \left(\mathbf{I}_N + (\delta_0 - \delta)\mathbf{S}(\mathbf{I}_N - \delta_0 \mathbf{S})^{-1}\right)\right]\boldsymbol{\varepsilon} = [\mathbf{I}_T \otimes \mathbf{P}(\delta)]\boldsymbol{\varepsilon} \\ \hat{\boldsymbol{\varepsilon}}(\delta) &= [\mathbf{M} \otimes (\mathbf{I}_N - \delta\mathbf{S})]\mathbf{X}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + [\mathbf{M} \otimes \mathbf{P}(\delta)]\boldsymbol{\varepsilon}\end{aligned}$$

where

$$\mathbf{P}(\delta) = \mathbf{I}_N + (\delta - \delta_0)\mathbf{S}(\mathbf{I}_N - \delta_0\mathbf{S})^{-1}. \tag{25}$$

To prove (8), note that

$$\begin{aligned}\frac{1}{NT}\hat{\boldsymbol{\varepsilon}}(\delta)'(\mathbf{I}_T \otimes \mathbf{A}_\ell)\hat{\boldsymbol{\varepsilon}}(\delta) &= \frac{1}{NT}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{X}'\left[\mathbf{M}\otimes(\mathbf{I}_N - \delta\mathbf{S})'\mathbf{A}_\ell(\mathbf{I}_N - \delta\mathbf{S})\right]\mathbf{X}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) \\ &\quad + \frac{2}{NT}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{X}'\left[\mathbf{M}\otimes(\mathbf{I}_N - \delta\mathbf{S})'\mathbf{A}_\ell\mathbf{P}(\delta)\right]\boldsymbol{\varepsilon} \\ &\quad + \boldsymbol{\varepsilon}'\left(\mathbf{M}\otimes\mathbf{P}(\delta)'\mathbf{A}_\ell\mathbf{P}(\delta)\right)\boldsymbol{\varepsilon} \\ &= \frac{1}{NT}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{X}'(\mathbf{M}\otimes\mathbf{B}_\ell)\mathbf{X}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) + \frac{2}{NT}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{X}'(\mathbf{M}\otimes\mathbf{C}_\ell)\boldsymbol{\varepsilon} \\ &\quad + \boldsymbol{\varepsilon}'(\mathbf{M}\otimes\mathbf{D}_\ell)\boldsymbol{\varepsilon}.\end{aligned}$$

where $\mathbf{B}_\ell = (\mathbf{I}_N - \delta\mathbf{S})'\mathbf{A}_\ell(\mathbf{I}_N - \delta\mathbf{S})$, $\mathbf{C}_\ell = (\mathbf{I}_N - \delta\mathbf{S})'\mathbf{A}_\ell\mathbf{P}(\delta)$, and $\mathbf{D}_\ell = \mathbf{P}(\delta)'\mathbf{A}_\ell\mathbf{P}(\delta)$. Under Assumptions 3-4 $\mathbf{B}_\ell, \mathbf{C}_\ell$, and $\mathbf{D}_\ell$ have row and column norms that are uniformly bounded. Given the $\sqrt{NT}$-consistency of $\hat{\boldsymbol{\beta}}$ we obtain

$$\begin{aligned}\frac{1}{NT}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{X}'(\mathbf{M}\otimes\mathbf{B}_\ell)\mathbf{X}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) &\leq \frac{K}{NT}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{X}'(\mathbf{M}\otimes\mathbf{I}_N)\mathbf{X}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) = O_p\left(\frac{1}{NT}\right), \\ \frac{2}{NT}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{X}'(\mathbf{M}\otimes\mathbf{C}_\ell)\boldsymbol{\varepsilon} &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{j=1}^{N} c_{\ell,ij}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right)'\mathbf{X}_i'\mathbf{M}\boldsymbol{\varepsilon}_j = O_p\left(\frac{1}{NT}\right).\end{aligned}$$

Further, we have

$$\frac{1}{NT}\boldsymbol{\varepsilon}'(\mathbf{M}\otimes\mathbf{D}_\ell)\boldsymbol{\varepsilon} = \frac{1}{NT}\boldsymbol{\varepsilon}'(\mathbf{I}_T\otimes\mathbf{D}_\ell)\boldsymbol{\varepsilon} - \frac{1}{NT}\boldsymbol{\varepsilon}'\left(\frac{\mathbf{ii}'}{T}\otimes\mathbf{D}_\ell\right)\boldsymbol{\varepsilon} \tag{26}$$

The second term in (26) has mean

$$\begin{aligned}E\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\frac{\mathbf{ii}'}{T}\otimes\mathbf{D}_\ell\right)\boldsymbol{\varepsilon}\right] &= \frac{1}{NT}E\left(\sum_{i=1}^{N}\sum_{j=1}^{N} d_{\ell,ij}\boldsymbol{\varepsilon}_i\frac{\mathbf{ii}'}{T}\boldsymbol{\varepsilon}_j\right) = \frac{1}{NT}\sum_{i=1}^{N} d_{\ell,ii}E\left(\boldsymbol{\varepsilon}_i\frac{\mathbf{ii}'}{T}\boldsymbol{\varepsilon}_i\right) \\ &\leq \frac{\sigma^2_{\max}}{NT}\sum_{i=1}^{N} d_{\ell,ii} = O\left(\frac{1}{T}\right),\end{aligned}$$

