Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances

Harry H. Kelejian and Ingmar R. Prucha*
Department of Economics
University of Maryland, College Park, MD 20742

Revised September 2005
Revised July 2007

Abstract

One important goal of this study is to develop a methodology of inference for a widely used Cliff-Ord type spatial model containing spatial lags in the dependent variable, exogenous variables, and the disturbance terms, while allowing for unknown heteroskedasticity in the innovations. We first generalize the generalized moments (GM) estimator suggested in Kelejian and Prucha (1998,1999) for the spatial autoregressive parameter in the disturbance process. We prove the consistency of our estimator; unlike in our earlier paper we also determine its asymptotic distribution, and discuss issues of efficiency. We then define instrumental variable (IV) estimators for the regression parameters of the model and give results concerning the joint asymptotic distribution of those estimators and the GM estimator under reasonable conditions. Much of the theory is kept general to cover a wide range of settings. We note the estimation theory developed by Kelejian and Prucha (1998, 1999) for GM and IV estimators and by Lee (2004) for the quasi-maximum likelihood estimator under the assumption of homoskedastic innovations does not carry over to the case of heteroskedastic innovations. The paper also provides a critical discussion of the usual specification of the parameter space.

Key Words: Spatial dependence, heteroskedasticity, Cliff-Ord model, two-stage least squares, generalized moments estimation, asymptotics

JEL Classification: C21, C31

---

*Corresponding Author: Ingmar R. Prucha, Tel.: 301-405-3499, Email: prucha@econ.umd.edu.

1 Introduction[1]

In recent years the economics literature has seen an increasing number of theoretical and applied econometric studies involving spatial issues.[2] While this increase in interest in spatial models in economics is relatively recent, spatial models have a long history in the regional science and geography literature.[3] One of the most widely referenced model of spatial interactions is one that was put forth by Cliff and Ord (1973, 1981). This model is a variant of the model considered by Whittle (1954). In its simplest (and original) form the model only considers spatial spillovers in the dependent variable, and specifies the endogenous variable corresponding to a cross sectional unit in terms of a weighted average of endogenous variables corresponding to other cross sectional units, plus a disturbance term. This model is typically referred to as a spatial autoregressive model, the weighted average is typically referred to as a spatial lag, the corresponding parameter as the autoregressive parameter, and the matrix containing the weights as the spatial weights matrix. A generalized version of this model also allows for the dependent variable to depend on a set of exogenous variables and spatial lags thereof. A further generalization allows for the disturbances to be generated by a spatial autoregressive process. Consistent with the terminology developed by Anselin and Florax (1995) we refer to the combined model as a spatial autoregressive model with autoregressive disturbances of order $(1, 1)$, for short SARAR$(1, 1)$. We note that this model is fairly general in that it allows for spatial spillovers in the endogenous variables, exogenous variables and disturbances.

Somewhat surprisingly, even though the SARAR$(1, 1)$ model has been a modeling tool for many years, until recently there has been a lack of formal results concerning estimation methods for this model. One method that has been employed to estimate this model is the (quasi) maximum likelihood (ML)

[2] Some recent applications of spatial models are, e.g., Audretsch and Feldmann (1996), Baltagi, Egger and Pfaffermayr (2005), Bell and Bockstael (2000), Besley and Case (1995), Betrand, Luttmer and Mullainathan (2000), Case (1991), Case, Hines, and Rosen (1993), Cohen and Morrison Paul (2004), Hanushek et al. (2003), Holtz-Eakin (1994), Sacerdote (2001), Shroder (1995), Topa (2001). Contributions to the theoretical econometric literature include, e.g., Baltagi and Li (2001a,b, 1999), Baltagi, Song, Jung and Koh (2005), Baltagi, Song and Koh (2003), Bao and Ullah (2003), Conley (1999), Das, Kelejian and Prucha (2003), Driscol and Kraay (1998), Kapoor, Kelejian and Prucha (2004), Kelejian and Prucha (2005, 2004, 2002, 2001, 1999, 1998), Korniotis (2005), Lee (2005, 2004, 2003, 2002), Pinkse and Slade (1998), Pinkse, Slade, and Brett (2002), Yang (2005).

