No. 89

UNIVERSITY OF COLOGNE

WORKING PAPER SERIES IN ECONOMICS

ALTERNATIVE GMM ESTIMATORS FOR SPATIAL REGRESSION MODELS

JÖRG BREITUNG
CHRISTOPH WIGGER

Department of Economics
University of Cologne
Albertus-Magnus-Platz
D-50923 Köln
Germany

http://www.wiso.uni-koeln.de

Alternative GMM Estimators for Spatial Regression Models

Jörg Breitung[a] and Christoph Wigger[b]

18th January 2017

Abstract

Using approximations of the score of the log-likelihood function we derive optimal moment conditions for estimating spatial regression models. Our approach results in computationally simple and robust estimators. The moment conditions resemble those proposed by Kelejian & Prucha (1999), hence we provide an intuitive interpretation of their estimator as a second order approximation to the log-likelihood function. Furthermore we propose simplified and efficient GMM estimators based on a convenient modification of the moment conditions. Heteroskedasticity robust versions of our estimators are also provided. Finally, a first order approximation for the spatial lag model is also considered. Monte Carlo results suggest that a simple just-identified estimator based on a quadratic moment derived from a first order approximation of the score of the log-likelihood function performs similar to the GMM estimator proposed by Kelejian & Prucha (2010).

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

JEL Classification Numbers: C01, C13, C31.

[a]Center 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*.

[b]Department 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`.

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 two 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".

Recently much progress has been made in the estimation of spatial models. In particular, starting with the pioneering work of Kelejian & Prucha (1999) Generalized Method of Moments (GMM) estimators have been developed as an alternative to Maximum-likelihood (ML) estimation of spatial regression models.[2] The main advantage of GMM methods is that the inversion of $n \times n$ matrices in each iteration step is avoided, which become computationally demanding whenever the sample size n is large. Furthermore, GMM estimators require weaker distributional assumptions and are robust to heteroskedasticity and deviations from normality, cf. (Kelejian & Prucha 1999, Anselin 1988, Lin & Lee 2010).

In this paper we 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 SARAR(0,1) model. Additionally we consider GMM estimation of the spatial lag (Durbin) model (SARAR(1,0)). However, the proposed estimators can easily be employed for estimating more general spatial models, such as the SARAR(1,1) model. Another possible extension is the estimation of spatial panel data models, see (Kapoor et al. 2007).

The main contribution of this paper is the idea to derive optimal moment conditions by applying approximations to the first order condition of the ML estimator. In

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

[2] The work of Kelejian & Prucha (1999) as well as Kelejian & Prucha (1998) is developed further in articles by Kelejian & Prucha (2001, 2007, 2010), Lee (2003, 2004, 2007), Lin & Lee (2010) Liu et al. (2010), Arnold & Wied (2010) and Drukker et al. (2013) among others.

the following we call the resulting estimators *Maximum Likelihood Approximate Moment* (MLAM) estimators. We apply two approximations and find that the resulting moment conditions 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) 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). Moreover, the results suggest that the proposed simple moment estimators based on a single quadratic moment condition perform very well in comparison to all other GMM estimators. This is remarkable as the proposed simple estimators are computationally less demanding than the overidentified GMM estimators involving optimal weighting of the moment conditions.

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 lag 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 autoregressive model

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

$$y = X\beta + u \tag{1}$$

$$u = \rho_0 M_n u + \varepsilon, \tag{2}$$

where the vector $y = [y_1, \ldots, y_n]'$ 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 4. *(a) $\sum_{j=1}^{n} |m_{ij,n}| < C_m$ and $\sum_{i=1}^{n} |m_{ij,n}| < C_m$ for all $1 \leq i, j \leq n$ and some positive constant $C_m < \infty$, where $m_{ij,n}$ denotes the (i, j)-element of M_n. (b) $\sum_{j=1}^{n} |b_{ij,n}| < C_b$ and $\sum_{i=1}^{n} |b_{ij,n}| < C_b$ for all $1 \leq i, j \leq n$ and some positive constant $C_b < \infty$, where $b_{ij,n}$ denotes the (i, j)-element of $B_n(\rho)$ with $|\rho| < 1$ and C_b may depend on ρ.*

These assumptions are standard in the literature on GMM estimation of the linear spatial error model (e.g. Kelejian & Prucha (1999), Lee (2003) and Lin & Lee (2010)). Typically the row sums of the spatial weight matrix M_n are normalized to unity. If 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). However, row-normalization also implies that M_n may depend on the sample size n and, thus, form triangular arrays, as noted by Kelejian & Prucha (1999).

