

Alternative GMM Estimators for Spatial Regression Models

Jörg Breitung^a and Christoph Wigger^b

24th July 2017

Abstract

Using approximations of the score of the log-likelihood function we derive moment conditions for estimating spatial regression models, starting with the spatial error model. Our approach results in computationally simple and robust estimators such as a new moment estimator derived from the first order approximation that is obtained from solving a quadratic moment equation and performs similarly to existing GMM estimators. As our estimator based on the second order approximation resembles the GMM estimator proposed by Kelejian & Prucha (1999), we provide an intuitive interpretation of their estimator. Additionally, we provide a convenient framework for computing the weighting matrix of the optimal GMM estimator. Heteroskedasticity robust versions of our estimators are also provided. Furthermore, a first order approximation for the spatial autoregressive model is considered, resulting in a computationally simple method of moment estimator.

Keywords: Spatial Econometrics, Spatial error correlation, GMM-estimation

JEL Classification Numbers: C01, C13, C31.

^aCenter of Econometrics and Statistics, University of Cologne, Address: Albertus Magnus Platz, D-50923 Cologne, Germany, telephone: +49 221 470 4266, e-mail: breitung@stat.uni-koeln.de.

^bDepartment of Economics, University of Cologne, Address: Albertus Magnus Platz, D-50923 Cologne, Germany, telephone: +49 221 470 8660, e-mail: wigger@wiso.uni-koeln.de, <http://www.ieam.uni-koeln.de>.

We would like to thank the editor-in-chief, Paul Elhorst, and two anonymous referees for their helpful comments and suggestions. Christoph Wigger gratefully acknowledges financial support provided by the German Research Foundation (DFG) in the context of the Priority Programme 1764 "The German Labor Market in a Globalized World: Challenges through Trade, Technology, and Demographics".

1 Introduction

Spatial regression models have gained increasing popularity in applied economics during the last three decades, for example, when estimating regional spillovers and peer effects among economic agents.¹ As Anselin (2010) puts it, the field of spatial econometrics has moved “from the margins in applied urban and regional science to the mainstream of economics and other social sciences”.

In the past 20 years much progress has been made in the estimation of spatial models. Besides a well developed Maximum Likelihood (ML) estimation framework (see e.g. Anselin & Rey 2014) and a Bayesian framework (see LeSage & Pace 2009), Generalized Method of Moments (GMM) estimators have been developed, starting with the pioneering work of Kelejian & Prucha (1998, 1999).² Initially the main motivation for the derivation of GMM estimators for spatial models has been its computational simplicity. In particular, the inversion of $n \times n$ matrices is avoided, which can be computationally demanding whenever the sample size n is large. Although the inversion of large matrices is no longer a severe problem with nowadays computer technologies it may hinder the application of simulation methods such as the bootstrap. Furthermore, GMM estimators require weaker distributional assumptions than the traditional ML estimators and are robust to (unknown) heteroskedasticity and deviations from normality (see Anselin 1988, Kelejian & Prucha 1999, Lin & Lee 2010). It should be noted however that quasi-ML estimators have been developed allowing for more general distributional assumptions (e.g. Lee 2004). Another advantage is that GMM methods can easily be adapted for estimating models with endogenous covariates or lagged

¹ Some recent applications are Lin (2010), Piras et al. (2012), de Dominicis et al. (2013), Kelejian et al. (2013), Miguélez & Moreno (2013), and Brady (2014). For a list of applications from 1991 until 2007 see Kelejian & Prucha (2010*b*).

² Important contributions to the development of the GMM framework include Kelejian & Prucha (2001, 2004, 2007, 2010*a,b*), Lee (2002, 2003, 2004, 2007*a,b*), Kelejian et al. (2004), Kapoor et al. (2007), Fingleton (2008*a,b*), Fingleton & Le Gallo (2008*a,b*), Arnold & Wied (2010), Arraiz et al. (2010), Elhorst (2010*b*), Lee & Liu (2010), Lin & Lee (2010), Liu & Lee (2010, 2013), Liu et al. (2010), Baltagi & Liu (2011), Drukker et al. (2013), Wang & Lee (2013), Kelejian & Piras (2014), Lee & Yu (2014, 2016), Qu & Lee (2015), and Qu et al. (2016).

dependent variables (see Fingleton & Le Gallo 2008 a,b and Elhorst 2010 a).

In this paper we point out a close relationship between the ML and GMM estimation methods. We argue that the moment conditions of the GMM estimator proposed by Kelejian & Prucha (1999) can be interpreted as a second order approximation of the ML scores. Accordingly, the GMM estimator is expected to perform similar to the ML estimator, which is typically observed in empirical practice. We also derive a just-identified method of moment estimator based on a first order approximation of the scores. The resulting estimator is very simple to compute by just solving a quadratic equation. Our Monte Carlo simulations suggest that this estimator is nearly as efficient as the usual GMM estimator based on a second order approximation.

We first focus on GMM estimation of regression models with exogenous regressors and spatial error correlation as introduced by Cliff & Ord (1973), often referred to as the spatial error model (SEM) model. We then extend the GMM methodology to models that accommodate spatial spillover effects. Specifically the proposed estimators can be employed for estimating the spatial autoregressive (SAR) and the spatial Durbin model (SDM). Another possible extension is the estimation of spatial panel data models, where the computational simplicity of our estimators should be particularly beneficial (see Kapoor et al. 2007).³

The main contribution of this paper is the idea to derive moment conditions by applying approximations to the first order condition of the ML estimator. The resulting estimators are labelled the first or second order *Maximum Likelihood Approximate Moment* (MLAM) estimators. The first order MLAM estimator is the simplest available estimator of the spatial autoregressive parameter that results from solving a quadratic moment equation. Since the estimator is based on just identified moment conditions, no weighting matrix is required. Furthermore the estimator is robust to

³ Elhorst (2010 a) provides a comprehensive discussion and comparison of different types of spatial regression models, whereas spatial panel models are analyzed in Elhorst (2003), Kapoor et al. (2007), and Elhorst (2012).

heteroscedastic and non-normal error distributions. For the second order MLAM estimator we obtain moment conditions that resemble the ones proposed by Kelejian & Prucha (1999). Hence, we argue that their estimator can be interpreted as an approximation of the score of the log-likelihood function. Second, we derive efficient GMM estimators with computationally simplified optimal weighting matrices. Following the approach of Kelejian & Prucha (2010*b*) as well as Lin & Lee (2010) we also propose heteroskedasticity-robust versions of all our estimators. Third, we carry out Monte Carlo simulations in order to investigate the performance of alternative estimators under different sample sizes as well as both, homoskedastic and heteroskedastic errors. As expected, the efficient GMM estimator based on our moment conditions performs similarly to the one using the moments of Kelejian & Prucha (2010*b*). Moreover, the results suggest that the extremely simple first order MLAM estimator performs only slight worse than the optimal estimators based on over-identified moment conditions.

The rest of the paper is structured as follows. In section 2 we present the spatial autoregressive model and outline the GMM approach of Kelejian & Prucha (1999). Approximate moment conditions underlying our MLAM estimators, as well as their asymptotic distribution are derived in section 3. In section 4 we propose simplified efficient GMM estimators, whose heteroskedasticity robust modifications are presented in section 5. GMM estimation of the spatial autoregressive model is considered in section 6. Some Monte Carlos results on the performance of the proposed estimators are discussed in section 7 and section 8 concludes.

2 The spatial error model (SEM)

We first focus on the linear regression model with spatially correlated errors given by

$$y = X\beta + u \quad (1)$$

$$u = \rho_0 M_n u + \varepsilon, \quad (2)$$

where the vector $y = [y_1, \dots, y_n]'$ comprises the observations of the dependent variable, X is a $n \times k$ matrix of exogenous regressors and u is the $n \times 1$ vector of disturbances with $\mathbb{E}(u|X) = \mathbb{E}(u) = 0$. The spatial dependence among the elements of the error vector u is represented by the spatial error model (2). In what follows, let $B_n(\rho) \equiv (I_n - \rho M_n)$, where I_n denotes the $n \times n$ identity matrix. The model assumptions can be summarized as follows:

Assumption 1. (a) The regressor matrix is strictly exogenous with $\mathbb{E}(u|X) = 0$ and (b) $n^{-1}X'X$ has full rank for all n .

Assumption 2. The elements ε_i of the vector ε are i.i.d. with $\mathbb{E}(\varepsilon_i) = 0$, $\mathbb{E}(\varepsilon_i^2) = \sigma_0^2 > 0$, and $\mathbb{E}(|\varepsilon_i|^{4+\delta}) < \infty$ for some positive constant δ .

Assumption 3. (a) The spatial weight matrix M_n has zeros on the leading diagonal. (b) $|\rho_0| < 1$. (c) The matrix $B_n(\rho)$ is non-singular for $|\rho| < 1$.

Assumption 3 is a normalization rule. Part (c) implies that M_n is normalized accordingly. For example, consider the non-normalized weight matrix M_n^r . Then, as noted by Kelejian & Prucha (2010b), assumption 3(c) will hold for $B_n(\rho) = (I_n - \rho M_n)$, where $M_n = M_n^r / \tau_n$, $\tau_n = \min\{\max_{1 \leq i \leq n} \sum_{j=1}^n |m_{ij}^r|, \max_{1 \leq j \leq n} \sum_{i=1}^n |m_{ij}^r|\}$, and where m_{ij}^r indicates the (i, j) -element of matrix M_n^r . Alternatively, assumption 3(c) will also hold if M_n is normalized such that each row sums to unity, as often done in empirical applications (see Kelejian & Prucha 2010b). In our Monte Carlo simulation we also use such a row-normalized weight matrix.

Assumption 4. (a) The row and column sums of the matrix M_n are bounded uniformly in absolute value. (b) The row and column sums of the matrix $B_n(\rho)^{-1}$ are bounded uniformly in absolute value.

Together assumptions 2 and 4, imply that the mean of $u = B(\rho_0)^{-1}\varepsilon$ is $\mathbb{E}(u) = 0$ and that the variance-covariance matrix $\mathbb{E}[uu'] = \sigma^2 B(\rho_0)^{-1}(B(\rho_0)')^{-1}$ is uniformly bounded, thus limiting the degree of correlation between elements of u (see Kelejian & Prucha 2010b). Assumptions 1 to 4 are standard in the literature on GMM estimation of the linear spatial error model (e.g. Kelejian & Prucha 1999, Lee 2003). If the weight matrix M_n is row-normalized and, in addition, all elements of M_n are non-negative, which is usually the case in empirical applications, assumptions 3 (c) and 4 (a) will hold as noted by Lee (2003).

If in model (1) β is unknown, i.e. u is unobserved, we need the following additional assumption for the identification of ρ_0 :

Assumption 5. Let $\hat{\beta}$ be an initial estimate of β in (1). $\hat{\beta}$ is estimated by a consistent estimator of β with $\hat{\beta} - \beta = O_p(n^{-\frac{1}{2}})$.