[3] See, e.g., Anselin (1988), Bennett and Hordijk (1986), Cliff and Ord (1973, 1981), and Cressie (1993) and the references cited therein.

procedure, where the likelihood function corresponds to the normal distribution. Formal results concerning the asymptotic properties of the ML estimator have been established only recently in an important contribution by Lee (2004). Given that the likelihood function involves the determinant of a matrix whose dimensions depend on the sample size and an unknown parameter, there can be significant difficulties in the practical computation of this estimator especially if the sample size is large, as it might be if the spatial units relate to counties, single family houses, etc. In part because of this Kelejian and Prucha (1999) introduced a generalized moments (GM) estimator for the autoregressive parameter of the disturbance process that is simple to compute and remains computationally feasible even for large sample sizes. In Kelejian and Prucha (1998) we used that GM estimator to introduce a generalized spatial two stage least squares estimator (GS2SLS) for the regression parameters of the spatial $SARAR(1,1)$ model that is again simple to compute, and demonstrated its consistency and asymptotic normality.[4]

All of the above estimators for the $SARAR(1,1)$ model were introduced and their asymptotic properties were derived under the assumption that the innovations in the disturbance process are homoskedastic. The lack of an estimation theory that allows for heteroskedasticity, and the lack of corresponding joint hypothesis tests for the presence of spatial dependencies in the endogenous variables, exogenous variables and/or disturbances, is a serious shortcoming. Spatial units are often heterogeneous in important characteristics, e.g., size, and hence the homoskedasticity assumption may not hold in many situations (conditionally and unconditionally). It is readily seen that if the innovations are heteroskedastic, the ML estimator considered in Lee (2004) is inconsistent, and the asymptotic distribution given in Kelejian and Prucha (1998) for the GS2SLS estimator is not appropriate. One important goal of this study is therefore to develop a methodology of inference for the $SARAR(1,1)$ model that allows for heteroskedastic innovations. In developing this theory we will adopt a modular approach such that much of the theory not only applies to the $SARAR(1,1)$ model, but can also be utilized in different settings in future research.

In more detail, in this paper we introduce a new class of GM estimators for the autoregressive parameter of a spatially autoregressive disturbance process that allows for heteroskedastic innovations. Our GM estimators are again computationally simple even in large samples. We determine their consistency; unlike in our earlier paper we also determine, under reasonably general conditions, their asymptotic distribution. Loosely speaking, in deriving those results we essentially only maintain that the disturbances are $n^{1/2}$-consistently estimated (where n is the sample size) and that the estimator of the model parameters employed in estimating the disturbances is asymptotically linear in the innova-

---

[4]The formulation of the GS2SLS estimator is based on an approximation of the ideal instruments. Recently Lee (2003) and Kelejian, Prucha, and Yuzefovich (2004) extended the analysis to include the use of ideal instruments. Das, Kelejian and Prucha (2003) analyzed the small sample properties of the GS2SLS (as well as those of other estimators). They find that in many situations the loss of efficiency due to the approximation of the ideal instruments is minor.

tions. As a result the methodology developed in this paper covers a wide range of (linear and nonlinear) models and estimators, in addition to the SARAR(1, 1) model and estimators specific for that model. We furthermore derive results concerning the joint distribution of the GM estimators and estimators of the regression parameters to facilitate joint tests. While the results are presented for the case of two step estimation procedures where the spatial autoregressive parameter and the regression parameters are estimated in separate steps, the analysis can be readily adapted to one step procedures where all parameters are estimated in a single (but numerically more involved) step.

The general theory is then applied to develop inference methodology for the SARAR(1, 1) model. In particular, we use the GM estimator in constructing a GS2SLS estimator for the regression parameters of the SARAR(1, 1) model and demonstrate the consistency and asymptotic normality of this estimator. We also provide results concerning the joint distribution of the GM estimator and the GS2SLS estimator, which permits, among other things, testing the joint hypothesis of the absence of spatial spillovers stemming from the endogenous variables, exogenous variables or disturbances.