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

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

$$\widehat{\beta}_{gls,n} = \left[X'(I_n - \widehat{\rho}M_n')(I_n - \widehat{\rho}M_n)X\right]^{-1} X'(I_n - \widehat{\rho}M_n')(I_n - \widehat{\rho}M_n)y.$$

According to Assumption 5 we have $\widehat{u} = u + O_p(n^{-1/2})$ and we treat u (resp. β) as known for now. The log-likelihood function of the model results as

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

Concentrating out σ^2 yields the first order condition for maximizing the log likelihood

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

Note that the first order condition involves the inverse of an $n \times n$ matrix $B_n(\rho_0)$ which becomes 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

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

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

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

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

where $\overline{M}_n = M_n'M_n - tr(n^{-1}M_n'M_n)I_n$. 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 conditions in (5) is equivalent to minimizing $S_{KP} = m_n(\rho)'\,W_{KP,n}\,m_n(\rho)$ with $m_n(\rho) = [m_{1,n}(\rho), \overline{m}_{2,n}(\rho)]'$ and the weighting matrix $W_{KP,n} = \operatorname{diag}\{1, v\}$ with $v = 1/\left[1 + \left(n^{-1}\operatorname{tr}\{M_n'M_n\}\right)^2\right]$. 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

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

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

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

$$B_n(\rho_0)^{-1} = (I - \rho_0 M_n)^{-1} = I_n + \rho_0 M_n + \rho_0^2 M_n^2 + \cdots . \tag{9}$$

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

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

which clearly holds under assumptions 2 and 3 (a). Replacing u by its empirical counterpart $\widehat{u}$ 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

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

where

$$p_n = \frac{u'M_n'M_nu + u'M_n^2u}{2\,u'M_n'M_n^2u}$$
$$q_n = \left(\frac{u'M_n'M_nu + u'M_n^2u}{2\,u'M_n'M_n^2u}\right)^2 - \frac{u'M_nu}{u'M_n'M_n^2u}.$$

The MLAM1 estimator $\widehat{\rho}_1$ is the root that satisfies Assumption 3 (b), where in the above expressions u is replaced by $\widehat{u}$, yielding $\widehat{p}_n$ and $\widehat{q}_n$. Since $\widehat{q}_n$ 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 $\widehat{m}_{1,n}(\rho)^2$ (the empirical counterpart of $m_{1,n}(\rho)^2$, where in (10) u is replaced by $\widehat{u}$) which is due to the fact that the moment function is symmetric around the minimum.[3]

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

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

Under assumptions 2 and 3 (a) this condition holds, since in the case of homoskedastic disturbances $\mathbb{E}\left[\varepsilon'\left(\mathrm{diag}\{\rho_0 M_n^2\} - \frac{1}{n}\,\mathrm{tr}\{\rho_0 M_n^2\}I_n\right)\varepsilon\right] = 0$, where $\mathrm{diag}\{\rho_0 M_n^2\}$ 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

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

where $\widetilde{M}_n := M_n^2 - n^{-1}\,\mathrm{tr}(M_n^2)I_n$. Note that for a symmetric spatial weight matrix

[3] 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.