An estimator that satisfies assumption 5 is the OLS estimator of model (1). Given $\hat{\beta}_{OLS}$, one can apply any of the estimators discussed below on the residuals $\hat{u}_{OLS} = y - X\hat{\beta}_{OLS} = u - X(\hat{\beta}_{OLS} - \beta)$ in order to obtain a consistent estimate $\hat{\rho}$ of the spatial autoregressive parameter ρ_0 . A consistent and asymptotically efficient estimate of β can then be obtained by the two-step GLS estimator

$$\hat{\beta}_{gl_s,n} = [X'(I_n - \hat{\rho}M_n')(I_n - \hat{\rho}M_n)X]^{-1} X'(I_n - \hat{\rho}M_n')(I_n - \hat{\rho}M_n)y.$$

According to assumption 5 we have $\hat{u} = u + O_p(n^{-1/2})$. Thus, for simplicity we treat β , and hence u , as known in the sections on estimating the spatial error model. The

log-likelihood function of the model results as

$$\ell(\rho, \sigma^2) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \sigma^2 + \ln |B_n(\rho)| - \frac{1}{2\sigma^2} u' B_n(\rho)' B_n(\rho) u. \quad (3)$$

Taking the first order derivative with respect to ρ and concentrating out σ^2 yields the first order condition for maximizing the log likelihood

$$\mathbb{E} \left[u' B_n(\rho_0)' \left(M_n B_n(\rho_0)^{-1} - \frac{1}{n} \text{tr} \{ M_n B_n(\rho_0)^{-1} \} I_n \right) B_n(\rho_0) u \right] = 0. \quad (4)$$

Note that the first order condition involves the inverse of the $n \times n$ matrix $B_n(\rho_0)$ which may become computationally demanding for sample sizes typically encountered in empirical practice. To sidestep this difficulty Kelejian & Prucha (1999) propose a method of moments approach for estimating ρ and σ^2 based on the three moment conditions

$$\mathbb{E} \left(\frac{1}{n} \varepsilon' \varepsilon \right) = \sigma_0^2, \quad \mathbb{E} \left(\frac{1}{n} \bar{\varepsilon}' \bar{\varepsilon} \right) = \sigma_0^2 \frac{1}{n} \text{tr} (M_n' M_n), \quad \mathbb{E} \left(\frac{1}{n} \bar{\varepsilon}' \varepsilon \right) = 0, \quad (5)$$

where $\varepsilon = u - \rho_0 \bar{u}$, $\bar{\varepsilon} = \bar{u} - \rho_0 \bar{\bar{u}}$, $\bar{u} = M_n u$ and $\bar{\bar{u}} = M_n \bar{u}$. Substituting the first of these moment conditions into the second and rewriting $\varepsilon = B_n(\rho_0) u$ yields the quadratic moment conditions

$$m_{1,n}(\rho_0) = \mathbb{E} \left[\frac{1}{n} u' B_n(\rho_0)' M_n B_n(\rho_0) u \right] = 0 \quad (6)$$

$$\text{and } \bar{m}_{2,n}(\rho_0) = \mathbb{E} \left[\frac{1}{n} u' B_n(\rho_0)' \bar{M}_n B_n(\rho_0) u \right] = 0, \quad (7)$$

where $\bar{M}_n = M_n' M_n - \text{tr}(n^{-1} M_n' M_n) I_n$. In Section 3 we analyze how these moment conditions are related to the score of the log-likelihood function (i.e. the first order condition for maximizing the log-likelihood). As outlined in Prucha (2014) it can be shown that minimizing the unweighted objective function based on the moment con-

ditions in (5) is equivalent to minimizing $S_{KP} = m_n(\rho)' W_{KP,n} m_n(\rho)$ with $m_n(\rho) = [m_{1,n}(\rho), \bar{m}_{2,n}(\rho)]'$ and the weighting matrix $W_{KP,n} = \text{diag}\{1, v\}$ with $v = 1/[1 + (n^{-1} \text{tr}\{M_n' M_n\})^2]$. We refer to this estimator as the original KP estimator in the following. Note that $W_{KP,n}$ is not the optimal weighting matrix. From the theory of GMM estimation we know that an estimator minimizing the objective function

$$S_{W,n}(\rho) = m_n(\rho)' Y_n(\rho_0) m_n(\rho) \quad \text{with } Y_n(\rho_0) = \left[\mathbb{E} \left(\frac{1}{n} m_n(\rho_0) m_n(\rho_0)' \right) \right]^{-1} \quad (8)$$

is asymptotically efficient. In Section 4 we propose a simple representation of the moment conditions that allows us to easily estimate the optimal weighting matrix $Y_n(\rho_0)$.

3 MLAM estimators

If ε is normally distributed, the efficient moment condition for estimating the parameter ρ_0 is the score of the log-likelihood function given in (4).

Under assumptions 3 and 4 we have

$$B_n(\rho_0)^{-1} = (I - \rho_0 M_n)^{-1} = I_n + \rho_0 M_n + \rho_0^2 M_n^2 + \dots \quad (9)$$

Using this expansion, truncating it after the first term such that $B_n(\rho)^{-1} \approx I_n$ (e.g. by assuming $\rho_0 \approx 0$) and dividing by n (4) yields the moment condition

$$\begin{aligned} m_{1,n}(\rho_0) &= \mathbb{E} \left[\frac{1}{n} u' B_n(\rho_0)' \left(M_n - \frac{1}{n} \text{tr}\{M_n\} I_n \right) B_n(\rho_0) u \right] \\ &= \mathbb{E} \left(\frac{1}{n} u' B_n(\rho_0)' M_n B_n(\rho_0) u \right) = 0, \end{aligned} \quad (10)$$

which clearly holds under assumptions 2 and 3 (a). Replacing u by its empirical counterpart \hat{u} this moment condition is sufficient to identify the spatial autoregressive parameter ρ_0 such that it is possible to construct a simple method of moments estimator.

In what follows we call the resulting estimator the *First Order Maximum Likelihood Approximate Moment* (MLAM1) estimator. The moment condition in (10) can be rewritten as a quadratic polynomial in ρ with the two roots

$$\begin{aligned}\hat{\rho}_+ &= p_n + \sqrt{q_n} \\ \text{and } \hat{\rho}_- &= p_n - \sqrt{q_n},\end{aligned}$$

where

$$\begin{aligned}p_n &= \frac{u' M_n' M_n u + u' M_n^2 u}{2 u' M_n' M_n^2 u} \\ q_n &= \left(\frac{u' M_n' M_n u + u' M_n^2 u}{2 u' M_n' M_n^2 u} \right)^2 - \frac{u' M_n u}{u' M_n' M_n^2 u}.\end{aligned}$$

The MLAM1 estimator $\hat{\rho}_1$ is the root that satisfies assumption 3 (b), where in the above expressions u is replaced by \hat{u} , yielding \hat{p}_n and \hat{q}_n . Since \hat{q}_n may be negative, it is possible that no real solution exists. In this case the unique estimator is found by setting q_n equal to zero. This is equivalent to minimizing the squared moment $\hat{m}_{1,n}(\rho)^2$ (the empirical counterpart of $m_{1,n}(\rho)^2$, where in (10) u is replaced by \hat{u}) because the moment function is symmetric around the minimum.⁴

In order to improve the approximation of (4) we may truncate the expansion (9) at the second term such that $B_n(\rho)^{-1} \approx I_n + \rho M_n$ yielding the *Second Order Maximum Likelihood Approximate Moment* (MLAM2) estimator based on the moment condition

$$m_{2,n}(\rho_0) = \mathbb{E} \left[\frac{1}{n} \varepsilon' \left(M_n (I_n + \rho_0 M_n) - \frac{1}{n} \text{tr} \{ (I_n + \rho_0 M_n) M_n \} I_n \right) \varepsilon \right] = 0.$$

Under assumptions 2 and 3 (a) this condition holds, since in the case of homoskedas-

⁴ Due to this symmetry, for any pair of values with $m_n(\rho_1) = m_n(\rho_2)$ it follows that the minimum is obtained as $m_n \left(\frac{\rho_1 + \rho_2}{2} \right)$. Since in our case ρ_1 and ρ_2 are complex conjugate roots with $m_n(\rho_1) = m_n(\rho_2) = 0$, it follows that $(\rho_1 + \rho_2)/2$ is just the real part of the two solutions.

tic disturbances $\mathbb{E} \left[\varepsilon' \left(\text{diag}\{\rho_0 M_n^2\} - \frac{1}{n} \text{tr}\{\rho_0 M_n^2\} I_n \right) \varepsilon \right] = 0$, where $\text{diag}\{\rho_0 M_n^2\}$ represents a diagonal matrix constructed by the diagonal elements of the matrix $\rho_0 M_n^2$. Moreover, it is easy to show that the moment condition can be rewritten as

$$m_{2,n}(\rho_0) = \mathbb{E} \left[\frac{1}{n} \varepsilon' (M_n + \rho_0 \tilde{M}_n) \varepsilon \right] = 0, \quad (11)$$

where $\tilde{M}_n \equiv M_n^2 - n^{-1} \text{tr}(M_n^2) I_n$. Note that for a symmetric spatial weight matrix $M_n = M_n'$ the moment is just a linear combination of the moments suggested by Kelejian & Prucha (1999).⁵ In the case that M_n is asymmetric, which is usually the case if the row sums of M_n are normalized, MLAM2 exploits very similar but not identical information as the estimator proposed by Kelejian & Prucha (1999). Substituting ε by $(I_n - \rho_0 M_n)u$, the moment condition can be written as a cubic polynomial in ρ :

$$m_{2,n}(\rho) = \mathbb{E} \left[\frac{1}{n} \left(u' M_n u + \rho_0 u' (\tilde{M}_n - M_n^2 - M_n' M_n) u + \rho_0^2 u' (M_n' M_n^2 - \tilde{M}_n M_n - M_n' \tilde{M}_n) u + \rho_0^3 u' M_n' \tilde{M}_n M_n u \right) \right] = 0. \quad (12)$$

Now let $\hat{m}_{2,n}(\rho)$ be the empirical counterpart of (12) with u replaced by \hat{u} . The MLAM2 estimator $\hat{\rho}_2$ is the real root of $\hat{m}_{2,n}(\rho)$ that satisfies assumption 3 (b). From this it is clear that we do not need an initial estimate of ρ_0 , as suggested by the moment condition in (11). Hence, the implementation of both the MLAM1 and the MLAM2 estimator is straight forward and computationally very simple.⁶ In a similar manner the m 'th order approximation to the scores can be obtained from replacing $B_n(\rho)$ by $I + \rho M_n + \dots + \rho^{m-1} M_n^{m-1}$. For $m \rightarrow \infty$ the m 'th order MLAM converges to the ML estimator. Since

⁵ In section 4 we will split up that linear combination into two moment conditions, yielding an over-identified GMM version of the MLAM2 estimator.

⁶ It should be noted that in principle all real roots of (12) may be out of the domain $(-1, 1)$. While this occurs very rarely and only for extreme values of ρ_0 a simple and tractable solution of this issue is to minimize $|\hat{m}_{2,n}(\rho)|$ under the constraint $|\rho| \leq 1$ whenever all real roots lie outside the domain $[-1, 1]$. In the even more exceptional case that the solution is not unique, because more than one real roots lie in the domain $(-1, 1)$, it is less straight forward to determine $\hat{\rho}_2$. However, unreported simulations suggest that this is not a relevant issue in practice.

the computational advantage over the ML estimator gets lost for MLAM estimators with $m > 2$ we focus on the first and second order approximation.