Another concern with the existing literature on Cliff-Ord type models, including in the above cited literature on the SARAR(1, 1) models, is the specification of the parameter space for spatial autoregressive parameters. In virtually all of the literature it is assumed that the parameter space for autoregressive parameters is the interval $(-1, 1)$, or a subset thereof. One may conjecture that this traditional specification of the parameter space received its motivation from the time series literature. However, as discussed in detail below, choosing the interval $(-1, 1)$ as the parameter space for the autoregressive parameter of a spatial model is not natural in the sense that the spatial autoregressive parameter always appears in those models in product form with the spatial weights matrix. Hence equivalent model formulations are obtained by applying an (arbitrary) scale factor to the autoregressive parameter and its inverse to the weights matrix. Of course, applying a scale factor to the autoregressive parameter leads to a corresponding re-scaling of its parameter space. In this paper we therefore allow for a more general specification of the parameter space. Even if a scale factor is used that results in the parameter space being the interval $(-1, 1)$, this scale factor and correspondingly the autoregressive parameter will then typically depend on the sample size. In contrast to the existing literature we thus allow for the parameters to depend on the sample size. Our discussion of the parameter space and possible normalizations of the spatial weights matrix also points out potential pitfalls with the frequently used approach of row-normalizing the spatial weights matrix.

The paper is organized as follows: The generalized SARAR(1, 1) model is specified and interpreted in Section 2. This section also contains a discussion of the parameter space of the autoregressive parameter. In Section 3 we define and establish the large sample properties of our suggested GM estimators for the autoregressive parameter of a spatially autoregressive disturbance process. In this section we also provide results concerning the joint large sample distribution of the GM estimators and a wide class of estimator of the regression parameters.

4

We also develop HAC type estimators for the large sample variance-covariance matrix of the suggested estimators. Section 4 contains results relating to the suggested instrumental variable estimators of the regression parameters of the SARAR(1, 1) model and their joint large sample distribution with GM estimators. Concluding remarks are given in the Section 5. Technical details are relegated to the appendices.

It proves helpful to introduce the following notation: Let $\mathbf{A}_n$ with $n \in \mathbf{N}$ be some matrix; we then denote the (i, j)-th element of $\mathbf{A}_n$ as $a_{ij,n}$. Similarly, if $\mathbf{v}_n$ is a vector, then $v_{i,n}$ denotes the i-th element of $\mathbf{v}_n$. An analogous convention is adopted for matrices and vectors that do not depend on the index n, in which case the index n is suppressed on the elements. If $\mathbf{A}_n$ is a square matrix, then $\mathbf{A}_n^{-1}$ denotes the inverse of $\mathbf{A}_n$. If $\mathbf{A}_n$ is singular, then $\mathbf{A}_n^{-1}$ should be interpreted as the generalized inverse of $\mathbf{A}_n$. At times it will also be helpful to denote the generalized inverse more explicitly as $\mathbf{A}_n^+$. With $\mathbf{a}_{i\cdot,n}$ and $\mathbf{a}_{\cdot i,n}$ we denote the i-th row and column of $\mathbf{A}_n$, respectively, and with $\mathbf{a}_n^{i\cdot}$ and $\mathbf{a}_n^{\cdot i}$ those of $\mathbf{A}_n^{-1}$. If $\mathbf{A}_n$ is a square symmetric nonnegative matrix, then $\mathbf{A}_n^{1/2}$ denotes the unique symmetric and nonnegative square root of $\mathbf{A}_n$. If $\mathbf{A}_n$ is nonsingular, then $\mathbf{A}_n^{-1/2}$ denotes $(\mathbf{A}_n^{-1})^{1/2}$. Further, we say the row and column sums of the (sequence of) matrices $\mathbf{A}_n$ are bounded uniformly in absolute value if there exists a constant $c_A < \infty$ (that does not dependent of n) such that

$$\max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij,n}| \leq c_A \text{ and } \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij,n}| \leq c_A \text{ for all } n \in \mathbf{N}$$

holds. As a point of interest, we note that the above condition is identical to the condition that the sequences of the maximum column sum matrix norms and maximum row sum matrix norms of $\mathbf{A}_n$ are bounded; cp. Horn and Johnson (1985, pp.294-5). For definiteness, let $\mathbf{A}$ be some vector or matrix, then $\|\mathbf{A}\| = [Tr(\mathbf{A}'\mathbf{A})]^{1/2}$. We note that this norm is submultiplicative, i.e., $\|\mathbf{AB}\| \leq \|\mathbf{A}\| \|\mathbf{B}\|$.