$M_n = M_n'$ the moment is equivalent to a linear combination of the moments suggested by Kelejian & Prucha (1999). In the case that M_n is asymmetric, which is usually the case if the row sums of M_n are normalized, MLAM2 exploits similar but not identical information. Substituting ε by $(I_n - \rho_0 M_n)u$, the moment condition can be written as a cubic polynomial in ρ:

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

Now let $\widehat{m}_{2,n}(\rho)$ be the empirical counterpart of (12) with u replaced by $\widehat{u}$. The MLAM2 estimator $\widehat{\rho}_2$ is the real root of $\widehat{m}_{2,n}(\rho)$ that satisfies Assumption 3 (b).[4] In a similar manner the m'th order approximation to the scores can be obtained from replacing $B_n(\rho)$ by $I + \rho M_n + \cdots + \rho^{m-1}M_n^{m-1}$. For $m \to \infty$ the m'th order MLAM converges to the ML estimator. Since the computational advantage over the ML estimator gets lost for MLAM estimators with $m > 2$ we focus on the first and second order approximation.

The moments of the MLAM estimator can be represented as a quadratic form given by

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

where

$$A_{1,n}(\rho_0) = M_n \quad (14)$$

$$A_{2,n}(\rho_0) = M_n + \rho_0 \widetilde{M}_n. \quad (15)$$

[4] In principle all real roots 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 $|\widehat{m}_{2,n}(\rho)|$ 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 $\widehat{\rho}_2$. However, unreported simulations suggest that this is not a relevant issue in practice.

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

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

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

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

Under Assumption 4 (a) the variance of $\xi_{k,i-1,n}$ is finite for all i and n.

Under assumption 2 it holds that $\mathbb{E}(\varepsilon_i|\xi_{k,i-1,n}) = 0$, $\mathbb{E}(z_i) = \mathbb{E}(z_i|\varepsilon_j) = 0$ for $j \neq i$ and that $\sum_{i=2}^{n}\varepsilon_i\xi_{k,i-1,n}$ and $\sum_{1=1}^{n} a_{k,ii,n}z_i$ 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 $\{\varepsilon_1, \ldots, \varepsilon_{i-1}\}$. 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

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

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 $n \to \infty$ the k'th order MLAM estimators are asymptotically distributed as*

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

where V_k is defined in (18) *and*

$$\psi_1 = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}u'(2\rho M_n' M_n^2 - M_n' M_n - M_n^2)u\right]$$

$$\psi_2 = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}u'(\widetilde{M}_n - M_n^2 - M_n' M_n)u + 2\rho u'(M_n' M_n^2 - \widetilde{M}_n M_n - M_n' \widetilde{M}_n)u \right.$$
$$\left. + 3\rho^2 u' M_n' \widetilde{M}_n M_n u\right].$$

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 $\widehat{u}$ and by inserting the estimator

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

where $\widehat{\varepsilon}_i$ is the i'th element of the vector $\widehat{\varepsilon} = (I - \widehat{\rho} M_n)\widehat{u}$. 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 (2010) and Drukker et al. (2013) propose GMM estimators with optimal weighting matrices W_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 $m_n(\rho) = [m_{1,n}(\rho), \overline{m}_{2,n}(\rho)]'$ or $m_n^*(\rho) = [m_{1,n}(\rho), \widetilde{m}_{2,n}(\rho)]'$ with $\overline{m}_{2,n}(\rho)$ defined in (7) and $\widetilde{m}_{2,n}(\rho) = \mathbb{E}(n^{-1}\varepsilon'\widetilde{M}_n\varepsilon)$. Since $m_{2,n}(\rho) = m_{1,n}(\rho) + \rho\widetilde{m}_{2,n}(\rho)$, which is easily seen in (11), the MLAM2 estimator is based on 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

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

with $\eta_{1,i,n} = \varepsilon_i \xi_{1,i-1,n}$, where $\xi_{1,i-1,n}$ is defined in (17), whereas the second moment condition can be represented as

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

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

It should be noted that $\mathbb{E}\left(\sum\limits_{i=1}^{n}\overline{m}_{ii,n}z_i\right) = 0$ 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 $\{\eta_{1,n}, \ldots, \eta_{i-1,n}\}$. Invoking the central limit theorem for martingale difference sequences yields