Next, we derive the limiting distribution of the MLAM estimators. The moment conditions of the estimators can be represented as a quadratic form given by

$$m_{k,n}(\rho_0) = \mathbb{E} \left[\frac{1}{n} \varepsilon' A_{k,n}(\rho_0) \varepsilon \right] = 0, \quad (13)$$

where for the MLAM1 estimator

$$A_{1,n}(\rho) = M_n \quad (14)$$

whereas for the MLAM2 estimator we have

$$A_{2,n}(\rho) = M_n + \rho \tilde{M}_n. \quad (15)$$

Let ε_i denote the i 'th element of the vector ε and $a_{k,ij,n}$ the (i, j) -element of $A_{k,n}(\rho_0)$, $k \in \{1, 2\}$. Using similar representations as Born & Breitung (2011) we can rewrite $m_{k,n}(\rho_0)$ as follows:

$$\begin{aligned} \frac{1}{n} \varepsilon' A_{k,n}(\rho_0) \varepsilon &= \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n a_{k,ij,n} \varepsilon_i \varepsilon_j \\ &= \frac{1}{n} \sum_{i=2}^n \varepsilon_i \tilde{\zeta}_{k,i,n} + \frac{1}{n} \sum_{i=1}^n a_{k,ii,n} z_i, \end{aligned} \quad (16)$$

where $z_i = \varepsilon_i^2 - \sigma_0^2$, $\sum_{i=1}^n a_{k,ii,n} = 0$, and

$$\tilde{\zeta}_{k,i,n} = \sum_{j=1}^{i-1} (a_{k,ij,n} + a_{k,ji,n}) \varepsilon_j \quad \text{for } i \geq 2. \quad (17)$$

Under assumption 4 (a) the variance of $\tilde{\zeta}_{k,i,n}$ is finite for all i and n .

Under assumption 2 it holds that $\mathbb{E}(\varepsilon_i|\xi_{k,i,n}) = 0$, $\mathbb{E}(z_i) = \mathbb{E}(z_i|\varepsilon_j) = 0$ for $j \neq i$ and that $\sum_{i=2}^n \varepsilon_i \xi_{k,i,n}$ and $\sum_{i=1}^n a_{k,ii,n} z_i$ are uncorrelated. Note also that the latter sum is equal to zero for $k = 1$ (the MLAM1 estimator) since all diagonal elements of the matrix $A_{1,n}(\rho_0)$ equal zero. Furthermore, $\xi_{k,i,n}$ is a martingale difference sequence with respect to the increasing sigma-algebra generated by $\{\varepsilon_1, \dots, \varepsilon_{i-1}\}$. The central limit theorem for martingale difference sequences yields

$$\sqrt{n} \varepsilon' A_{k,n}(\rho_0) \varepsilon \xrightarrow{d} \mathcal{N}(0, V_k),$$

where

$$V_k = \sigma_0^4 \lim_{n \rightarrow \infty} \left[\left(\frac{1}{n} \sum_{i=2}^n \sum_{j=1}^{i-1} (a_{k,ij,n} + a_{k,ji,n})^2 \right) + \kappa_4 \left(\frac{1}{n} \sum_{i=1}^n a_{k,ii,n}^2 \right) \right], \quad (18)$$

where $\kappa_4 = \mathbb{E}(\varepsilon_i^4)/\sigma_0^4 - 1$ (with $\kappa_4 = 2$ for normally distributed errors). With these results the limiting distribution of the MLAM estimator can be derived.

Theorem 1. *Under assumptions 1 to 5 and $n \rightarrow \infty$ the k 'th order MLAM estimators are asymptotically distributed as*

$$\sqrt{n}(\hat{\rho}_k - \rho_0) \xrightarrow{d} \mathcal{N}(0, V_k/\psi_k^2), \quad \text{for } k = 1, 2$$

where V_k is defined in (18) and

$$\begin{aligned} \psi_1 &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} u' (2\rho M_n' M_n^2 - M_n' M_n - M_n^2) u \right] \\ \psi_2 &= \lim_{n \rightarrow \infty} \mathbb{E} \left[\frac{1}{n} u' (\tilde{M}_n - M_n^2 - M_n' M_n) u + 2\rho u' (M_n' M_n^2 - \tilde{M}_n M_n - M_n' \tilde{M}_n) u \right. \\ &\quad \left. + 3\rho^2 u' M_n' \tilde{M}_n M_n u \right]. \end{aligned}$$

The proof of Theorem 1 is provided in the appendix.

In practice, the asymptotic variances can be consistently estimated by the respective sample moments based on \hat{u} and by inserting the estimator

$$\hat{\kappa}_4 = \left(\frac{1}{n\hat{\sigma}^4} \sum_{i=1}^n \hat{\varepsilon}_i^4 \right) - 1, \quad (19)$$

where $\hat{\varepsilon}_i$ is the i 'th element of the vector $\hat{\varepsilon} = (I - \hat{\rho}M_n)\hat{u}$. Note also that the limiting distribution is invariant to the error variance σ_0^2 since the factor σ_0^4 drops from the asymptotic variance V_k/ψ_k^2 .

4 Efficient GMM estimators

As outlined in Section 2 the original GMM estimator suggested by Kelejian & Prucha (1999) based on the moment conditions (6) and (7) is not efficient. Kelejian & Prucha (2010b) and Drukker et al. (2013) propose GMM estimators with optimal weighting matrices Y_n as defined in (8). In this section we employ a simpler approach to obtain an asymptotically efficient GMM estimator based on the empirical counterpart of the moment vector $m_n(\rho) = [m_{1,n}(\rho), \bar{m}_{2,n}(\rho)]'$, proposed by Kelejian & Prucha (1999), or the MLAM2 version $m_n^*(\rho) = [m_{1,n}(\rho), \tilde{m}_{2,n}(\rho)]'$ with $\bar{m}_{2,n}(\rho)$ defined in (7) and $\tilde{m}_{2,n}(\rho) = \mathbb{E}(n^{-1}\varepsilon' \tilde{M}_n \varepsilon)$. Since $m_{2,n}(\rho) = m_{1,n}(\rho) + \rho \tilde{m}_{2,n}(\rho)$, which is easily seen in (11), the single moment condition of the simple MLAM2 estimator proposed in section 3 is just a linear combination of the moments in $m_n^*(\rho)$.

As in Section 3 we represent the first moment condition of the original KP estimator as

$$m_{1,n}(\rho_0) = \mathbb{E} \left(\frac{1}{n} \varepsilon' M_n \varepsilon \right) = \mathbb{E} \left(\frac{1}{n} \sum_{i=2}^n \eta_{1,i,n} \right) = 0 \quad (20)$$

with $\eta_{1,i,n} = \varepsilon_i \tilde{\zeta}_{1,i,n}$, where $\tilde{\zeta}_{1,i,n}$ is defined in (17), whereas the second moment condi-

tion can be represented as

$$\bar{m}_{2,n}(\rho_0) = \mathbb{E} \left(\frac{1}{n} \varepsilon' \bar{M}_n \varepsilon \right) = \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \eta_{2,i,n} \right) = 0, \quad (21)$$

$\eta_{2,1,n} = \bar{m}_{11,n} z_1$, $\eta_{2,i,n} = \sum_{j=1}^{i-1} \varepsilon_i (\bar{m}_{ij,n} + \bar{m}_{ji,n}) \varepsilon_j + \bar{m}_{ii,n} z_i$ for $i \geq 2$, where $\bar{m}_{ij,n}$ is the (i, j) -element of \bar{M}_n and $z_i = \varepsilon_i^2 - \sigma_0^2$.

It should be noted that $\mathbb{E} \left(\sum_{i=1}^n \bar{m}_{ii,n} z_i \right) = 0$ so that this term can be neglected when minimizing the criterion function. It is required only for computing the weighting matrix, which is typically held fix during the iterative minimization.

Using the results of Section 3 it is not difficult to show that $\eta_{i,n} = [\eta_{1,i,n}, \eta_{2,i,n}]'$ is a martingale difference sequence with respect to the increasing sigma-algebra generated by $\{\eta_{1,n}, \dots, \eta_{i-1,n}\}$. Invoking the central limit theorem for martingale difference sequences yields

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \eta_{i,n} \xrightarrow{d} \mathcal{N}(0, \bar{\mathcal{V}}_\eta), \quad (22)$$

where $\bar{\mathcal{V}}_\eta = \lim_{n \rightarrow \infty} \mathcal{V}_{\eta,n}$ and

$$\mathcal{V}_{\eta,n} = \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \eta_{i,n} \eta_{i,n}' \right). \quad (23)$$

Let $\hat{\eta}_{i,n}(\rho)$ be constructed as $\eta_{i,n}$, where ε_i is replaced by the i 'th element of $\hat{\varepsilon}_n(\rho) = (I - \rho M_n) \hat{u}$. An asymptotically efficient GMM estimator based on the KP moment conditions results from minimizing the objective function

$$\hat{Q}_n(\rho) = \left(\frac{1}{n} \sum_{i=1}^n \hat{\eta}_{i,n}(\rho) \right)' \hat{Y}_n(\hat{\rho}) \left(\frac{1}{n} \sum_{i=1}^n \hat{\eta}_{i,n}(\rho) \right), \quad (24)$$

with

$$\hat{Y}_n(\hat{\rho}) = \left(\hat{V}_{\eta,n} \right)^{-1} = \left[\frac{1}{n} \sum_{i=1}^n \hat{\eta}_{i,n}(\hat{\rho}) \hat{\eta}_{i,n}(\hat{\rho})' \right]^{-1}$$

From the law of large numbers it follows that under assumptions 1 to 5 the estimator $\hat{Y}_n(\hat{\rho})$ is consistent for the optimal weighting matrix $Y_n(\rho_0)$ whenever the estimate $\hat{\rho}$ is consistent for ρ_0 . For example, a consistent initial estimate may be obtained by letting $Y_n = W_{KP,n}$ as in the original approach by Kelejian & Prucha (1999).

The asymptotic distribution of the efficient GMM estimator is presented in the following theorem.

Theorem 2. *Under assumptions 1 to 5 and $n \rightarrow \infty$ the efficient GMM estimator $\hat{\rho}_{opt} = \operatorname{argmin}\{\hat{Q}_n(\rho)\}$ with $\hat{Q}_n(\rho)$ based on the moment vector $\hat{m}_n(\rho) = [\hat{m}_{1,n}(\rho), \hat{m}_{2,n}(\rho)]'$ and defined in (24) possesses the limiting distribution*

$$\sqrt{n}(\hat{\rho}_{opt} - \rho_0) \xrightarrow{d} \mathcal{N}(0, [\bar{\delta}(\rho_0)' \bar{Y}(\rho_0) \bar{\delta}(\rho_0)]^{-1}),$$

where $\bar{Y}(\rho_0) = \lim_{n \rightarrow \infty} Y_n(\rho_0)$ and $\bar{\delta}(\rho_0) = \lim_{n \rightarrow \infty} \delta_n(\rho_0)$ with

$$\delta_n(\rho_0) = \frac{1}{n} \begin{bmatrix} u'(2\rho_0 M_n' M_n^2 - M_n' M_n - M_n^2) u \\ u'(2\rho_0 M_n' \bar{M}_n M_n - M_n' \bar{M}_n - \bar{M}_n M_n) u \end{bmatrix}.$$

The proof of theorem 2 is provided in the appendix.