and variance

$$
\begin{aligned}
Var\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\frac{\mathbf{ii}'}{T}\otimes\mathbf{D}_\ell\right)\boldsymbol{\varepsilon}\right] &= \frac{1}{(NT)^2}E\left(\sum_{i=1}^N\sum_{j=1}^N\sum_{h=1}^N\sum_{k=1}^N d_{\ell,ij}d_{\ell,hk}\boldsymbol{\varepsilon}_i'\frac{\mathbf{ii}'}{T}\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_h'\frac{\mathbf{ii}'}{T}\boldsymbol{\varepsilon}_k\right) \\
&= \frac{1}{(NT)^2}\sum_{i=1}^N\sum_{j=1}^N E\left(d_{\ell,ii}d_{\ell,jj}\boldsymbol{\varepsilon}_i'\frac{\mathbf{ii}'}{T}\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j'\frac{\mathbf{ii}'}{T}\boldsymbol{\varepsilon}_j + d_{\ell,ij}^2\boldsymbol{\varepsilon}_i'\frac{\mathbf{ii}'}{T}\boldsymbol{\varepsilon}_j\boldsymbol{\varepsilon}_j'\frac{\mathbf{ii}'}{T}\boldsymbol{\varepsilon}_i\right) \\
&\leq \frac{\sigma_{\max}^2}{(NT)^2}\sum_{i=1}^N\sum_{j=1}^N\left(d_{\ell,ii}d_{\ell,jj}+d_{\ell,ij}^2\right) = O\left(\frac{1}{NT^2}\right).
\end{aligned}
\tag{27}
$$

It follows that

$$
\frac{1}{NT}\hat{\boldsymbol{\varepsilon}}(\delta)'(\mathbf{I}_T\otimes\mathbf{A}_\ell)\hat{\boldsymbol{\varepsilon}}(\delta) = \frac{1}{NT}\boldsymbol{\varepsilon}(\delta)'(\mathbf{I}_T\otimes\mathbf{A}_\ell)\boldsymbol{\varepsilon}(\delta) + O_p\left(\frac{1}{T\sqrt{N}}\right) + O_p\left(\frac{1}{T}\right),
$$

which proves (8). To prove (9), notice that, at $\boldsymbol{\theta}_0$,

$$
\frac{1}{NT}\boldsymbol{\varepsilon}'(\mathbf{M}\otimes\mathbf{D}_\ell)\boldsymbol{\varepsilon} = \frac{1}{NT}\boldsymbol{\varepsilon}'(\mathbf{M}\otimes\mathbf{A}_\ell)\boldsymbol{\varepsilon} = \frac{1}{NT}\boldsymbol{\varepsilon}'(\mathbf{I}_T\otimes\mathbf{A}_\ell)\boldsymbol{\varepsilon} - \frac{1}{NT}\boldsymbol{\varepsilon}'\left(\frac{\mathbf{ii}'}{T}\otimes\mathbf{A}_\ell\right)\boldsymbol{\varepsilon}. \tag{28}
$$

Given that the elements $a_{\ell,ii}$ are zero by assumptions, the latter term in (28) satisfies

$$
\frac{1}{NT}E\left[\boldsymbol{\varepsilon}'\left(\frac{\mathbf{ii}'}{T}\otimes\mathbf{A}_\ell\right)\boldsymbol{\varepsilon}\right] = \frac{1}{NT}\sum_{i=1}^N a_{\ell,ii}E\left(\boldsymbol{\varepsilon}_i\frac{\mathbf{ii}'}{T}\boldsymbol{\varepsilon}_i\right) = 0, \tag{29}
$$

while its variance is $O\left(\frac{1}{NT^2}\right)$, as shown in (27). This proves (9).

6.1 Proof of Proposition 2

Consider the quadratic form

$$
\frac{1}{NT}\boldsymbol{\varepsilon}(\delta)'(\mathbf{I}_T\otimes\mathbf{A}_\ell)\boldsymbol{\varepsilon}(\delta) = \frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T\otimes\mathbf{P}(\delta)'\mathbf{A}_\ell\mathbf{P}(\delta)\right)\boldsymbol{\varepsilon} \tag{30}
$$

where $\mathbf{P}(\delta)$ is given by (25), and has uniformly bounded row and column norms. The mean of (30) satisfies (see Lemma 4)

$$
E\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T\otimes\mathbf{P}(\delta)'\mathbf{A}_\ell\mathbf{P}(\delta)\right)\boldsymbol{\varepsilon}\right] = \frac{1}{N}Tr\left(\boldsymbol{\Sigma}\mathbf{P}(\delta)'\mathbf{A}_\ell\mathbf{P}(\delta)\right) = O(1)
$$

Let $\mathbf{W}_\ell = \mathbf{P}(\delta)'\mathbf{A}_\ell\mathbf{P}(\delta)$ with elements $w_{ij,\ell}$, the variance of (30) satisfies

$$
\begin{aligned}
Var\left[\frac{1}{NT}\boldsymbol{\varepsilon}'\left(\mathbf{I}_T\otimes\mathbf{P}(\delta)'\mathbf{A}_\ell\mathbf{P}(\delta)\right)\boldsymbol{\varepsilon}\right] &= \frac{1}{N^2T}\sum_{i=1}^N w_{ii,\ell}^2\left[E\left(\varepsilon_{it}^4\right)-3\sigma_i^4\right] \\
+\frac{1}{N^2T}Tr\left[(\boldsymbol{\Sigma}\mathbf{W}_\ell)^2+\boldsymbol{\Sigma}\mathbf{W}_\ell\mathbf{W}_\ell'\boldsymbol{\Sigma}\right] &= O\left(\frac{1}{NT}\right)
\end{aligned}
$$

which proves (10).

6.2 Proof of Theorem 3

6.2.1 Consistency

We now sketch the proof of consistency of $\hat{\delta}$. See Kelejian and Prucha (2008), Lee and Liu (2006a), Lee (2007), and Kelejian and Prucha (1999) for further details on consistency of GMM estimators of spatial models. Consider the following functions

$$\begin{aligned} R(\delta) &= \mathbf{M}_{NT}(\delta)' \mathbf{Q}_{NT} \mathbf{M}_{NT}(\delta) \\ Z(\delta) &= \mathcal{M}(\delta)' \mathbf{Q} \mathcal{M}(\delta) \end{aligned}$$

Consistency of GMM estimator can be showed by proving the following two conditions:

1. *Identification uniqueness*: for all N, T, and for $K > 0$

$$\inf_{\delta:\|\delta-\delta_0\|_2 \geq K} |Z(\delta) - Z(\delta_0)| > 0$$

2. *Uniform convergence*:

$$\lim_{N,T\to\infty} \sup_{\delta\in[c_l,c_u]} |R(\delta) - Z(\delta)| = 0$$

To prove point 1, first note that, for $\|\delta - \delta_0\|_2 \geq K > 0$, under the identificability condition provided in Assumption 6 we have

$$\begin{aligned} \mathcal{M}_\ell(\delta) - \mathcal{M}_\ell(\delta_0) &= \frac{1}{NT} E\left[\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{P}(\delta)' \mathbf{A}_\ell \mathbf{P}(\delta)\right)\boldsymbol{\varepsilon}\right] - \frac{1}{NT} E\left[\boldsymbol{\varepsilon}'\left(\mathbf{I}_T \otimes \mathbf{A}_\ell\right)\boldsymbol{\varepsilon}\right] \\ &= \frac{2(\delta-\delta_0)}{NT} E\left\{\boldsymbol{\varepsilon}'\left[\mathbf{I}_T \otimes \left(\mathbf{A}_\ell \mathbf{S}(\mathbf{I}_N - \delta_0 \mathbf{S})^{-1}\right)\right]\boldsymbol{\varepsilon}\right\} \\ &\quad + (\delta-\delta_0)^2 E\left\{\boldsymbol{\varepsilon}'\left[\mathbf{I}_T \otimes \left(\left(\mathbf{I}_N - \delta_0 \mathbf{S}'\right)^{-1} \mathbf{S}' \mathbf{A}_\ell \mathbf{S}(\mathbf{I}_N - \delta_0 \mathbf{S})^{-1}\right)\right]\right\}\boldsymbol{\varepsilon} \\ &= \frac{2(\delta-\delta_0)}{N} \sigma_0^2 Tr\left[\boldsymbol{\Sigma} \mathbf{A}_\ell \mathbf{S}(\mathbf{I}_N - \delta_0 \mathbf{S})^{-1}\right] \\ &\quad + (\delta-\delta_0)^2 Tr\left[\left(\boldsymbol{\Sigma}\left(\mathbf{I}_N - \delta_0 \mathbf{S}'\right)^{-1} \mathbf{S}' \mathbf{A}_\ell \mathbf{S}(\mathbf{I}_N - \delta_0 \mathbf{S})^{-1}\right)\right] \\ &\neq 0. \end{aligned}$$

Consider $Z(\delta) - Z(\delta_0)$. Adding and subtracting the term $\mathcal{M}(\delta)' \mathbf{Q} \mathcal{M}(\delta_0)$ we obtain

$$\begin{aligned} |Z(\delta) - Z(\delta_0)| &= \left|\mathcal{M}(\delta)' \mathbf{Q}\left[\mathcal{M}(\delta) - \mathcal{M}(\delta_0)\right] - \left[\mathcal{M}(\delta_0) - \mathcal{M}(\delta)\right]' \mathbf{Q} \mathcal{M}(\delta_0)\right| \\ &= \left|\left[\mathcal{M}(\delta) - \mathcal{M}(\delta_0)\right]' \mathbf{Q}\left[\mathcal{M}(\delta) - \mathcal{M}(\delta_0)\right]\right| > 0, \end{aligned}$$

given that $|\mathcal{M}(\delta) - \mathcal{M}(\delta_0)| > 0$. Point 2 follows from the fact that, from (8) and (10), $\mathbf{M}_{NT}(\delta) - \mathcal{M}(\delta) \xrightarrow{p} 0$, as N and $T \to \infty$, for all $\delta \in [c_l, c_u]$.

6.2.2 Asymptotic normality

We now prove the asymptotic normality of the GMM estimator. Let

$$q_{NT}(\delta) = \mathbf{M}_{NT}(\hat{\delta})'\mathbf{Q}_{NT}\mathbf{M}_{NT}(\hat{\delta}) \tag{31}$$

The minimisation of (31) with respect to δ implies its first derivative at $\hat{\delta}$ is zero

$$\mathbf{0} = \frac{\partial q_N(\delta)}{\partial \delta} \mid_{\delta=\hat{\delta}} = 2\left(\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta'} \mid_{\delta=\hat{\delta}}\right)' \mathbf{Q}_{NT}\mathbf{M}_{NT}(\hat{\delta}). \tag{32}$$

where $\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta}$ has elements given by (15). Consider the mean value expansion of $\mathbf{M}_{NT}(\hat{\delta})$ around δ_0

$$\mathbf{M}_{NT}(\hat{\delta}) = \mathbf{M}_{NT}(\delta_0) + \frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta'} \mid_{\delta=\bar{\delta}} \left(\hat{\delta} - \delta_0\right), \tag{33}$$

where $\bar{\delta}$ lies between $\hat{\delta}$ and δ_0. Substituting (33) in (32) we obtain

$$\mathbf{0} = \left(\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta} \mid_{\delta=\hat{\delta}}\right)' \mathbf{Q}_{NT}\mathbf{M}_{NT}(\delta_0) + \left(\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta} \mid_{\delta=\hat{\delta}}\right)' \mathbf{Q}_{NT}\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta} \mid_{\delta=\bar{\boldsymbol{\theta}}} \left(\hat{\delta} - \delta_0\right).$$

Solving for $\left(\hat{\delta} - \delta_0\right)$ and multiplying by $\sqrt{NT}$ yields

$$\sqrt{NT}\left(\hat{\delta} - \delta_0\right) = -\left[\left(\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta} \mid_{\delta=\hat{\delta}}\right)' \mathbf{Q}_{NT}\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta} \mid_{\delta=\bar{\boldsymbol{\theta}}}\right]^{-1} \left(\frac{\partial \mathbf{M}_{N}(\delta)}{\partial \delta} \mid_{\delta=\hat{\delta}}\right)' \mathbf{Q}_{NT}\sqrt{NT}\mathbf{M}_{NT}(\delta_0).$$