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

where $\overline{\mathcal{V}}_\eta = \lim_{n \to \infty} \mathcal{V}_{\eta,n}$ and

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

Let $\widehat{\varepsilon}_n(\rho) = (I_n - \rho M_n)\widehat{u}$ and $\widehat{\eta}_{i,n}(\rho)$ is constructed as $\eta_{i,n} \equiv \eta_{i,n}(\rho_0)$, where ε_i is replaced by the i'th element of $\widehat{\varepsilon}_n = I - \widehat{\rho}M_n$. An asymptotically efficient GMM estimator based on the KP moment conditions results from minimizing the objective function

$$\widehat{Q}_n(\rho) = \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{\eta}_{i,n}(\rho)\right)' \widehat{W}_n(\widehat{\rho}) \left(\frac{1}{n}\sum_{i=1}^{n} \widehat{\eta}_{i,n}(\rho)\right), \tag{24}$$

with

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

From the law of large numbers it follows that under assumptions 1 to 5 the estimator $\widehat{W}_n(\widehat{\rho})$ is consistent for the optimal weighting matrix $W_n(\rho_0)$ whenever the estimate $\widehat{\rho}$ is consistent for ρ_0. For example, a consistent initial estimate may be obtained by letting $W_n = W_{KP,n}$ 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 $n \to \infty$ the efficient GMM estimator $\widehat{\rho}_{opt} =$ argmin$\{\widehat{Q}_n(\rho)\}$ with $\widehat{Q}_n(\rho)$ based on the moment vector $\widehat{m}_n(\rho) = [\widehat{m}_{1,n}(\rho), \widehat{\overline{m}}_{2,n}(\rho)]'$ and defined in* (24) *possesses the limiting distribution*

$$\sqrt{n}(\widehat{\rho}_{opt} - \rho_0) \xrightarrow{d} \mathcal{N}\left(0, [\overline{\delta}(\rho_0)'\overline{W}(\rho_0)\overline{\delta}(\rho_0)]^{-1}\right),$$

where $\overline{W}(\rho_0) = \lim_{n\to\infty} W_n(\rho_0)$ *and* $\bar{\delta}(\rho_0) = \lim_{n\to\infty} \delta_n(\rho_0)$ *with*

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

The proof of theorem 2 is provided in the appendix.

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

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

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

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

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 (2010) as well as Liu et al. (2010) and Lin & Lee (2010) propose the alternative moment condition $\overline{m}_{2,n}^h(\rho) = \mathbb{E}\left[\varepsilon'\left(M_n'M_n - \operatorname{diag}\{M_n'M_n\}\right)\varepsilon\right] = \mathbb{E}(\varepsilon'\overline{M}_{0,n}\varepsilon)$, where $\operatorname{diag}\{M_n'M_n\}$ represents a diagonal matrix constructed by the diagonal elements of the matrix $M_n'M_n$. Accordingly, the matrix $\overline{M}_{0,n}$ is obtained by setting the diagonal elements of $M_n'M_n$ equal to zero. Similarly, a heteroskedasticity robust modification for the MLAM2 moment condition $m_{2,n}(\rho)$ is given by $m_{2,n}^h(\rho) = \mathbb{E}[\varepsilon'(M_n + \rho\widetilde{M}_{0,n})\varepsilon]$ with $\widetilde{M}_{0,n} = M_n^2 - \operatorname{diag}\{M_n^2\}$.

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 MLAM2 estimator the robust moment condition results as

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

where $\widetilde{\eta}_{2,i,n}^h = \varepsilon_i\xi_{2,i-1,n}$ and $\xi_{2,i-1,n}$ is defined in (17). The heteroscedasticity robust version of the moment used by the original KP estimator is given by

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

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

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

in (18) is replaced by

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

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 lag model

An alternative specification is the spatial lag (Durbin) model given by

$$y = \gamma W_n y + X\beta + \varepsilon,$$

where γ and W_n correspond to ρ and M_n in the spatial error model. Accordingly, Assmptions 3 and 4 apply to γ and W_n. Assuming $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ yields the log-likelihood function