In practice the variance of $\hat{\rho}_{opt}$ can be consistently estimated by

$$V_{\hat{\rho}_{opt}} = \frac{1}{n} \left[\hat{\delta}_n(\hat{\rho}_{opt})' \hat{Y}_n(\hat{\rho}_{opt}) \hat{\delta}_n(\hat{\rho}_{opt}) \right]^{-1} \quad (25)$$

$$\text{where } \hat{\delta}_n(\hat{\rho}_{opt}) = \frac{1}{n} \begin{bmatrix} \hat{u}'(2\hat{\rho}_{opt}M'_nM_n^2 - M'_nM_n - M_n^2)\hat{u} \\ \hat{u}'(2\hat{\rho}_{opt}M'_n\bar{M}_nM_n - M'_n\bar{M}_n - \bar{M}_nM_n)\hat{u} \end{bmatrix}.$$

The asymptotic distribution of the GMM version of the MLAM estimator based on the moment vector $m_n^*(\rho) = [m_{1,n}(\rho), \tilde{m}_{2,n}(\rho)]'$ results easily from replacing the elements of \bar{M}_n by the elements of \tilde{M}_n when constructing $\eta_{2,i,n}$ and $\delta_n(\rho)$. Asymptotically this estimator may yield a smaller variance than the simple MLAM2 estimator if there exists some superior linear combination of $m_{1,n}(\rho)$ and $\tilde{m}_{2,n}(\rho)$ than $m_{1,n}(\rho) + \rho\tilde{m}_{2,n}(\rho)$.

5 Heteroskedastic and non-Gaussian errors

The MLAM1 estimator is based on the empirical counterpart of the moment $m_{1,n}(\rho_0) = \mathbb{E}(\varepsilon' M_n \varepsilon)$ which remains valid under heteroskedastic and non-Gaussian errors. In contrast, the original KP estimator and the MLAM2 estimator, including their efficient GMM variants presented in section 4, are inconsistent if the errors are heteroskedastic. This is due to the fact that $\mathbb{E}(z_i) = \mathbb{E}(\varepsilon_i^2) - \sigma_0^2$ may be different from zero (see eqs. (16) and (21)). Furthermore, the asymptotic variances of the estimators presented in Theorems 1 and 2 require homoskedastic errors (see assumption 2).

In order to cope with this shortcoming, Kelejian & Prucha (2010b) propose the alternative moment condition $\bar{m}_{2,n}^h(\rho) = \mathbb{E}[\varepsilon'(M'_nM_n - \text{diag}\{M'_nM_n\})\varepsilon] = \mathbb{E}(\varepsilon'\bar{M}_{0,n}\varepsilon)$, where $\text{diag}\{M'_nM_n\}$ represents a diagonal matrix constructed by the diagonal elements of the matrix M'_nM_n . Accordingly, the matrix $\bar{M}_{0,n}$ is obtained by setting the diagonal elements of M'_nM_n equal to zero. Similarly, a heteroskedasticity robust modification for the MLAM2 moment condition $m_{2,n}(\rho)$ is given by $m_{2,n}^h(\rho) = \mathbb{E}[\varepsilon'(M_n + \rho\tilde{M}_{0,n})\varepsilon]$ with $\tilde{M}_{0,n} = M_n^2 - \text{diag}\{M_n^2\}$.

These heteroskedasticity robust moments can easily be constructed by dropping the terms depending on $z_i = \varepsilon_i^2 - \sigma_0^2$ in (16) and (21). For the simple (i.e. just identified)

MLAM2 estimator the robust moment condition results as

$$m_{2,n}^h(\rho_0) = \mathbb{E} \left(\frac{1}{n} \sum_{i=2}^n \tilde{\eta}_{2,i,n}^h \right) = 0 \quad (26)$$

where $\tilde{\eta}_{2,i,n}^h = \varepsilon_i \tilde{\zeta}_{2,i,n}$ and $\tilde{\zeta}_{2,i,n}$ is defined in (17). The heteroscedasticity robust version of the moment used by the original KP estimator is given by

$$\bar{m}_{2,n}^h(\rho_0) = \mathbb{E} \left(\frac{1}{n} \sum_{i=2}^n \eta_{2,i,n}^h \right) = 0, \quad (27)$$

where $\eta_{2,i,n}^h = \sum_{j=1}^{i-1} (\bar{m}_{ij,n} + \bar{m}_{ji,n}) \varepsilon_i \varepsilon_j$. For the GMM version of the MLAM2 estimator the corresponding moment $\tilde{m}_{2,n}^h(\rho_0)$ is constructed equivalently, replacing the elements of \bar{M}_n by those of \tilde{M}_n .

The asymptotic distributions of the resulting estimators are easily derived by setting the diagonal elements of $A_{2,n}(\rho_0)$ (for the simple MLAM2 estimator), respectively \bar{M}_n and \tilde{M}_n (for the GMM estimators), equal to zero. For example, the asymptotic variance V_2 in (18) is replaced by

$$V_2 = \sigma_0^4 \lim_{n \rightarrow \infty} \left(\frac{1}{n} \sum_{i=2}^n \sum_{j=1}^{i-1} (a_{2,ij,n} + a_{2,ji,n})^2 \right). \quad (28)$$

The respective modifications of the limiting distributions presented in Theorems 1 and 2 are obvious and we therefore do not derive the modified limiting distributions.

6 GMM estimators for the spatial autoregressive model (SAR) and the Durbin model (SDM)

In empirical practice, the spatial autoregressive model is more relevant as it offers the opportunity to estimate spatial spillover effects (see e.g. Elhorst 2010a). The basic SAR

model is given by

$$y = \gamma W_n y + X\beta + \varepsilon, \quad (29)$$

where γ and W_n correspond to ρ and M_n in the spatial error model. Furthermore, let $C_n(\gamma) \equiv (I_n - \gamma W_n)$ in what follows. Accordingly, we now apply assumptions 3 and 4 to γ , W_n , and $C_n(\gamma)$ instead of ρ , M_n , and $B_n(\rho)$. Assuming further that $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_n)$ yields the log-likelihood function

$$\ell(\beta, \gamma, \sigma^2) = \text{const} - \frac{n}{2} \ln(\sigma^2) + \ln|C_n(\gamma)| - \frac{1}{2\sigma^2} (y - \gamma W_n y - X\beta)' (y - \gamma W_n y - X\beta)$$

with the gradients⁷

$$\begin{aligned} \frac{\partial \ell(\cdot)}{\partial \beta} &= \frac{1}{\sigma^2} X' (y - \gamma W_n y - X\beta) \\ \frac{\partial \ell(\cdot)}{\partial \gamma} &= -\text{tr}\{W_n C_n(\gamma)^{-1}\} + \frac{1}{\sigma^2} (y - \gamma W_n y - X\beta)' W_n y. \end{aligned}$$

The reduced form representation of the model

$$y = C_n(\gamma)^{-1} X\beta + u$$

with $u = C_n(\gamma)^{-1} \varepsilon$ yields

$$\begin{aligned} \frac{\partial \ell(\cdot)}{\partial \gamma} &= -\text{tr}\{W_n C_n(\gamma)^{-1}\} + \frac{1}{\sigma^2} (y - \gamma W_n y - X\beta)' W_n C_n(\gamma)^{-1} (X\beta + \varepsilon) \\ &= -\text{tr}\{W_n C_n(\gamma)^{-1}\} + \left[\frac{1}{\sigma^2} (y - \gamma W_n y - X\beta)' W_n C_n(\gamma)^{-1} X\beta \right] \\ &\quad + \left[\frac{1}{\sigma^2} (y - \gamma W_n y - X\beta)' W_n C_n(\gamma)^{-1} \varepsilon \right]. \end{aligned}$$

⁷For convenience we assume σ^2 to be known. As usual a consistent estimator can be obtained from the residuals of the model.

Concentrating out σ^2 yields the moment conditions $g_1(\beta, \gamma) = \varepsilon' W_n C_n(\gamma)^{-1} X \beta$ and $g_2(\beta, \gamma) = \varepsilon' [W_n C_n(\gamma)^{-1} - \frac{1}{n} \text{tr}\{W_n C_n(\gamma)^{-1}\} I_n] \varepsilon$ along with the moment condition obtained from the gradient with respect to β , $g_0(\beta, \gamma) = X' \varepsilon$. These are the same moment conditions as the ones proposed by Lee (2007a) for the best GMM estimator (BGMME).

Using the first order approximation $C_n(\gamma)^{-1} \approx I_n$ gives rise to the following set of moments for estimating β and γ

$$\tilde{g}_0(\beta, \gamma) = X' \varepsilon \quad (30)$$

$$\tilde{g}_1(\beta, \gamma) = \beta' X' W_n' \varepsilon \quad (31)$$

$$\tilde{g}_2(\beta, \gamma) = \varepsilon' W_n \varepsilon. \quad (32)$$

The first and third moment conditions are also employed for the spatial error model and only the second moment is added for the spatial autoregressive model. Accordingly, the treatment of the spatial autoregressive model is straightforward.

Note that our approach yields not only computationally very simple but also heteroskedasticity robust moment conditions. In particular, $\tilde{g}_2(\beta, \gamma)$ in (32) is heteroskedasticity robust, while the original moment condition $g_2(\beta, \gamma)$ is not. An alternative heteroskedasticity robust, yet computationally more demanding version of $g_2(\beta, \gamma)$ is $\varepsilon' [W_n C_n(\gamma)^{-1} - \text{diag}\{W_n C_n(\gamma)^{-1}\}] \varepsilon$, as proposed by Lin & Lee (2010).