2   Model

In this section we specify the generalized SARAR(1, 1) model and discuss the underlying assumptions.

2.1   Specification

Suppose a cross section of n spatial units is observed, and the interactions between those spatial units can be described by the following model:

$$\begin{aligned} \mathbf{y}_n &= \mathbf{X}_n \beta_n + \lambda_n \mathbf{W}_n \mathbf{y}_n + \mathbf{u}_n \\ &= \mathbf{Z}_n \boldsymbol{\delta}_n + \mathbf{u}_n \end{aligned} \tag{1}$$

and

$$\mathbf{u}_n = \rho_n \mathbf{M}_n \mathbf{u}_n + \boldsymbol{\varepsilon}_n, \tag{2}$$

with $\mathbf{Z}_n = [\mathbf{X}_n, \mathbf{W}_n \mathbf{y}]$ and $\boldsymbol{\delta}_n = [\beta_n', \lambda_n]'$. Here $\mathbf{y}_n$ denotes the $n \times 1$ vector of observations of the dependent variable, $\mathbf{X}_n$ denotes the $n \times k$ matrix of non-stochastic (exogenous) regressors, $\mathbf{W}_n$ and $\mathbf{M}_n$ are $n \times n$ non-stochastic matrices, $\mathbf{u}_n$ denotes the $n \times 1$ vector of regression disturbances, $\boldsymbol{\varepsilon}_n$ is an $n \times 1$ vector of innovations, λ_n and ρ_n are unknown scalar parameters, and β_n is a $k \times 1$ vector of unknown parameters. The matrices $\mathbf{W}_n$ and $\mathbf{M}_n$ are typically referred to as spatial weights matrices, and λ_n and ρ_n are typically called spatial autoregressive parameters. The analysis allows for $\mathbf{W}_n = \mathbf{M}_n$, which will frequently be the case in applications. All quantities are allowed to depend on the sample size.

The vectors $\overline{\mathbf{y}}_n = \mathbf{W}_n \mathbf{y}_n$ and $\overline{\mathbf{u}}_n = \mathbf{M}_n \mathbf{u}_n$ are typically referred to as spatial lags of $\mathbf{y}_n$ and $\mathbf{u}_n$, respectively. We note that all quantities are allowed to depend on the sample size and so some of the exogenous regressors may be spatial lags of exogenous variables. Thus the model is fairly general in that it allows for spatial spillovers in the endogenous variables, exogenous variables and disturbances.

The spatial weights matrices and the autoregressive parameters are assumed to satisfy the following assumption.

Assumption 1 *(a) All diagonal elements of $\mathbf{W}_n$ and $\mathbf{M}_n$ are zero. (b) $\lambda_n \in (-\underline{a}_n^\lambda, \overline{a}_n^\lambda)$, $\rho_n \in (-\underline{a}_n^\rho, \overline{a}_n^\rho)$ with $0 < \underline{a}_n^\lambda, \overline{a}_n^\lambda \le a^\lambda < \infty$ and $0 < \underline{a}_n^\rho, \overline{a}_n^\rho \le a^\rho < \infty$. (c) The matrices $\mathbf{I}_n - \lambda \mathbf{W}_n$ and $\mathbf{I}_n - \rho \mathbf{M}_n$ are nonsingular for all $\lambda \in (-\underline{a}_n^\lambda, \overline{a}_n^\lambda)$, and $\rho \in (-\underline{a}_n^\rho, \overline{a}_n^\rho)$.*

Assumption 1(a) is clearly a normalization rule. Assumption 1(b) concerning the parameter space of λ_n and ρ_n will be discussed in the next subsection. Assumption 1(c) ensures that $\mathbf{y}_n$ and $\mathbf{u}_n$ are uniquely defined by (1) and (2) as

$$\mathbf{y}_n = (\mathbf{I}_n - \lambda_n \mathbf{W}_n)^{-1} \mathbf{X}_n \beta_n + (\mathbf{I}_n - \lambda_n \mathbf{W}_n)^{-1} \mathbf{u}_n, \qquad (3)$$
$$\mathbf{u}_n = (\mathbf{I}_n - \rho_n \mathbf{M}_n)^{-1} \boldsymbol{\varepsilon}_n.$$

As remarked in the Introduction, spatial units are often heterogeneous in important characteristics, e.g., size. For that reason it is important to develop an estimation theory that allows for the innovations to be heteroskedastic. Therefore, we maintain the following set of assumptions with respect to the innovations.