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

with the gradients[5]

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

The reduced form representation of the model

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

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

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

$$\begin{aligned}
\frac{\partial \ell(\cdot)}{\partial \gamma} &= -tr[W_n(I-\gamma W_n)^{-1}] + \frac{1}{\sigma^2}(y-\gamma W_n y - X\beta)' W_n (I-\gamma W_n)^{-1}(X\beta + \varepsilon) \\
&= \mu(\gamma) + \left[\frac{1}{\sigma^2}(y-\gamma W_n y - X\beta)' W_n (I-\gamma W_n)^{-1} X\beta\right] \\
&\quad + \left[\frac{1}{\sigma^2}(y-\gamma W_n y - X\beta)' W_n (I-\gamma W_n)^{-1}\varepsilon\right] \\
&\equiv \mu(\gamma) + g_1(\beta,\gamma) + g_2(\beta,\gamma),
\end{aligned}$$

where we split the gradient into the deterministic term $\mu(\gamma)$ and two stochastic terms $g_1(\beta,\gamma) = \varepsilon' W_n (I-\gamma W_n)^{-1} X\beta/\sigma^2$ and $g_2(\beta,\gamma) = \varepsilon' W_n (I-\gamma W_n)^{-1}\varepsilon/\sigma^2$. Using the first order approximation $(I-\gamma W_n)^{-1} \approx I_n$ gives rise to the following set of moment conditions for estimating β and γ

$$\widetilde{g}_0(\beta,\gamma) = X'\varepsilon = 0 \tag{29}$$

$$\widetilde{g}_1(\beta,\gamma) = \beta' X' W_n' \varepsilon = 0 \tag{30}$$

$$\widetilde{g}_2(\beta,\gamma) = \varepsilon' W_n \varepsilon = 0. \tag{31}$$

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

The moment condition $\widetilde{g}_1(\beta,\gamma)$ can be interpreted as a linear combination of the k moment conditions $\widetilde{g}_1^v(\beta,\gamma) = X' W_n' \varepsilon = 0$. While the set of moment conditions $\widetilde{g}(\beta,\gamma) = (\widetilde{g}_0(\beta,\gamma)', \widetilde{g}_1^v(\beta,\gamma)', \widetilde{g}_2(\beta,\gamma))'$ leads to an over-identified GMM estimator whenever the number of regressors is larger than one, the linear moment conditions in $\widetilde{g}_1^v(\beta,\gamma)$ are computationally simpler than the nonlinear moment condition $\widetilde{g}_1(\beta,\gamma)$. The optimal weight matrix can be constructed as in section 4. In the following we refer to the resulting estimator as the MLAM1 estimator for the spatial lag model.

It is interesting to note that the two-stage instrumental variable estimator of Anselin (1988) and the GMM estimator of Kelejian & Prucha (1998) are based on 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 $(X, W_n X,\ W_n^2 X, \ldots, W_n^p X)$ whereas the term $g_2(\beta, \gamma)$ is ignored. As pointed out by Lee (2003) this may result in a dramatic loss of efficiency, in particular if β is close to zero.

7 Monte Carlo comparison

We now compare the properties of alternative MM and GMM estimators for the spatial error model 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" $(i - r_i, \ldots, i-1)$ and r_i "behind" $(i+1, \ldots, i+r_i)$.[6] 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 the 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:

[6] Since we consider a "circular world" observation 1 and observation n are direct neighbors. Hence, for observation $i = n$, the jth "behind"-neighbor $i + j$ is observation j.

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 $\varepsilon \sim \mathcal{N}(0, \sigma_0^2 I_n)$ 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). Specifically $\varepsilon_i = \sigma_{0,i}\xi_i$, where ξ_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 moment estimators based on a first and second order approximation of the likelihood function (see Section 3). 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 MLAM moment vectors and the 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.[7] The MC results allow us 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 SARAR(0,1) 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

[7] This is done by running a t-test for each estimate $\widehat{\rho}$ 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. Hence, for each estimator we expect a (wrong) rejection to occur in 5% of the MC replications.