The moment condition $\tilde{g}_1(\beta, \gamma)$ can be interpreted as a linear combination of the k vector of moments

$$\tilde{g}_1^+(\beta, \gamma) = X' W_n' \varepsilon. \quad (33)$$

While the vector of moments $\tilde{g}(\beta, \gamma) = (\tilde{g}_0(\beta, \gamma)', \tilde{g}_1^+(\beta, \gamma)', \tilde{g}_2(\beta, \gamma)')$ leads to an over-identified GMM estimator whenever the number of regressors is larger than one, the linear moment conditions in $\tilde{g}_1^+(\beta, \gamma)$ are computationally simpler than the nonlinear

moment condition $\tilde{g}_1(\beta, \gamma)$. The optimal weight matrix can be constructed as in section 4. Specifically we rewrite the moment condition (32) as

$$\tilde{g}_2(\beta, \gamma) = \sum_{i=2}^N \zeta_{i,n} \varepsilon_i \quad (34)$$

where $\zeta_{i,n} = \sum_{j=1}^{i-1} (w_{ij,n} + w_{ji,n}) \varepsilon_j$ for $i \geq 2$, and $w_{ij,n}$ denotes the (i, j) element of W_n . Accordingly the set of moment conditions can be written as

$$\tilde{g}(\beta, \gamma) = \sum_{i=1}^N z_{i,n} \varepsilon_i$$

where

$$z_{i,n} = \begin{pmatrix} x_i \\ X' w'_{i,n} \\ \zeta_{i,n} \end{pmatrix}$$

and $w_{i,n}$ denotes the i 'th row of W_n . Let $\hat{\varepsilon}_i$ denote the i 'th element of the vector $\hat{\varepsilon} = (I_n - \hat{\gamma}W_n)\hat{u}$, where \hat{u} denotes the residuals from a consistent initial estimation of the model (e.g. by using the identity matrix as the weight matrix). The optimal weight matrix of the GMM estimator can be easily estimated as

$$\hat{Y}_n = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 \hat{z}_{i,n} \hat{z}'_{i,n}$$

where $\hat{z}_{i,n}$ is constructed as $z_{i,n}$ but using $\hat{\varepsilon}_i$ instead of ε_i when computing $\zeta_{i,n}$. We refer to the resulting estimator as the MLAM1 estimator for the spatial autoregressive model.

Note that the Generalized Spatial Two-Stage Least Squares (GS2SLS) estimator of Kelejian & Prucha (1998) is based on $g_0(\beta, \gamma)$ and a p 'th order approximation of $g_1(\beta, \gamma)$, resulting in the instrumental variable matrix that is given by the linearly independent

columns of $(X, W_n X, W_n^2 X, \dots, W_n^p X)$ whereas the term $g_2(\beta, \gamma)$ is ignored. As pointed out by Lee (2007a) this may result in a dramatic loss of efficiency if β is close to zero.

Let us finally consider the SDM given by

$$y = \gamma W_n y + X\beta + W_n X\alpha + \varepsilon. \quad (35)$$

In this case the MLAM1 estimator based on the vector of moments $\tilde{g}(\beta, \gamma)$ and treating $W_n X$ as additional exogenous regressors is no longer identified because $W_n X$ is already employed as instrument for the lagged dependent variable in (33). A potential remedy is the use of higher order approximations of $C_n(\gamma)^{-1}$. For example, a second order approximation of the moment $g_2(\beta, \gamma)$ yields

$$\begin{aligned} g_1^*(\beta, \gamma) &= \varepsilon' W_n (I_n + \gamma W_n) X \beta \\ &= \tilde{g}_1(\beta, \gamma) + \gamma \varepsilon' W_n^2 X \beta. \end{aligned}$$

This suggest to employ

$$\begin{aligned} \tilde{g}_{1a}^+(\alpha, \beta, \gamma) &= X' W_n' \varepsilon \\ \tilde{g}_{1b}^+(\alpha, \beta, \gamma) &= X' W_n^{2'} \varepsilon \end{aligned}$$

for identifying the parameters α and γ , where $\varepsilon = y - \gamma W_n y - X\beta - W_n X\alpha$. As pointed out by Gibbons & Overman (2012) the identification by powers of W_n typically results in weak identification. Accordingly, the loss in efficiency may become large relative to the ML estimator when estimating the SDM.⁸

⁸We are grateful to an anonymous referee for pointing out this potential problem with GMM estimates in the SDM.

7 Monte Carlo comparison

We now compare the properties of alternative MM and GMM estimators for the SEM and SAR in a Monte Carlo (MC) simulation distinguishing the case of small and large samples as well as homo- and heteroskedastic disturbances. To this end we employ two variants of a row standardized “ahead-behind” spatial weight matrix similar to the one used by Kelejian & Prucha (2007) in which observation i has $d_i = 2r_i$ neighbors, r_i “ahead” ($i - r_i, \dots, i - 1$) and r_i “behind” ($i + 1, \dots, i + r_i$).⁹ The corresponding elements of the matrix are initially set to one and then the matrix is row-normalized. In the first variant, $M_{1,n}$, the first and third quarter of observations have $d_i = 8$ neighbors each whereas the second and last quarter have $d_i = 2$ neighbors each. Hence, all nonzero elements of the weight matrix equal $1/8$ or $1/2$. In the second variant, $M_{2,n}$, the first and third quarter of observations have $d_i = 6$ neighbors while the second and last quarter of observations have $d_i = 4$ neighbors, with corresponding nonzero elements equal to $1/6$ or $1/4$. Note that the resulting spatial weight matrices are not symmetric. The characteristics of both matrices for sample sizes $n = 100$ and $n = 1000$ are summarized in Table 1:

Table 1: Characteristics of spatial weight matrices.

	No. of neighbors			percent nonzero	
	avg.	max	min	$n = 100$	1000
$M_{1,n}$	5.0	8	2	5.0	0.5
$M_{2,n}$	5.0	6	4	5.0	0.5

For the spatial autoregressive parameter ρ_0 we consider the values $-0.8, -0.4, 0, 0.4$ and 0.8 . In our baseline specification the error vector is generated as $\varepsilon \sim \mathcal{N}(0, \sigma_0^2 I_n)$

⁹ Since we consider a “circular world” observation 1 and observation n are direct neighbors. Hence, for observation $i = n$, the j th “behind”-neighbor $i + j$ is observation j .

with $\sigma_0^2 = 1$. As noted by Kelejian & Prucha (1999) the choice of the error variance σ_0^2 does not affect the performance of the estimators. Results are presented for small samples with $n = 100$ and large samples with $n = 1000$.

In addition to the baseline specification we also consider a heteroskedastic setup similar to Kelejian & Prucha (2010*b*). Specifically $\varepsilon_i = \sigma_{0,i}\varepsilon_i$, where ε_i is generated by independent draws from a standard normal distribution and $\sigma_{0,i}^2 = d_i/5$, where, as defined above, the factor d_i denotes the number of neighbors of individual i . The scaling is such that the average error variance is approximately 1.

The following estimators are included in our MC study: KP-NLS refers to the inefficient and not heteroskedasticity robust original KP estimator computed by running a nonlinear least-squares estimator based on the three moment conditions in (5). In the heteroskedastic-error setup we compare this estimator to the heteroskedasticity robust but inefficient GMM version of the original KP estimator using the weighting matrix $W_{KP,n}$ (see section 2), labeled KP-GMM. MLAM1 and MLAM2 indicate the simple moment estimators based on a first and second order approximation of the likelihood function (see Section 3). Recall that these estimators involve no weighting matrix as they are based on a single moment condition. For both estimators we use the heteroskedasticity robust variants (described in Section 5) in the setup with heteroskedastic errors. KP-eff and MLAM-eff refer to the efficient GMM estimators based on the two-dimensional KP and MLAM2 moment vector, respectively, and the corresponding optimal weighting matrix proposed in section 4.

Besides comparing the estimation performance in terms of the bias and root mean squared error (RMSE) of the estimators, actual sizes of the t -statistics are reported in order to assess whether the asymptotic properties regarding the variance-covariance structure of the estimators remain valid in small samples.¹⁰ The MC results allow us

¹⁰ This is done by running a t-test for each estimate $\hat{\rho}$ of whether it equals the true parameter ρ_0 . In doing so we use the corresponding standard error estimate of ρ which is based on its asymptotic distribution. We indicate a rejection if the (true) null hypothesis of equality is rejected at the 5% level.

to assess how our simple MM estimators perform compared to the original estimator by Kelejian & Prucha (1999) and its efficient and heteroskedasticity robust counterpart. For our analysis we focus on the disturbance vector of the SEM model given in equation (2). All MC results are based on 1000 replications as in comparable simulation studies (e.g. Lin & Lee 2010 and Liu et al. 2010).

Table 2 summarizes the MC results for both spatial weight matrices and sample size $n = 100$ under homoskedasticity. There is no clear tendency for any estimator to yield the smallest bias if the sample size is that small. This holds true for both types of the spatial weight matrix, $M_{1,n}$ and $M_{2,n}$. However, the MLAM2 estimator tends to yield the largest bias in cases where ρ_0 is large in absolute value. The efficient GMM estimators (KP-eff and MLAM-eff) do not yield smaller RMSE than the other estimators if the variation in the number of neighbors per observation is large ($M_{1,n}$). In the case of a more moderate variation in the number of neighbors per observation ($M_{2,n}$) the efficient GMM estimator with KP moments improves slightly in terms of RMSE relative to the original KP estimator (KP-NLS). Overall, the MLAM2 estimator yields the smallest RMSE for samples of size 100. Furthermore, the results in Table 2 show that the KP-NLS estimator fits the desired rejection rate of 5% best in the case of small samples with large variation in the number of neighbors ($M_{1,n}$) whereas the efficient GMM estimators tend to reject too often. If the number of neighbors varies less ($M_{2,n}$) the actual size of the MLAM1 approach is close to the nominal size of 5%.

[TABLE 2 ABOUT HERE]

Increasing the sample size to 1000 but maintaining homoskedastic errors changes the relative performance of the estimators only little, see Table 3. Still there is no clear tendency which estimator yields the smallest bias if the number of neighbors per observation varies a lot ($M_{1,n}$). For the spatial weight matrix with moderate variation

Hence, for each estimator we expect a (wrong) rejection to occur in 5% of the MC replications.

($M_{2,n}$) the efficient GMM estimators yield the smallest bias whereas the KP-NLS estimator possesses the largest bias. In specification $M_{1,n}$ the efficient GMM estimator with KP moments yields a smaller RMSE than the KP-NLS estimator while the efficient GMM variant of the MLAM2 estimator does not improve relative to the simpler variants. This changes in specification M2 where both variants of the efficient GMM yield a smaller or equal RMSE than the other estimators. Generally the differences are not very large, however. Overall the estimators fit the desired rejection rate of 5% adequately, an exception being the efficient GMM estimators in the case of specification $M_{1,n}$ with $\rho_0 = -0.8$.

[TABLE 3 ABOUT HERE]

Incorporating heteroskedasticity that depends on the spatial dependence structure, which is arguably more realistic than the case of homoskedastic errors, leads to more pronounced results. Table 4 contains the results for small samples ($n = 100$). Not surprisingly the original KP estimator, which relies on homoskedasticity, yields much larger biases than the heteroskedasticity robust estimators for almost all values of ρ . Among the latter there is, however, again no clear ranking in terms of size distortions. The efficient GMM estimators with KP moments yield smaller RMSE than the inefficient counterpart (KP-GMM) in more cases than under homoskedasticity and for both spatial weight matrices. Yet the MLAM2 and even the MLAM1 estimator still yield the smallest RMSE in most cases. The actual sizes are least reliable for the efficient GMM estimators and the KP-NLS estimator.

[TABLE 4 ABOUT HERE]

The MC results presented above do not reveal a notable advantage of the efficient GMM estimators. This picture changes in the case of heteroskedastic errors and a sample size of 1000. These results are summarized in Table 5. The efficient GMM estimators

yield the smallest bias across all specifications. As expected, the bias is almost indistinguishable between these two estimators. Not surprisingly the non-robust KP-NLS estimator yields the largest bias. Also in terms of the RMSE the efficient GMM estimators perform best. However, for some values of ρ the RMSE of the simple MLAM estimators is only marginally larger. While again the KP-NLS estimator generally yields the largest RMSE, the inefficient GMM variant yields similar results as the MLAM estimators for positive values of ρ . The fit of the rejection rate is best for the efficient GMM estimators in most specifications. In some cases the MLAM1 estimator performs even better. Generally the fit is adequate for all estimators but the KP-NLS estimator.