Assumption 2 *The innovations $\{\varepsilon_{i,n} : 1 \le i \le n, \; n \ge 1\}$ satisfy $E\varepsilon_{i,n} = 0$, $E(\varepsilon_{i,n}^2) = \sigma_{i,n}^2$ with $0 < \underline{a}^\sigma \le \sigma_{i,n}^2 \le \overline{a}^\sigma < \infty$, and $\sup_{1 \le i \le n, n \ge 1} E |\varepsilon_{i,n}|^{4+\eta} < \infty$ for some $\eta > 0$. Furthermore, for each $n \ge 1$ the random variables $\varepsilon_{1,n}, \ldots, \varepsilon_{n,n}$ are totally independent.*

The above assumption also allows for the innovations to depend on the sample size n, i.e., to form triangular arrays. We note that even if the innovations do not depend on n, the elements of $\mathbf{y}_n$ and $\mathbf{u}_n$ would still depend on n in light of (3) since the elements of the inverse matrices involved would generally depend on n. We maintain the following assumption concerning the spatial weights matrices.

Assumption 3 *The row and column sums of the matrices* $\mathbf{W}_n$, $\mathbf{M}_n$, $(\mathbf{I}_n - \lambda_n \mathbf{W}_n)^{-1}$ *and* $(\mathbf{I}_n - \rho_n \mathbf{M}_n)^{-1}$ *are bounded uniformly in absolute value.*

Given (3), Assumption 2 implies that $E\mathbf{u}_n = 0$, and that the VC matrix of $\mathbf{u}_n$ is given by

$$E\mathbf{u}_n \mathbf{u}_n' = (\mathbf{I}_n - \rho_n \mathbf{M}_n)^{-1} \mathbf{\Sigma}_n (\mathbf{I}_n - \rho_n \mathbf{M}_n)^{-1}$$

where $\mathbf{\Sigma}_n = diag(\sigma_{i,n}^2)$. This specification allows for fairly general patterns of autocorrelation and heteroskedasticity of the disturbances. It is readily seen that the row and column sums of products of matrices, whose row and column sums are bounded uniformly in absolute value, are again uniformly bounded in absolute value; see, e.g., Kelejian and Prucha (2004), Remark A.1. Because of this, Assumptions 2 and 3 imply that the row and column sums of the variance-covariance (VC) matrix of $\mathbf{u}_n$ (and similarly those of $\mathbf{y}_n$) are uniformly bounded in absolute value, thus limiting the degree of correlation between, respectively, the elements of $\mathbf{u}_n$ (and of $\mathbf{y}_n$). That is, making an analogy to the time series literature, these assumptions ensure that the disturbance process and the process for the dependent variable exhibit a "fading" memory.[5]

2.2   Parameter Space for an Autoregressive Parameter

Assumption 1(b) defines the parameter space for the autoregressive parameters. In discussing this assumption we focus on $\mathbf{W}_n$ and λ_n. (An analogous discussion applies to $\mathbf{M}_n$ and ρ_n.) In the existing literature relating to Cliff-Ord models the parameter space for the autoregressive parameter is typically taken to be the interval $(-1, 1)$ and the autoregressive parameter is assumed not to depend on the sample size. However, in applications it is typically found that for unnormalized spatial weights matrices, $\mathbf{I}_n - \lambda \mathbf{W}_n$ is singular for some values of $\lambda \in (-1, 1)$. To avoid this situation, many applied researchers normalize each row of their spatial weights matrices in such a way that $\mathbf{I}_n - \lambda \mathbf{W}_n$ is non-singular for all $\lambda \in (-1, 1)$. We now discuss the implications of various normalizations of the spatial weight matrix.