Observe that, given the consistency of $\hat{\delta}$, as N and T tend to infinity we have

$$\left[\left(\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta} \mid_{\delta=\hat{\delta}}\right)' \mathbf{Q}_{NT}\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta} \mid_{\delta=\bar{\boldsymbol{\theta}}}\right]^{-1} \left(\frac{\partial \mathbf{M}_{NT}(\delta)}{\partial \delta} \mid_{\delta=\hat{\delta}}\right)' \mathbf{Q}_{NT} \xrightarrow{p} \left(\mathbf{d}'\mathbf{Q}\mathbf{d}\right)^{-1}\mathbf{d}'\mathbf{Q}.$$

Further, from result (9), the term $\sqrt{NT}\mathbf{M}_{NT}(\delta_0)$ converges to a vector of quadratic forms, with generic element $\frac{1}{\sqrt{NT}}\boldsymbol{\varepsilon}(\delta_0)'(\mathbf{I}_T \otimes \mathbf{A}_\ell)\boldsymbol{\varepsilon}(\delta_0)$ as N and T go to infinity, and having $\mathbf{V}(\delta_0)$ (see (11)) as covariance matrix. Since the hypotheses of Theorem 1 in Kelejian and Prucha (2001) (p. 227), and of Theorem A1 in Kelejian and Prucha (2008) (p. 25) are satisfied for the elements $\frac{1}{\sqrt{NT}}\boldsymbol{\varepsilon}(\delta_0)'(\mathbf{I}_T \otimes \mathbf{A}_\ell)\boldsymbol{\varepsilon}(\delta_0)$, the following result holds:

$$\mathbf{V}(\delta_0)^{-1/2}\left[\sqrt{NT}\cdot\mathbf{M}_{NT}(\delta_0)\right] \overset{a}{\sim} N(\mathbf{0}, \mathbf{I}_r), \quad \text{as } (N,T) \to \infty$$

Hence, we obtain

$$\sqrt{NT}\left(\hat{\delta} - \delta_0\right) \overset{a}{\sim} N\left(0, \left(\mathbf{d}'\mathbf{Q}\mathbf{d}\right)^{-1}\mathbf{d}'\mathbf{Q}\mathbf{V}\mathbf{Q}\mathbf{d}\left(\mathbf{d}'\mathbf{Q}\mathbf{d}\right)^{-1}\right).$$

6.3 The elements of $\mathbf{d}$

We now derive the elements of the vector $\mathbf{d}$, introduced in Theorem 3. First notice that

$$\begin{aligned}\frac{\partial}{\partial\delta}\boldsymbol{\varepsilon}\left(\delta\right) &= \left[\mathbf{I}_T\otimes\frac{\partial}{\partial\delta}\left(\mathbf{I}_N+\left(\delta_0-\delta\right)\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right)\right]\boldsymbol{\varepsilon}\\ &= -\left[\mathbf{I}_T\otimes\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right]\boldsymbol{\varepsilon}\end{aligned}$$

Hence, we have

$$\begin{aligned}p\lim\left[\frac{\partial}{\partial\delta}\mathbf{M}_{NT}\left(\delta\right)\right] &= -\frac{1}{NT}\boldsymbol{\varepsilon}\left(\delta\right)'\left[\mathbf{I}_T\otimes\mathbf{A}_\ell\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right]\boldsymbol{\varepsilon}-\frac{1}{NT}\boldsymbol{\varepsilon}\left[\mathbf{I}_T\otimes\left(\mathbf{I}_N-\delta_0\mathbf{S}'\right)^{-1}\mathbf{S}'\mathbf{A}_\ell\right]\boldsymbol{\varepsilon}\left(\delta\right)\\ &= -\frac{1}{NT}\boldsymbol{\varepsilon}'\left[\mathbf{I}_T\otimes\left(\mathbf{I}_N+\left(\delta_0-\delta\right)\left(\mathbf{I}_N-\delta_0\mathbf{S}'\right)^{-1}\mathbf{S}'\right)\mathbf{A}_\ell\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right]\boldsymbol{\varepsilon}\\ &\quad -\frac{1}{NT}\boldsymbol{\varepsilon}'\left[\mathbf{I}_T\otimes\left(\mathbf{I}_N-\delta_0\mathbf{S}'\right)^{-1}\mathbf{S}'\mathbf{A}_\ell\left(\mathbf{I}_N+\left(\delta_0-\delta\right)\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right)\right]\boldsymbol{\varepsilon}\end{aligned}$$

Thus, at δ_0,

$$\begin{aligned}\left.\frac{\partial}{\partial\delta}\mathbf{M}_{NT}\left(\delta\right)\right|_{\delta=\delta_0} &= \frac{\partial}{\partial\delta}M_{NT,\ell}\left(\delta_0\right)=-\frac{1}{NT}\boldsymbol{\varepsilon}'\left[\mathbf{I}_T\otimes\mathbf{A}_\ell\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right]\boldsymbol{\varepsilon}\\ &\quad -\frac{1}{NT}\boldsymbol{\varepsilon}'\left[\mathbf{I}_T\otimes\left(\mathbf{I}_N-\delta_0\mathbf{S}'\right)^{-1}\mathbf{S}'\mathbf{A}_\ell\right]\boldsymbol{\varepsilon}\end{aligned}$$

and the ℓth element of $\mathbf{d}$ is

$$\begin{aligned}d_\ell &= p\lim E\left[\left.\frac{\partial}{\partial\delta}\mathbf{M}_{NT}\left(\delta\right)\right|_{\delta=\delta_0}\right]=-\frac{1}{NT}E\left\{\boldsymbol{\varepsilon}'\left[\mathbf{I}_T\otimes\mathbf{A}_\ell\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right]\boldsymbol{\varepsilon}\right\}\\ &\quad -\frac{1}{NT}E\left\{\boldsymbol{\varepsilon}'\left[\mathbf{I}_T\otimes\left(\mathbf{I}_N-\delta_0\mathbf{S}'\right)^{-1}\mathbf{S}'\mathbf{A}_\ell'\right]\boldsymbol{\varepsilon}\right\}\\ &= -\frac{1}{N}Tr\left(\boldsymbol{\Sigma}\mathbf{A}_\ell\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right)-\frac{1}{N}Tr\left(\boldsymbol{\Sigma}\left(\mathbf{I}_N-\delta_0\mathbf{S}'\right)^{-1}\mathbf{S}'\mathbf{A}_\ell\right)\\ &= -\frac{1}{N}Tr\left(\boldsymbol{\Sigma}\left(\mathbf{A}_\ell+\mathbf{A}_\ell'\right)\mathbf{S}\left(\mathbf{I}_N-\delta_0\mathbf{S}\right)^{-1}\right).\end{aligned}$$