[TABLE 5 ABOUT HERE]

The results of these MC simulations show that both, the efficient GMM variant with KP moments and the one with MLAM2 moments tend to outperform the other estimators in terms of bias and RMSE if the sample size is sufficiently large. The advantage is more pronounced in the case of heteroskedastic errors. Comparing both efficient GMM estimators with each other the performance is indistinguishable, as expected.¹¹

The simple MLAM estimators, based on a single moment condition, perform very well in general. For smaller samples they even outperform the (overidentified) efficient GMM estimators. This observation is remarkable and makes these simple moment estimators, not involving the estimation of a weighting matrix, a reasonable alternative to conventional GMM estimators. This holds in particular for the MLAM1 estimator, which has an analytical solution. Overall, the results also reassure the validity of the applied approximations empirically.

We close our MC simulation study by comparing the MLAM1 estimator for the spatial autoregressive model to the ML estimator and the GS2SLS estimator of Kelejian & Prucha (1998) which is based on the instrumental variable matrix $[X, WX, W^2X]$. The

¹¹ Recall that these estimators differ only in the second element of the moment vector given in (27). In the case of a symmetric spatial weight matrix they are equal, since then $\bar{M}_n \equiv M_n' M_n = M_n M_n \equiv \tilde{M}_n$.

regressor is generated as $x_i \stackrel{i.i.d.}{\sim} \mathcal{N}(0, 1)$ and the spatial dependence is generated by the weight matrix $W_n = M_{2,n}$. The sample size is $n = 100$ and various combinations of the parameters β and γ are considered. From the results presented in Table 6 it turns out that the MLAM1 estimator performs much better than the GS2SLS estimator, in particular if β is small. The reason is that for small β the instruments $W^k X$ are weak leading to a severe bias and large standard errors. Unfortunately, in empirical applications the R^2 is typically small, which is the scenario where the GS2SLS estimator performs poorly. As expected, the ML estimator performs best but the MLAM1 estimator is also unbiased and nearly as efficient as the ML estimator. Note that the MLAM1 estimator is robust against heteroskedastic and non-Gaussian errors which is not the case for the ML estimator.

[TABLE 6 ABOUT HERE]

8 Conclusion

In this paper we reconsider (approximately) optimal moment conditions for the estimation of the spatial parameter in regression models with spatial error correlation (SEM). These are directly derived from the first order condition for the maximization of the log-likelihood function. The resulting moment conditions yield computationally simple and robust estimators. Illustrating the similarity of our moment conditions to those used by Kelejian & Prucha (1999) we provide an intuitive interpretation for their popular method of moments estimator. In addition we derive simplified efficient GMM estimators based on a modification of the moment conditions. Following Kelejian & Prucha (2010b) and Lin & Lee (2010) we also propose heteroskedasticity robust versions of all our estimators. Finally, we extend the idea underlying our estimators to the GMM estimation of the spatial autoregressive (SAR) and the spatial Durbin model (SDM).

Our MC results suggest that the efficient GMM estimators are (slightly) more efficient than the original KP estimator if the errors are heteroskedastic and the sample is large. As expected, the KP moments and the MLAM2 moments perform equally well, confirming our interpretation of the estimator by Kelejian & Prucha (1999) as an approximation of the score of the log-likelihood function. Most importantly the simplest MLAM1 estimator performs similar to the more demanding GMM or ML variants suggesting that this estimator is particularly attractive in empirical practice. Our MLAM1 estimator for the spatial autoregressive model performs well in comparison to both the GS2SLS estimator proposed by Kelejian & Prucha (1998) and also the (efficient) ML estimator.

References

- Anselin, L. (1988), *Spatial Econometrics: Methods and Models*, Kluwer Academic Publishers, Dordrecht, NL.
- Anselin, L. (2010), 'Thirty years of spatial econometrics', *Papers in Regional Science* **89**, 3–25.
- Anselin, L. & Rey, S. J. (2014), *Modern Spatial Econometrics in Practice: A Guide to GeoDa, GeoDaSpace and PySAL*, GeoDa Press LLC, Chicago, IL.
- Arnold, M. & Wied, D. (2010), 'Improved GMM estimation of the spatial autoregressive error model', *Economics Letters* **108**, 65–68.
- Arraiz, I., Drukker, D. M., Kelejian, H. H. & Prucha, I. R. (2010), 'A Spatial Cliff-Ord-Type Model With Heteroskedastic Innovations: Small and Large Sample Results', *Journal of Regional Science* **50**, 592–614.
- Baltagi, B. H. & Liu, L. (2011), 'Instrumental variable estimation of a spatial autoregressive panel model with random effects', *Economics Letters* **111**, 135–137.
- Born, B. & Breitung, J. (2011), 'Simple regression-based tests for spatial dependence', *Econometrics Journal* **14**, 330–342.
- Brady, R. R. (2014), 'The spatial diffusion of regional housing prices across U.S. states', *Regional Science and Urban Economics* **46**, 150–166.
- Cliff, A. & Ord, J. (1973), *Spatial Autocorrelation*, Pion, London, UK.
- de Dominicis, L., Florax, R. J. G. M. & de Groot, H. (2013), 'Regional clusters of innovative activity in Europe: are social capital and geographical proximity key determinants?', *Applied Economics* **45**, 2325–2335.
- Drukker, D. M., Egger, P. & Prucha, I. R. (2013), 'On Two-Step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors', *Econometric Reviews* **32**, 686–733.
- Elhorst, J. P. (2003), 'Specification And Estimation of Spatial Panel Data Models', *International Regional Science Review* **26**, 244–268.
- Elhorst, J. P. (2010a), 'Applied Spatial Econometrics: Raising the Bar', *Spatial Economic Analysis* **5**, 9–28.

- Elhorst, J. P. (2010b), 'Dynamic panels with endogenous interaction effects when T is small', *Regional Science and Urban Economics* **40**, 272–282.
- Elhorst, J. P. (2012), 'Dynamic spatial panels: models, methods, and inferences', *Journal of Geographical Systems* **14**, 5–28.
- 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**, 35–57.
- 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.
- Fingleton, B. & Le Gallo, J. (2008a), 'Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: Finite sample properties', *Papers in Regional Science*, **87**, 319–339.
- Fingleton, B. & Le Gallo, J. (2008b), 'Finite Sample Properties of Estimators of Spatial Models with Autoregressive, or Moving Average, Disturbances and System Feedback', *Annales D'Économie et de Statistique* **87**.
- Gibbons, S. & Overman, H. G. (2012), 'Mostly Pointless Spatial Econometrics?', *Journal of Regional Science* **52**, 172–191.
- Kapoor, M., Kelejian, H. H. & Prucha, I. R. (2007), 'Panel data models with spatially correlated error components', *Journal of Econometrics* **140**, 97–130.
- Kelejian, H. H., Murrell, P. & Shepotylo, O. (2013), 'Spatial spillovers in the development of institutions', *Journal of Development Economics* **101**, 297–315.
- Kelejian, H. H. & Piras, G. (2014), 'Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes', *Regional Science and Urban Economics* **46**, 140–149.
- Kelejian, H. H. & Prucha, I. R. (1998), 'A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances', *Journal of Real Estate Finance and Economics* **17**, 99–121.
- Kelejian, H. H. & Prucha, I. R. (1999), 'A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model', *International Economic Review* **40**, 509–533.

- Kelejian, H. H. & Prucha, I. R. (2001), 'On the asymptotic distribution of the Moran I test statistic with applications', *Journal of Econometrics* **104**, 219–257.
- Kelejian, H. H. & Prucha, I. R. (2004), 'Estimation of simultaneous systems of spatially interrelated cross sectional equations', *Journal of Econometrics* **118**, 27–50.
- Kelejian, H. H. & Prucha, I. R. (2007), 'HAC estimation in a spatial framework', *Journal of Econometrics* **140**, 131–154.
- Kelejian, H. H. & Prucha, I. R. (2010a), 'Spatial models with spatially lagged dependent variables and incomplete data', *Journal of Geographic Systems* **12**, 241–257.
- Kelejian, H. H. & Prucha, I. R. (2010b), 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics* **157**, 53–67.
- Kelejian, H. H., Prucha, I. R. & Yuzefovich, Y. (2004), Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results, in 'Spatial and Spatiotemporal Econometrics, Advances in Econometrics, Volume 18', Elsevier, New York, NY, pp. 163–198.
- Lee, L.-f. (2002), 'Consistency and Efficiency of Least Squares Estimation for Mixed Regressive, Spatial Autoregressive Models', *Econometric Theory* **18**, 252–277.
- Lee, L.-f. (2003), 'Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances', *Econometric Reviews* **22**, 307–335.
- Lee, L.-f. (2004), 'Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models', *Econometrica* **72**, 1899–1925.
- Lee, L.-f. (2007a), 'GMM and 2SLS estimation of mixed regressive, spatial autoregressive models', *Journal of Econometrics* **137**, 489–514.
- Lee, L.-f. (2007b), 'The method of elimination and substitution in the GMM estimation of mixed regressive, spatial autoregressive models', *Journal of Econometrics* **140**, 155–189.
- Lee, L.-F. & Liu, X. (2010), 'Efficient GMM Estimation of High Order Spatial Autoregressive Models with Autoregressive Disturbances', *Econometric Theory* **26**, 187–230.
- Lee, L.-f. & Yu, J. (2014), 'Efficient GMM estimation of spatial dynamic panel data models with fixed effects', *Journal of Econometrics* **180**, 174–197.

- Lee, L.-f. & Yu, J. (2016), 'Identification of Spatial Durbin Panel Models', *Journal of Applied Econometrics* **31**, 133–162.
- LeSage, J. & Pace, R. K. (2009), *Introduction to Spatial Econometrics*, CRC Press Inc., Taylor and Francis Group, Broken Sound Parkway, NW.
- Lin, X. (2010), 'Identifying Peer Effects in Student Academic Achievement by Spatial Autoregressive Models with Group Unobservables', *Journal of Labor Economics* **28**, 825–860.
- Lin, X. & Lee, L.-F. (2010), 'GMM estimation of spatial autoregressive models with unknown heteroskedasticity', *Journal of Econometrics* **157**, 34–52.
- Liu, X. & Lee, L.-f. (2010), 'GMM estimation of social interaction models with centrality', *Journal of Econometrics* **159**, 99–115.
- Liu, X. & Lee, L.-F. (2013), 'Two-Stage Least Squares Estimation of Spatial Autoregressive Models with Endogenous Regressors and Many Instruments', *Econometric Reviews* **32**, 734–753.
- Liu, X., Lee, L.-f. & Bollinger, C. R. (2010), 'An efficient GMM estimator of spatial autoregressive models', *Journal of Econometrics* **159**, 303–319.
- Miguélez, E. & Moreno, R. (2013), 'Research Networks and Inventors' Mobility as Drivers of Innovation: Evidence from Europe', *Regional Studies* **47**, 1668–1685.
- Newey, W. & McFadden, D. (1994), Large Sample Estimation and Hypothesis Testing, in R. Engle & D. McFadden, eds, 'Handbook of Econometrics, Volume 4', North-Holland, Amsterdam, NL, pp. 2111–2245.
- Piras, G., Postiglione, P. & Aroca, P. (2012), 'Specialization, R&D and productivity growth: evidence from EU regions', *Annals of Regional Science* **49**, 35–51.
- Prucha, I. R. (2014), Generalized Method of Moments Estimation of Spatial Models, in M. M. Fischer & P. Nijkamp, eds, 'Handbook of Regional Science', Springer Verlag, Berlin, DE, pp. 1597–1618.
- Qu, X. & Lee, L.-f. (2015), 'Estimating a spatial autoregressive model with an endogenous spatial weight matrix', *Journal of Econometrics* **184**, 209–232.
- Qu, X., Wang, X. & Lee, L.-f. (2016), 'Instrumental variable estimation of a spatial dynamic panel model with endogenous spatial weights when T is small', *Econometrics Journal* **19**, 261–290.