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 ($M_{2,n}$) 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 MLAM 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. Still, however the MLAM2 or even the MLAM1 estimator 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 MLAM 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.[8]

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

[8] 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 $\overline{M}_n \equiv M_n' M_n = M_n M_n \equiv \widetilde{M}_n$.

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 lag 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 regressor is generated as $x_i \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$ 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 perfoms much better than the GS2SLS estimator, in particular if β is small. The reason is that for small β the instruments W^kX 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-Gausian 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. 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 (2010) 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 lag (Durbin) model.

Our MC results suggest that the efficient GMM estimators are (slightly) more efficient 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 lag 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, Drodrecht.

Anselin, L. (2010), 'Thirty years of spatial econometrics', *Papers in Regional Science* **89**, 3–25.

Arnold, M. & Wied, D. (2010), ' Improved GMM estimation of the spatial autoregressive error model', *Economics Letters* **108**, 65–68.

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**, 1879–2308.

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 innovation activity in Europe: are social capital and geographical proximity key determinants? ', *Applied Economics* **45**, 2325–2335.

Drukker, D., 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.

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. & Prucha, I. R. (1998), 'A Generalized Spatial Two Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autorgressive 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. (2007), 'HAC estimation in a spatial framework', *Journal of Econometrics* **140**, 131–154.

Kelejian, H. H. & Prucha, I. R. (2010), 'Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances', *Journal of Econometrics* **157**, 53–67.

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. (2007), 'GMM and 2SLS estiamtion of mixed regressive, spatial autoregressive models', *Journal of Econometrics* **137**, 489–514.

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 estiamtion of spatial autoregressive models with unknown heteroskedasticity ', *Journal of Econometrics* **157**, 34–52.

Liu, X., Lee, L.-f. & Bollinger, C. R. (2010), ' An efficient GMM estimator of spatial autoregressive models', *Journal of Econometrics* **159**, 303–319.

Newey, W. & McFadden, D. (1994), large sample estimation and hypothesis testing, *in* R. Engle & D. McFadden, eds, 'Handbook of Econometrics. Volume IV', North-Holland, Amsterdam, 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. Fischer & P. Nijkamp, eds, 'Handbook of Regional Science, Spatial Econometrics', Springer Verlag, pp. 1597–1618.

Tables

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

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

Notes: The number of MC replications is 1000. The errors are generated as $u = \rho M u + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, I)$. 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
spatial weight matrix: $M_{1,n}$ (8 or 2 neighbors)					
-0.8	0.0003	0.0000	0.0008	0.0006	0.0006
	(0.0197)	(0.0217)	(0.0182)	(0.0193)	(0.0193)
	{0.0480}	{0.0440}	{0.0510}	{0.0690}	{0.0670}
-0.4	-0.0004	-0.0003	0.0005	0.0004	0.0004
	(0.0359)	(0.0362)	(0.0350)	(0.0357)	(0.0357)
	{0.0500}	{0.0450}	{0.0480}	{0.0540}	{0.0530}
0.0	-0.0012	-0.0008	-0.0008	-0.0004	-0.0004
	(0.0391)	(0.0390)	(0.0388)	(0.0392)	(0.0392)
	{0.0450}	{0.0450}	{0.0450}	{0.0480}	{0.0480}
0.4	-0.0014	-0.0010	-0.0015	-0.0009	-0.0009
	(0.0321)	(0.0330)	(0.0320)	(0.0324)	(0.0324)
	{0.0440}	{0.0460}	{0.0410}	{0.0440}	{0.0430}
0.8	-0.0009	-0.0007	-0.0010	-0.0008	-0.0008
	(0.0160)	(0.0175)	(0.0156)	(0.0159)	(0.0159)
	{0.0420}	{0.0460}	{0.0400}	{0.0440}	{0.0450}
spatial weight matrix: $M_{2,n}$ (6 or 4 neighbors)					
-0.8	-0.0015	-0.0012	-0.0003	0.0002	0.0002
	(0.0548)	(0.0526)	(0.0509)	(0.0503)	(0.0503)
	{0.0450}	{0.0470}	{0.0500}	{0.0550}	{0.0550}
-0.4	-0.0019	-0.0016	-0.0011	-0.0005	-0.0005
	(0.0549)	(0.0533)	(0.0528)	(0.0528)	(0.0527)
	{0.0450}	{0.0470}	{0.0520}	{0.0570}	{0.0570}
0.0	-0.0020	-0.0017	-0.0017	-0.0012	-0.0012
	(0.0487)	(0.0481)	(0.0481)	(0.0481)	(0.0481)
	{0.0430}	{0.0470}	{0.0470}	{0.0500}	{0.0510}
0.4	-0.0018	-0.0016	-0.0018	-0.0015	-0.0015
	(0.0364)	(0.0369)	(0.0364)	(0.0364)	(0.0364)
	{0.0410}	{0.0470}	{0.0430}	{0.0450}	{0.0450}
0.8	-0.0010	-0.0009	-0.0010	-0.0010	-0.0010
	(0.0173)	(0.0184)	(0.0172)	(0.0170)	(0.0170)
	{0.0410}	{0.0470}	{0.0420}	{0.0430}	{0.0440}