Table 2: Small sample properties of GMM and ML estimators. Q=Qstar. normal errors, p=1

N	T	$\hat{\delta}^{KP}$	$\hat{\delta}^{L}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(2)}$	$\hat{\delta}^{(3)}$	$\hat{\delta}_{ML}$	$\hat{\delta}^{KP}$	$\hat{\delta}^{L}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(2)}$	$\hat{\delta}^{(3)}$	$\hat{\delta}_{ML}$
							δ_0= 0.0						
		Bias						RMSE					
20	10	0.08	0.11	0.09	0.08	0.09	0.03	7.36	7.36	7.43	7.39	7.39	3.97
20	20	0.07	0.08	0.07	0.07	0.07	0.06	5.22	5.22	5.24	5.23	5.25	2.21
20	50	0.13	0.14	0.13	0.13	0.14	0.01	3.20	3.19	3.20	3.20	3.19	1.08
50	10	-0.06	-0.05	-0.07	-0.06	-0.06	-0.11	4.90	4.89	4.92	4.91	4.91	3.55
50	20	-0.11	-0.11	-0.11	-0.11	-0.11	0.02	3.34	3.34	3.35	3.34	3.34	2.30
50	50	-0.01	-0.01	-0.01	-0.01	-0.01	0.05	2.03	2.03	2.03	2.03	2.03	1.38
100	10	0.09	0.08	0.09	0.09	0.08	0.09	3.27	3.28	3.28	3.28	3.27	2.49
100	20	0.00	0.00	0.00	0.00	0.00	0.01	2.21	2.21	2.21	2.21	2.21	1.54
100	50	0.06	0.06	0.06	0.06	0.06	0.03	1.37	1.36	1.37	1.37	1.37	0.97
		Size						Power					
20	10	7.60	7.20	7.60	7.60	6.60	14.40	31.80	32.40	33.00	32.00	32.80	83.00
20	20	6.40	6.20	6.40	6.40	6.40	8.40	54.40	53.80	54.40	54.40	53.80	98.60
20	50	5.60	6.00	5.80	5.60	6.00	7.80	89.60	90.00	89.60	89.60	90.00	100.0
50	10	7.20	7.20	7.20	7.20	7.20	15.40	60.00	59.60	59.80	60.00	59.80	89.40
50	20	5.60	5.60	5.60	5.60	6.00	8.60	87.60	87.80	87.60	87.60	87.60	99.00
50	50	5.80	5.80	5.80	5.80	5.80	8.00	100.0	100.0	100.0	100.0	100.0	100.0
100	10	5.80	5.60	5.80	5.80	5.60	14.80	87.80	87.60	87.60	87.80	87.60	99.40
100	20	6.20	6.20	6.20	6.20	6.00	7.60	99.60	99.60	99.60	99.60	99.60	100.0
100	50	5.20	5.20	5.20	5.20	5.40	6.60	100.0	100.0	100.0	100.0	100.0	100.0
							δ_0= 0.3						
		Bias						RMSE					
20	10	-0.46	-0.44	-0.14	-0.30	-0.39	-0.60	6.90	6.86	6.93	6.90	6.92	3.88
20	20	-0.27	-0.25	-0.10	-0.18	-0.15	-0.31	4.86	4.84	4.86	4.85	4.92	2.16
20	50	-0.02	0.00	0.05	0.02	0.05	-0.16	2.92	2.91	2.92	2.91	2.97	1.06
50	10	-0.33	-0.34	-0.21	-0.27	-0.26	-0.36	4.54	4.55	4.55	4.54	4.58	3.33
50	20	-0.22	-0.23	-0.17	-0.20	-0.20	-0.10	3.10	3.10	3.10	3.10	3.13	2.16
50	50	-0.06	-0.06	-0.04	-0.05	-0.04	0.00	1.87	1.87	1.87	1.87	1.90	1.27
100	10	-0.02	-0.03	0.03	0.01	-0.02	-0.03	3.07	3.07	3.07	3.06	3.06	2.32
100	20	-0.05	-0.05	-0.03	-0.04	-0.04	-0.06	2.05	2.05	2.06	2.05	2.06	1.40
100	50	0.03	0.03	0.04	0.03	0.04	0.00	1.25	1.25	1.25	1.25	1.27	0.89
		Size						Power					
20	10	7.80	7.80	8.40	8.20	8.00	14.60	35.20	35.00	37.60	36.00	35.40	79.20
20	20	7.00	6.80	6.40	6.60	6.60	6.80	56.80	56.40	59.00	58.00	57.80	97.60
20	50	5.60	5.80	6.00	5.80	6.20	7.60	92.40	92.80	92.60	92.60	92.20	100.0
50	10	7.80	7.40	7.40	7.40	7.00	15.40	63.20	63.00	64.40	63.60	63.40	91.60
50	20	6.40	6.60	6.40	6.40	5.60	8.60	90.00	89.60	90.00	90.00	90.60	99.40
50	50	5.60	5.20	5.60	5.40	6.00	7.00	100.0	100.0	100.0	100.0	100.0	100.0
100	10	6.40	6.60	6.60	6.60	5.40	15.00	90.40	90.40	90.80	90.80	89.80	99.40
100	20	5.40	6.00	6.00	5.80	5.80	7.00	100.0	100.0	100.0	100.0	100.0	100.0
100	50	4.80	4.60	4.80	4.80	5.60	7.20	100.0	100.0	100.0	100.0	100.0	100.0
							δ_0= 0.8						
		Bias						RMSE					
20	10	-1.48	-1.54	-1.08	-1.29	-1.91	-1.84	3.78	3.74	3.65	3.66	4.43	3.58
20	20	-0.80	-0.82	-0.57	-0.69	-0.93	-1.07	2.49	2.41	2.34	2.35	2.88	1.90
20	50	-0.26	-0.27	-0.18	-0.22	-0.30	-0.45	1.37	1.30	1.28	1.28	1.59	0.86
50	10	-0.76	-0.77	-0.59	-0.69	-0.87	-0.37	2.35	2.32	2.28	2.29	2.63	5.07
50	20	-0.38	-0.37	-0.29	-0.34	-0.46	-0.30	1.53	1.47	1.45	1.46	1.73	2.10
50	50	-0.14	-0.15	-0.11	-0.13	-0.16	-0.14	0.88	0.86	0.85	0.86	1.00	0.61
100	10	-0.29	-0.29	-0.20	-0.25	-0.38	-0.31	1.53	1.47	1.45	1.45	1.67	2.02
100	20	-0.17	-0.16	-0.12	-0.10	-0.21	-0.22	1.00	0.96	0.96	1.28	1.09	0.71
100	50	-0.05	-0.06	-0.04	-0.05	-0.05	-0.09	0.59	0.58	0.58	0.58	0.66	0.41
		Size						Power					
20	10	8.40	8.00	8.80	8.00	9.60	17.80	73.00	75.60	75.00	72.80	61.80	88.80
20	20	7.80	7.80	7.20	7.80	7.40	13.20	95.00	97.00	96.80	96.00	91.40	99.40
20	50	5.20	5.00	4.40	4.20	6.20	9.20	100.0	100.0	100.0	100.0	100.0	100.0
50	10	8.60	9.00	8.60	8.00	9.00	21.00	98.20	98.60	98.20	98.20	95.20	99.80
50	20	7.00	7.00	6.40	6.80	6.40	11.80	100.0	100.0	100.0	100.0	99.60	100.0
50	50	5.60	5.40	5.00	5.00	5.80	5.60	100.0	100.0	100.0	100.0	100.0	100.0
100	10	7.60	7.40	7.00	6.80	6.80	18.00	100.0	100.0	100.0	100.0	100.0	100.0
100	20	6.40	6.00	6.60	6.20	6.40	10.00	100.0	100.0	100.0	100.0	100.0	100.0
100	50	5.00	5.00	5.00	5.20	5.40	8.00	100.0	100.0	100.0	100.0	100.0	100.0