Wang, W. & Lee, L.-f. (2013), 'Estimation of spatial panel data models with randomly missing data in the dependent variable', *Regional Science and Urban Economics* **43**, 521–538.

Tables

Table 2: Bias, (RMSE) and {sizes} for homoskedastic errors, $n = 100$

ρ	KP-NLS	MLAM1	MLAM2	KP-eff	MLAM-eff
<i>spatial weight matrix: $M_{1,n}$ (8 or 2 neighbors)</i>					
-0.80	0.0064 (0.0747) {0.0430}	0.0048 (0.0796) {0.0430}	0.0121 (0.0696) {0.0420}	0.0104 (0.0743) {0.0810}	0.0104 (0.0742) {0.0800}
-0.40	0.0022 (0.1214) {0.0500}	0.0038 (0.1229) {0.0500}	0.0110 (0.1163) {0.0510}	0.0100 (0.1221) {0.0740}	0.0099 (0.1219) {0.0730}
0.00	-0.0046 (0.1293) {0.0510}	-0.0005 (0.1298) {0.0510}	-0.0002 (0.1256) {0.0540}	0.0031 (0.1307) {0.0710}	0.0030 (0.1306) {0.0690}
0.40	-0.0086 (0.1081) {0.0500}	-0.0041 (0.1102) {0.0510}	-0.0087 (0.1066) {0.0530}	-0.0036 (0.1100) {0.0710}	-0.0037 (0.1100) {0.0720}
0.80	-0.0064 (0.0566) {0.0500}	-0.0039 (0.0608) {0.0430}	-0.0073 (0.0555) {0.0570}	-0.0054 (0.0580) {0.0760}	-0.0054 (0.0579) {0.0770}
<i>spatial weight matrix: $M_{2,n}$ (6 or 4 neighbors)</i>					
-0.80	0.0035 (0.1549) {0.0250}	0.0064 (0.1495) {0.0240}	0.0123 (0.1472) {0.0190}	0.0119 (0.1519) {0.0330}	0.0118 (0.1515) {0.0330}
-0.40	-0.0174 (0.1830) {0.0590}	-0.0130 (0.1770) {0.0580}	-0.0057 (0.1684) {0.0540}	-0.0045 (0.1746) {0.0810}	-0.0044 (0.1741) {0.0810}
0.00	-0.0194 (0.1641) {0.0550}	-0.0154 (0.1611) {0.0570}	-0.0145 (0.1580) {0.0600}	-0.0114 (0.1614) {0.0740}	-0.0113 (0.1612) {0.0720}
0.40	-0.0176 (0.1252) {0.0520}	-0.0147 (0.1256) {0.0550}	-0.0172 (0.1242) {0.0530}	-0.0146 (0.1255) {0.0690}	-0.0145 (0.1255) {0.0690}
0.80	-0.0100 (0.0621) {0.0490}	-0.0087 (0.0650) {0.0520}	-0.0105 (0.0617) {0.0490}	-0.0102 (0.0618) {0.0670}	-0.0101 (0.0617) {0.0650}

Notes: The number of MC replications is 1000. The errors are generated as $u = \rho Mu + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, I)$. Entries report bias (without brackets), RMSE in round brackets, empirical sizes of t -tests in curly brackets (nominal size = 0.050). For details on the spatial weight matrices see Table 1.

Table 3: Bias, (RMSE) and {sizes} for homoskedastic errors, $n = 1000$

ρ	KP-NLS	MLAM1	MLAM2	KP-eff	MLAM-eff
<i>spatial weight matrix: $M_{1,n}$ (8 or 2 neighbors)</i>					
-0.8	0.0003 (0.0197) {0.0480}	0.0000 (0.0217) {0.0440}	0.0008 (0.0182) {0.0510}	0.0006 (0.0193) {0.0690}	0.0006 (0.0193) {0.0670}
-0.4	-0.0004 (0.0359) {0.0500}	-0.0003 (0.0362) {0.0450}	0.0005 (0.0350) {0.0480}	0.0004 (0.0357) {0.0540}	0.0004 (0.0357) {0.0530}
0.0	-0.0012 (0.0391) {0.0450}	-0.0008 (0.0390) {0.0450}	-0.0008 (0.0388) {0.0450}	-0.0004 (0.0392) {0.0480}	-0.0004 (0.0392) {0.0480}
0.4	-0.0014 (0.0321) {0.0440}	-0.0010 (0.0330) {0.0460}	-0.0015 (0.0320) {0.0410}	-0.0009 (0.0324) {0.0440}	-0.0009 (0.0324) {0.0430}
0.8	-0.0009 (0.0160) {0.0420}	-0.0007 (0.0175) {0.0460}	-0.0010 (0.0156) {0.0400}	-0.0008 (0.0159) {0.0440}	-0.0008 (0.0159) {0.0450}
<i>spatial weight matrix: $M_{2,n}$ (6 or 4 neighbors)</i>					
-0.8	-0.0015 (0.0548) {0.0450}	-0.0012 (0.0526) {0.0470}	-0.0003 (0.0509) {0.0500}	0.0002 (0.0503) {0.0550}	0.0002 (0.0503) {0.0550}
-0.4	-0.0019 (0.0549) {0.0450}	-0.0016 (0.0533) {0.0470}	-0.0011 (0.0528) {0.0520}	-0.0005 (0.0528) {0.0570}	-0.0005 (0.0527) {0.0570}
0.0	-0.0020 (0.0487) {0.0430}	-0.0017 (0.0481) {0.0470}	-0.0017 (0.0481) {0.0470}	-0.0012 (0.0481) {0.0500}	-0.0012 (0.0481) {0.0510}
0.4	-0.0018 (0.0364) {0.0410}	-0.0016 (0.0369) {0.0470}	-0.0018 (0.0364) {0.0430}	-0.0015 (0.0364) {0.0450}	-0.0015 (0.0364) {0.0450}
0.8	-0.0010 (0.0173) {0.0410}	-0.0009 (0.0184) {0.0470}	-0.0010 (0.0172) {0.0420}	-0.0010 (0.0170) {0.0430}	-0.0010 (0.0170) {0.0440}

Notes: See Table 2.

Table 4: Bias, (RMSE) and {sizes} for heteroskedastic errors, $n = 100$

ρ	KP-NLS	KP-GMM	MLAM1	MLAM2	KP-eff	MLAM-eff
<i>het. robust</i>	No	Yes	Yes		Yes	
<i>spatial weight matrix: $M_{1,n}$ (8 or 2 neighbors)</i>						
-0.8	0.0029 (0.1182) {0.0210}	-0.0061 (0.1093) {0.0480}	-0.0048 (0.1082) {0.0410}	0.0046 (0.0840) {0.0360}	0.0102 (0.0858) {0.0510}	0.0104 (0.0849) {0.0460}
-0.4	-0.0624 (0.1921) {0.0230}	-0.0119 (0.1618) {0.0500}	-0.0101 (0.1553) {0.0510}	0.0014 (0.1370) {0.0580}	0.0123 (0.1377) {0.0640}	0.0134 (0.1368) {0.0620}
0.0	-0.0852 (0.1911) {0.0270}	-0.0149 (0.1591) {0.0490}	-0.0126 (0.1532) {0.0560}	-0.0087 (0.1490) {0.0650}	0.0052 (0.1470) {0.0700}	0.0068 (0.1463) {0.0710}
0.4	-0.0823 (0.1586) {0.0420}	-0.0153 (0.1278) {0.0450}	-0.0129 (0.1247) {0.0570}	-0.0156 (0.1271) {0.0570}	-0.0034 (0.1255) {0.0750}	-0.0020 (0.1249) {0.0740}
0.8	-0.0499 (0.0868) {0.0690}	-0.0096 (0.0667) {0.0440}	-0.0080 (0.0673) {0.0590}	-0.0113 (0.0675) {0.0510}	-0.0070 (0.0685) {0.0730}	-0.0066 (0.0686) {0.0750}
<i>spatial weight matrix: $M_{2,n}$ (6 or 4 neighbors)</i>						
-0.8	-0.0029 (0.1566) {0.0250}	0.0027 (0.1577) {0.0310}	0.0050 (0.1520) {0.0330}	0.0097 (0.1490) {0.0270}	0.0119 (0.1537) {0.0310}	0.0116 (0.1531) {0.0290}
-0.4	-0.0269 (0.1878) {0.0510}	-0.0193 (0.1862) {0.0630}	-0.0154 (0.1798) {0.0650}	-0.0084 (0.1710) {0.0660}	-0.0048 (0.1767) {0.0780}	-0.0048 (0.1760) {0.0730}
0.0	-0.0281 (0.1676) {0.0520}	-0.0208 (0.1658) {0.0660}	-0.0173 (0.1624) {0.0650}	-0.0162 (0.1595) {0.0660}	-0.0116 (0.1627) {0.0770}	-0.0115 (0.1623) {0.0770}
0.4	-0.0245 (0.1276) {0.0510}	-0.0185 (0.1259) {0.0650}	-0.0160 (0.1260) {0.0640}	-0.0181 (0.1250) {0.0660}	-0.0149 (0.1262) {0.0740}	-0.0148 (0.1261) {0.0730}
0.8	-0.0137 (0.0634) {0.0470}	-0.0104 (0.0624) {0.0630}	-0.0093 (0.0649) {0.0590}	-0.0109 (0.0621) {0.0600}	-0.0105 (0.0621) {0.0670}	-0.0104 (0.0621) {0.0670}

Notes: The errors are generated as $u = \rho Mu + \varepsilon$, $\varepsilon_i = \sigma_i \varepsilon_i$, $\varepsilon_i \sim i.i.d. \mathcal{N}(0,1)$, $\sigma_i^2 = d_i/5$. See Table 2 for more details.