Notes: See Table 2.

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

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

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

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

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

Notes: See Table 4.

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

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

Notes: The data are generated by the model $y = \gamma W_n y + X\beta + \varepsilon$, where $\varepsilon \sim \mathcal{N}(0, I)$. 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 (29)–(31). 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

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

where

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

A first order Taylor expansion around ρ_0 yields

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

where

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

It follows that

$$\mathbb{E}[n(\widehat{\rho}-\rho_0)^2] = \mathbb{E}\left[\frac{n m_{k,n}(\rho_0)^2}{\psi_{k,n}^2}\right] \underset{n\to\infty}{\longrightarrow} \frac{V_k}{\psi_k^2},$$

where V_k is defined in (18) and $\psi_k = \lim_{n\to\infty} \mathbb{E}(\psi_{k,n})$. 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

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

where

$$\begin{aligned}
\mathcal{V}_\rho &= \left[D(\rho_0)'\overline{W}(\rho_0)D(\rho_0)\right]^{-1} , \\
D(\rho_0) &= \lim_{n\to\infty} \mathbb{E}\left(\frac{\partial m_n(\rho)}{\partial\rho}\right)\Big|_{\rho=\rho_0} \quad \text{and} \quad \overline{W}(\rho_0) = \lim_{n\to\infty} W_n(\rho_0).
\end{aligned}$$

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

$$\begin{aligned}
\frac{\partial m_n(\rho)}{\partial\rho}\Big|_{\rho=\rho_0} &= \frac{\partial}{\partial\rho}\begin{bmatrix} m_{1,n}(\rho) \\ \overline{m}_{2,n}(\rho)\end{bmatrix}\Big|_{\rho=\rho_0} \\
&= \frac{1}{n}\frac{\partial}{\partial\rho}\begin{bmatrix} u'B_n(\rho)'M_nB_n(\rho)u \\ u'B_n(\rho)'\overline{M}_nB_n(\rho)u\end{bmatrix}\Big|_{\rho=\rho_0} \\
&= \frac{1}{n}\frac{\partial}{\partial\rho}\begin{bmatrix} (u-\rho M_nu)'M_n(u-\rho M_nu) \\ (u-\rho M_nu)'\overline{M}_n(u-\rho M_nu)\end{bmatrix}\Big|_{\rho=\rho_0} \\
&=\frac{1}{n}\begin{bmatrix} u'(2\widehat{\rho}M_n'M_n^2 - M_n'M_n - M_n^2)u \\ u'(2\widehat{\rho}M_n'\overline{M}_nM_n - M_n'\overline{M}_n - \overline{M}_nM_n)u\end{bmatrix} .
\end{aligned}$$

The asymptotic variance follows straightforwardly.