Table 3: Small sample properties of GMM and ML estimators. $Q = I$, normal errors, p=1

N	T	$\hat{\delta}^L$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(2)}$	$\hat{\delta}^L$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(2)}$
				δ_0= 0.0			
		Bias			RMSE		
20	20	0.35	0.64	0.41	7.88	9.11	8.10
20	50	0.14	0.20	0.15	5.44	6.00	5.54
20	100	0.18	0.23	0.19	3.36	3.75	3.45
50	20	-0.01	0.06	0.01	5.23	5.85	5.37
50	50	-0.09	-0.03	-0.09	3.49	3.91	3.58
50	100	-0.03	-0.06	-0.04	2.21	2.47	2.28
100	20	0.15	0.21	0.17	3.52	3.98	3.63
100	50	-0.06	-0.08	-0.06	2.42	2.80	2.51
100	100	0.03	-0.01	0.02	1.47	1.70	1.51
		Size			Power		
20	20	9.60	13.00	10.60	33.00	35.00	34.00
20	50	8.00	11.40	9.20	55.60	57.20	56.60
20	100	7.60	9.40	7.80	89.60	86.60	89.20
50	20	9.40	14.00	10.60	61.20	60.40	60.00
50	50	6.40	10.20	7.20	88.00	85.40	87.00
50	100	7.80	11.40	8.40	99.80	98.60	99.60
100	20	7.80	12.60	9.40	86.20	84.00	86.00
100	50	8.20	11.00	8.00	98.80	97.40	98.00
100	100	7.00	10.40	7.40	100.0	100.0	100.0
				δ_0= 0.3			
		Bias			RMSE		
20	20	-0.14	0.17	-0.07	7.34	8.55	7.52
20	50	-0.09	-0.02	-0.08	5.09	5.60	5.19
20	100	0.09	0.14	0.10	3.14	3.51	3.21
50	20	-0.20	-0.11	-0.17	4.90	5.49	5.03
50	50	-0.17	-0.11	-0.17	3.27	3.70	3.36
50	100	-0.05	-0.08	-0.06	2.07	2.32	2.13
100	20	0.05	0.11	0.06	3.27	3.69	3.37
100	50	-0.08	-0.09	-0.09	2.25	2.59	2.33
100	100	0.01	-0.02	0.00	1.36	1.58	1.41
		Size			Power		
20	20	10.00	12.40	10.80	36.40	37.40	38.00
20	50	8.00	10.40	8.80	59.80	58.00	59.40
20	100	8.00	9.40	8.20	92.60	89.80	91.80
50	20	9.00	12.40	10.00	65.20	62.20	64.20
50	50	7.20	9.60	8.40	89.00	87.20	88.20
50	100	7.60	11.00	8.40	100.0	99.20	100.0
100	20	7.80	11.40	8.80	89.20	86.20	89.00
100	50	8.40	10.40	8.80	99.40	98.40	99.40
100	100	6.80	9.20	7.40	100.0	100.0	100.0
				δ_0= 0.8			
		Bias			RMSE		
20	20	-1.93	-1.88	-1.86	4.72	5.31	4.76
20	50	-1.03	-0.96	-1.00	2.91	3.22	2.92
20	100	-0.25	-0.21	-0.24	1.67	1.82	1.65
50	20	-0.79	-0.75	-0.76	2.62	3.08	2.65
50	50	-0.52	-0.52	-0.50	1.78	2.05	1.79
50	100	-0.20	-0.23	-0.20	1.09	1.28	1.11
100	20	-0.47	-0.48	-0.46	1.79	2.08	1.82
100	50	-0.28	-0.29	-0.27	1.19	1.39	1.20
100	100	-0.11	-0.13	-0.12	0.74	0.85	0.75
		Size			Power		
20	20	16.20	18.80	16.60	66.40	64.80	66.80
20	50	12.80	13.40	11.80	94.20	92.80	93.40
20	100	12.00	12.20	10.80	100.0	100.0	100.0
50	20	14.20	15.20	14.00	96.60	94.40	96.40
50	50	13.40	15.60	12.40	100.0	99.20	100.0
50	100	11.20	14.40	11.60	100.0	100.0	100.0
100	20	12.60	15.80	12.60	100.0	99.20	100.0
100	50	10.80	13.20	11.40	100.0	100.0	100.0
100	100	12.20	14.40	13.00	100.0	100.0	100.0