Table 5: Bias, (RMSE) and {sizes} for heteroskedastic errors, $n = 1000$

ρ <i>het. robust</i>	KP-NLS No	KP-GMM Yes	MLAM1 Yes	MLAM2 Yes	KP-eff Yes	MLAM-eff Yes
<i>spatial weight matrix: $M_{1,n}$ (8 or 2 neighbors)</i>						
-0.8	0.0221 (0.0424) {0.0640}	-0.0019 (0.0342) {0.0480}	-0.0018 (0.0348) {0.0400}	-0.0005 (0.0250) {0.0520}	0.0003 (0.0242) {0.0500}	0.0002 (0.0241) {0.0490}
-0.4	-0.0455 (0.0727) {0.0650}	-0.0025 (0.0514) {0.0410}	-0.0024 (0.0493) {0.0410}	-0.0012 (0.0445) {0.0430}	0.0003 (0.0430) {0.0570}	0.0003 (0.0429) {0.0560}
0.0	-0.0734 (0.0914) {0.1880}	-0.0028 (0.0509) {0.0360}	-0.0026 (0.0488) {0.0400}	-0.0023 (0.0487) {0.0400}	-0.0004 (0.0468) {0.0490}	-0.0004 (0.0467) {0.0490}
0.4	-0.0713 (0.0830) {0.3490}	-0.0025 (0.0400) {0.0320}	-0.0024 (0.0390) {0.0380}	-0.0026 (0.0402) {0.0400}	-0.0011 (0.0390) {0.0440}	-0.0011 (0.0390) {0.0460}
0.8	-0.0417 (0.0468) {0.5210}	-0.0014 (0.0197) {0.0290}	-0.0013 (0.0199) {0.0400}	-0.0016 (0.0198) {0.0400}	-0.0012 (0.0197) {0.0380}	-0.0011 (0.0197) {0.0370}
<i>spatial weight matrix: $M_{2,n}$ (6 or 4 neighbors)</i>						
-0.8	-0.0081 (0.0575) {0.0440}	-0.0018 (0.0568) {0.0420}	-0.0015 (0.0544) {0.0480}	-0.0008 (0.0522) {0.0540}	-0.0001 (0.0516) {0.0570}	-0.0001 (0.0515) {0.0570}
-0.4	-0.0095 (0.0573) {0.0450}	-0.0021 (0.0564) {0.0430}	-0.0019 (0.0546) {0.0480}	-0.0014 (0.0538) {0.0580}	-0.0008 (0.0537) {0.0560}	-0.0008 (0.0537) {0.0560}
0.0	-0.0096 (0.0508) {0.0460}	-0.0022 (0.0497) {0.0430}	-0.0020 (0.0490) {0.0480}	-0.0020 (0.0490) {0.0490}	-0.0014 (0.0489) {0.0500}	-0.0014 (0.0489) {0.0500}
0.4	-0.0081 (0.0381) {0.0470}	-0.0019 (0.0371) {0.0440}	-0.0018 (0.0374) {0.0470}	-0.0019 (0.0371) {0.0460}	-0.0016 (0.0370) {0.0410}	-0.0016 (0.0370) {0.0410}
0.8	-0.0044 (0.0183) {0.0470}	-0.0011 (0.0176) {0.0430}	-0.0010 (0.0186) {0.0460}	-0.0011 (0.0175) {0.0410}	-0.0011 (0.0174) {0.0410}	-0.0011 (0.0174) {0.0410}

Notes: See Table 4.

Table 6: Bias and (RMSE) for the spatial autoregressive model, $n = 100$

ρ	$\hat{\rho}$			$\hat{\beta}$		
	ML	MLAM1	GS2SLS	ML	MLAM1	GS2SLS
$\beta = 1$						
-0.8	0.0105 (0.1250)	-0.0100 (0.1527)	-0.0175 (0.2984)	0.0105 (0.1033)	-0.0100 (0.1039)	-0.0175 (0.1171)
-0.4	0.0042 (0.1374)	0.0003 (0.1536)	-0.0389 (0.2966)	-0.0043 (0.0994)	-0.0051 (0.0992)	-0.0164 (0.1090)
0.0	-0.0118 (0.1326)	-0.0111 (0.1402)	-0.0249 (0.2755)	-0.0001 (0.1005)	-0.0008 (0.1004)	-0.0077 (0.1063)
0.4	-0.0200 (0.1016)	-0.0107 (0.1056)	-0.0372 (0.2430)	-0.0118 (0.1028)	-0.0111 (0.1031)	-0.0249 (0.1078)
0.8	-0.0108 (0.1059)	-0.0044 (0.1066)	-0.0201 (0.1091)	0.0047 (0.0523)	0.0015 (0.0565)	-0.0028 (0.1313)
$\beta = 0.5$						
-0.8	0.0159 (0.1442)	-0.0075 (0.1718)	-0.0774 (0.9158)	0.0020 (0.1022)	0.0003 (0.1019)	-0.0270 (0.1317)
-0.4	-0.0015 (0.1593)	-0.0083 (0.1750)	-0.1404 (1.0009)	-0.0010 (0.1028)	-0.0016 (0.1030)	-0.0249 (0.1500)
0.0	-0.0144 (0.1413)	-0.0080 (0.1455)	-0.1819 (0.9676)	-0.0020 (0.0970)	-0.0021 (0.0971)	-0.0272 (0.1419)
0.4	-0.0175 (0.1160)	-0.0045 (0.1182)	-0.1646 (0.9583)	-0.0056 (0.0997)	-0.0062 (0.0997)	-0.0241 (0.1368)
0.8	-0.0158 (0.0565)	-0.0051 (0.0613)	-0.0589 (0.8154)	0.0014 (0.0999)	-0.0011 (0.1000)	-0.0078 (0.1249)
$\beta = 0.1$						
-0.8	0.0164 (0.1517)	-0.0114 (0.1812)	0.5711 (4.3060)	-0.0021 (0.0990)	-0.0024 (0.0990)	-0.0378 (0.2910)
-0.4	-0.0020 (0.1665)	-0.0086 (0.1773)	0.7920 (4.1379)	0.0014 (0.1012)	0.0015 (0.1014)	-0.0185 (0.2764)
0.0	0.0006 (0.1613)	0.0006 (0.1666)	-0.0130 (3.6199)	-0.0134 (0.1020)	-0.0072 (0.1021)	0.4847 (0.2428)
0.4	0.0034 (0.1230)	0.0033 (0.1347)	-0.0069 (4.6602)	-0.0211 (0.0986)	-0.0082 (0.0988)	0.2432 (0.2474)
0.8	-0.0125 (0.0585)	0.0040 (0.0910)	0.2275 (1.3377)	0.0026 (0.1005)	0.0020 (0.0990)	0.0007 (0.1320)

Notes: The data are generated by the model $y = \gamma W_n y + X\beta + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, I)$ and $W_n = M_{2,n}$ (see table 1). Entries report bias (without brackets) and RMSE (round brackets). ML indicates the maximum likelihood estimator as implemented in the Matlab Econometrics Toolbox developed by James LeSage. The MLAM1 estimator is based on the moment equations (30)–(32). GS2SLS is the GMM estimator based on the matrix of instruments $Z = [X, WX, W^2X]$.

Appendix

Proof of Theorem 1

The moments of the MLAM estimator can be represented as

$$\begin{aligned} m_k(\rho) &= \mathbb{E} \left[\frac{1}{n} u' (I - \rho M_n) A_{k,n}(\rho) (I - \rho M_n') u \right] \\ &= \mathbb{E} \left[\frac{1}{n} u' \tilde{A}_{k,n}(\rho) u \right], \end{aligned}$$

where

$$\begin{aligned} \tilde{A}_1(\rho) &= M_n - \rho(M_n' M_n + M_n^2) + \rho^2 M_n' M_n^2 \\ \tilde{A}_2(\rho) &= M_n + \rho(\tilde{M}_n - M_n^2 - M_n' M_n) + \rho^2(M_n' M_n^2 - \tilde{M}_n M_n - M_n' \tilde{M}_n) + \rho^3 M_n' \tilde{M}_n M_n. \end{aligned}$$

A first order Taylor expansion around ρ_0 yields

$$0 = m_{k,n}(\hat{\rho}) = m_{k,n}(\rho_0) + \psi_{k,n}(\hat{\rho} - \rho_0) + o_p(n^{-1/2}),$$

where

$$\begin{aligned} \psi_{k,n} &= \frac{1}{n} u' D_{k,n}(\rho_0) u \\ \text{and } D_{k,n}(\rho_0) &= \left. \frac{\partial \tilde{A}_{k,n}(\rho)}{\partial \rho} \right|_{\rho=\rho_0}. \end{aligned}$$

It follows that

$$\mathbb{E}[n(\hat{\rho} - \rho_0)^2] = \mathbb{E} \left[\frac{nm_{k,n}(\rho_0)^2}{\psi_{k,n}^2} \right] \xrightarrow{n \rightarrow \infty} \frac{V_k}{\psi_k^2},$$

where V_k is defined in (18) and $\psi_k = \lim_{n \rightarrow \infty} \mathbb{E}(\psi_{k,n})$. The derivatives $D_{k,n}(\rho_0)$ can easily be found by differentiating (14) and (15) yielding the results for $\psi_{k,n}$ as presented in the theorem.

Proof of Theorem 2

Following Newey & McFadden (1994) the asymptotic distribution of the nonlinear GMM estimator is given by

$$\sqrt{n}(\hat{\rho}_{opt} - \rho_0) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}_\rho),$$

where

$$\mathcal{V}_\rho = [D(\rho_0)' \bar{Y}(\rho_0) D(\rho_0)]^{-1},$$

$$D(\rho_0) = \lim_{n \rightarrow \infty} \mathbb{E} \left(\frac{\partial m_n(\rho)}{\partial \rho} \right) \Big|_{\rho=\rho_0} \quad \text{and} \quad \bar{Y}(\rho_0) = \lim_{n \rightarrow \infty} Y_n(\rho_0).$$

Using the representation of the moment conditions as in 6 and 7 it follows that

$$\begin{aligned} \frac{\partial m_n(\rho)}{\partial \rho} \Big|_{\rho=\rho_0} &= \frac{\partial}{\partial \rho} \left[\begin{matrix} m_{1,n}(\rho) \\ \bar{m}_{2,n}(\rho) \end{matrix} \right] \Big|_{\rho=\rho_0} \\ &= \frac{1}{n} \frac{\partial}{\partial \rho} \left[\begin{matrix} u' B_n(\rho)' M_n B_n(\rho) u \\ u' B_n(\rho)' \bar{M}_n B_n(\rho) u \end{matrix} \right] \Big|_{\rho=\rho_0} \\ &= \frac{1}{n} \frac{\partial}{\partial \rho} \left[\begin{matrix} (u - \rho M_n u)' M_n (u - \rho M_n u) \\ (u - \rho M_n u)' \bar{M}_n (u - \rho M_n u) \end{matrix} \right] \Big|_{\rho=\rho_0} \\ &= \frac{1}{n} \left[\begin{matrix} u' (2\hat{\rho} M_n' M_n^2 - M_n' M_n - M_n^2) u \\ u' (2\hat{\rho} M_n' \bar{M}_n M_n - M_n' \bar{M}_n - \bar{M}_n M_n) u \end{matrix} \right]. \end{aligned}$$

The asymptotic variance follows straightforwardly.