Table 4: Small sample properties of GMM and ML estimators. $Q = QStar$, non-normal errors, p=1

N	T	$\hat{\delta}^{KP}$	$\hat{\delta}^{L}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(2)}$	$\hat{\delta}^{(3)}$	$\hat{\delta}_{ML}$	$\hat{\delta}^{KP}$	$\hat{\delta}^{L}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(2)}$	$\hat{\delta}^{(3)}$	$\hat{\delta}_{ML}$
							δ_0= 0.0						
		Bias						RMSE					
20	10	0.04	0.00	0.04	0.04	0.01	0.21	7.49	7.45	7.57	7.53	7.43	6.64
20	20	-0.29	-0.29	-0.29	-0.29	-0.28	-0.31	5.21	5.20	5.23	5.22	5.21	4.78
20	50	0.12	0.12	0.12	0.12	0.12	0.10	3.37	3.37	3.37	3.37	3.38	3.23
50	10	0.00	0.01	0.00	0.00	0.00	0.05	4.59	4.58	4.61	4.60	4.59	3.91
50	20	-0.15	-0.16	-0.15	-0.15	-0.16	-0.10	3.17	3.17	3.18	3.18	3.17	2.98
50	50	-0.08	-0.08	-0.08	-0.08	-0.08	-0.06	1.99	1.98	1.99	1.99	1.98	1.86
100	10	-0.08	-0.08	-0.08	-0.08	-0.08	0.03	3.26	3.25	3.26	3.26	3.25	2.66
100	20	-0.13	-0.13	-0.13	-0.13	-0.13	-0.19	2.19	2.19	2.19	2.19	2.18	2.08
100	50	-0.03	-0.03	-0.03	-0.03	-0.03	0.02	1.41	1.42	1.41	1.41	1.42	1.37
		Size						Power					
20	10	8.20	7.60	8.40	8.40	7.60	13.60	33.60	33.60	33.80	33.60	33.60	48.60
20	20	7.00	7.20	7.00	7.00	7.20	7.20	50.40	50.20	50.40	50.40	49.80	62.60
20	50	7.40	7.60	7.40	7.40	7.60	7.40	88.40	88.20	88.40	88.40	88.40	90.40
50	10	6.80	6.40	7.00	6.80	6.60	12.80	61.20	61.60	61.40	61.40	61.80	83.00
50	20	4.60	4.60	4.60	4.60	4.80	7.60	89.40	89.20	89.40	89.40	89.00	94.40
50	50	4.80	4.60	4.80	4.80	4.60	4.40	100.0	100.0	100.0	100.0	100.0	100.0
100	10	6.00	6.00	6.00	6.00	5.80	10.40	89.20	89.20	89.00	89.20	89.40	98.40
100	20	4.20	4.00	4.20	4.20	3.80	6.20	99.40	99.40	99.40	99.40	99.40	99.20
100	50	5.20	5.00	5.20	5.20	5.60	5.40	100.0	100.0	100.0	100.0	100.0	100.0
							δ_0= 0.3						
		Bias						RMSE					
20	10	-0.46	-0.57	-0.20	-0.33	-0.46	-0.37	6.87	6.83	6.93	6.90	6.94	6.07
20	20	-0.53	-0.57	-0.39	-0.46	-0.48	-0.59	4.72	4.74	4.72	4.72	4.85	4.39
20	50	-0.01	-0.02	0.05	0.02	0.02	-0.05	3.11	3.10	3.11	3.10	3.14	2.97
50	10	-0.22	-0.25	-0.12	-0.17	-0.19	-0.20	4.28	4.26	4.28	4.28	4.28	3.65
50	20	-0.22	-0.23	-0.16	-0.19	-0.24	-0.19	2.91	2.91	2.92	2.91	2.96	2.74
50	50	-0.11	-0.12	-0.09	-0.10	-0.11	-0.09	1.82	1.81	1.82	1.81	1.85	1.70
100	10	-0.14	-0.15	-0.08	-0.11	-0.16	-0.08	3.00	2.99	3.00	2.99	3.03	2.46
100	20	-0.12	-0.12	-0.09	-0.10	-0.16	-0.20	2.00	2.00	2.00	2.00	2.04	1.92
100	50	-0.05	-0.06	-0.04	-0.05	-0.04	-0.02	1.30	1.30	1.30	1.30	1.32	1.26
		Size						Power					
20	10	7.60	7.20	8.20	7.60	7.40	14.00	36.40	35.00	37.40	37.00	35.60	51.00
20	20	6.20	5.80	6.00	6.00	7.20	7.40	53.80	53.20	54.60	53.60	53.40	65.40
20	50	7.80	7.80	8.00	8.00	7.60	7.20	90.20	90.00	90.40	90.40	89.80	92.40
50	10	6.00	6.00	6.40	6.20	6.20	12.20	66.60	66.80	67.80	67.40	65.20	85.20
50	20	4.60	4.60	4.60	4.60	4.80	7.20	93.40	93.40	93.80	93.60	92.00	96.40
50	50	5.20	5.00	5.00	5.20	4.60	4.40	100.0	100.0	100.0	100.0	100.0	100.0
100	10	5.40	5.20	5.60	5.40	6.00	11.60	91.40	91.20	91.80	91.60	91.00	98.80
100	20	4.40	4.80	4.60	4.60	3.80	6.20	99.40	99.40	99.40	99.40	99.40	99.60
100	50	4.40	4.40	4.40	4.40	5.20	4.40	100.0	100.0	100.0	100.0	100.0	100.0
							δ_0= 0.8						
		Bias						RMSE					
20	10	-1.48	-1.64	-1.21	-1.39	-1.90	-1.69	3.67	3.67	3.55	3.58	4.34	3.63
20	20	-0.86	-0.91	-0.69	-0.80	-1.04	-0.99	2.39	2.36	2.28	2.31	2.76	2.29
20	50	-0.26	-0.29	-0.19	-0.24	-0.31	-0.33	1.47	1.40	1.39	1.39	1.67	1.36
50	10	-0.66	-0.71	-0.53	-0.62	-0.81	-0.76	2.18	2.13	2.09	2.11	2.42	2.02
50	20	-0.34	-0.35	-0.25	-0.30	-0.49	-0.38	1.41	1.40	1.39	1.39	1.64	1.34
50	50	-0.14	-0.16	-0.13	-0.14	-0.18	-0.14	0.85	0.83	0.83	0.83	0.98	0.79
100	10	-0.30	-0.31	-0.20	-0.26	-0.45	-0.37	1.44	1.37	1.36	1.36	1.65	1.26
100	20	-0.15	-0.15	-0.10	-0.13	-0.28	-0.22	0.96	0.94	0.93	0.93	1.11	0.91
100	50	-0.13	-0.13	-0.08	-0.11	-0.25	-0.21	0.94	0.93	0.92	0.92	1.08	0.90
		Size						Power					
20	10	6.80	7.60	8.40	8.00	8.80	18.20	72.20	73.80	76.60	75.40	61.40	82.60
20	20	6.60	5.60	5.40	5.40	8.60	10.20	95.80	97.20	97.40	97.40	92.80	99.00
20	50	7.60	7.40	8.20	7.40	8.40	6.20	100.0	100.0	100.0	100.0	100.0	100.0
50	10	6.80	8.20	7.60	7.60	7.60	15.60	98.60	99.00	99.00	99.00	95.60	100.0
50	20	5.00	5.80	5.80	5.60	6.40	8.80	100.0	100.0	100.0	100.0	100.0	100.0
50	50	5.00	4.40	4.80	4.40	5.00	4.80	100.0	100.0	100.0	100.0	100.0	100.0
100	10	5.60	5.40	5.80	5.40	6.20	14.00	100.0	100.0	100.0	100.0	100.0	100.0
100	20	4.80	4.20	4.40	4.60	4.60	7.20	100.0	100.0	100.0	100.0	100.0	100.0
100	50	4.80	4.40	4.60	4.80	4.20	7.40	100.0	100.0	100.0	100.0	100.0	100.0

Table 5: Small sample properties of GMM and ML estimators. Use of real-world matrices $Q = Q$, $\delta = 0.3$, normal errors, p=1

		$\hat{\delta}^{KP}$	$\hat{\delta}^{L}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(2)}$	$\hat{\delta}^{(3)}$	$\hat{\delta}_{ML}$	$\hat{\delta}^{KP}$	$\hat{\delta}^{L}$	$\hat{\delta}^{(1)}$	$\hat{\delta}^{(2)}$	$\hat{\delta}^{(3)}$	$\hat{\delta}_{ML}$
N	T	Bias						RMSE					
12	10	-1.19	-1.40	-0.84	-0.42	-1.18	-0.90	9.50	9.47	9.64	9.58	9.47	8.45
12	20	-0.60	-0.68	-0.44	-0.30	-0.54	-0.51	6.86	6.85	6.89	6.86	6.87	5.55
12	50	-0.38	-0.42	-0.30	-0.24	-0.36	-0.21	3.93	3.92	3.94	3.95	3.97	3.17
32	10	-0.22	-0.50	0.00	0.19	-0.36	-0.64	7.29	7.30	7.34	7.34	7.42	4.68
32	20	-0.52	-0.64	-0.40	-0.34	-0.53	-0.29	5.15	5.16	5.16	5.14	5.24	2.99
32	50	-0.18	-0.21	-0.12	-0.10	-0.17	-0.13	2.97	2.98	2.98	2.98	2.99	1.70
148	10	-0.11	-0.19	-0.08	-0.03	-0.15	-0.15	3.56	3.54	3.55	3.55	3.55	2.33
148	20	0.02	-0.01	0.04	0.06	0.02	-0.04	2.42	2.43	2.42	2.42	2.43	1.52
148	50	-0.06	-0.09	-0.06	-0.05	-0.06	-0.11	1.67	1.68	1.67	1.67	1.67	1.06
		Size						Power					
12	10	7.60	7.20	8.40	9.00	7.60	13.00	23.20	21.60	23.60	25.60	23.00	35.40
12	20	6.40	6.40	7.20	7.00	6.20	7.40	37.00	37.60	39.00	39.40	37.20	51.60
12	50	5.60	6.00	5.60	5.60	5.60	5.20	67.40	67.80	68.00	68.60	68.00	87.40
32	10	6.60	6.80	7.00	7.80	8.20	10.60	34.40	32.80	35.20	36.00	34.20	66.80
32	20	6.00	6.40	6.20	6.00	6.80	8.00	51.60	50.00	52.60	53.00	50.20	91.40
32	50	4.60	4.60	4.80	4.60	3.80	6.80	89.80	89.20	89.40	90.00	89.60	100.0
148	10	5.40	5.00	5.60	5.40	5.40	14.80	80.60	80.00	81.00	81.00	80.00	99.60
148	20	5.60	5.60	5.80	5.60	5.60	7.80	97.40	97.20	97.40	97.40	97.40	100.0
148	50	4.80	5.20	5.20	5.20	5.20	7.60	99.60	99.60	99.60	99.60	99.60	100.0