Suppose c_n denotes a scalar normalization factor. Clearly, this normalization factor may depend on the sample size. For example, some of our results below relate to the case in which c_n corresponds to the maximal row or column sum of the absolute values of the elements of $\mathbf{W}_n$. Given such a normalizing factor, an equivalent specification of model (1) for $\mathbf{y}_n$ is obtained if $\lambda_n \mathbf{W}_n$ is replaced by $\lambda_n^* \mathbf{W}_n^*$ where $\lambda_n^* = c_n \lambda_n$ and $\mathbf{W}_n^* = \mathbf{W}_n/c_n$. It is important to observe that even if λ_n and its corresponding parameter space do not depend on n, λ_n^* and its implied parameter space will depend on the sample size as a result of the normalization of the spatial weights matrix.[6] It is for this reason that we allow in Assumption 1 for the elements of the spatial weights matrices, and the

---

[5]Of course, the extent of correlation is limited in virtually all large sample analysis; see, e.g., Amemiya (1985), ch. 3,4, and Pötscher and Prucha (1997), ch. 5,6.

[6]The parameter space for λ_n^* is given by $(-c_n \underline{a}_n^\lambda, c_n \overline{a}_n^\lambda)$.

autoregressive parameters and the corresponding parameter spaces to depend on n. Of course, Assumption 1 also covers the case where the true data generating process corresponds to a model where autoregressive parameters do not depend on n.

Assumption 1 defines the parameter space for λ_n as an interval around zero such that $\mathbf{I}_n - \lambda \mathbf{W}_n$ is nonsingular for values λ in that interval. The following trivial lemma gives bounds for that interval.

Lemma 1 *Let τ_n denote the spectral radius of $\mathbf{W}_n$; i.e.,*

$$\tau_n = \max\{|\nu_{1,n}|, \ldots, |\nu_{n,n}|\}$$

where $\nu_{1,n}, \ldots, \nu_{n,n}$ denote the eigenvalues of $\mathbf{W}_n$. Then $\mathbf{I}_n - \lambda \mathbf{W}_n$ is nonsingular for all values of λ in the interval $(-1/\tau_n, 1/\tau_n)$.[7]

Clearly, if we select $(-1/\tau_n, 1/\tau_n)$ as the parameter space for λ_n, then all eigenvalues of $\lambda_n \mathbf{W}_n$ are less than one in absolute value. Thus if we interpret (1) as an equilibrium relationship, then this choice of the parameter space rules out unstable Nash equilibria. Of course, we obtain an equivalent specification of the model if instead of working with $\mathbf{W}_n$ we work with the normalized weights matrix $\mathbf{W}_n^* = \mathbf{W}_n/\tau_n$ and select the interval $(-1, 1)$ as the parameter space for $\lambda_n^* = \lambda_n \tau_n$. Assumption 1 is sufficiently general to cover both cases.

For large sample sizes the computation of the eigenvalues on $\mathbf{W}_n$ is difficult. The following lemma gives boundaries, which are simple to compute, for a (sub)set of values of λ for which $\mathbf{I}_n - \lambda \mathbf{W}_n$ is nonsingular.

Lemma 2 *Let*

$$\tau_n^* = \min\{\max_{1 \le i \le n} \sum_{j=1}^{n} |w_{ij,n}|, \max_{1 \le j \le n} \sum_{i=1}^{n} |w_{ij,n}|\}.$$

Then $\tau_n \le \tau_n^*$ and consequently $\mathbf{I}_n - \lambda \mathbf{W}_n$ is nonsingular for all values of λ in the interval $(-1/\tau_n^*, 1/\tau_n^*)$.

The above lemma suggests $(-1/\tau_n^*, 1/\tau_n^*)$ as an alternative (although somewhat more restrictive) specification of the parameter space. Of course, we obtain an equivalent model specification if we normalize the spatial weights matrix by τ_n^* and if we choose $(-1, 1)$ as the parameter space for the autoregressive parameter. Since the spectral radius is bound by any matrix norm, other norms in place of the maximum absolute row and column sum norms can be used, but τ_n^* is especially easy to compute.

Rather than to normalize $\mathbf{W}_n$ by τ_n or τ_n^*, in much of the empirical literature the spatial weights matrices are normalized such that each row sums to unity.

---

[7]In some of the spatial literature the following closely related claim can be found: $\mathbf{I}_n - \lambda \mathbf{W}_n$ is nonsingular for all values of λ in the interval $(1/\nu_{n,\min}, 1/\nu_{n,\max})$, where $\nu_{n,\min}$ and $\nu_{n,\max}$ denote the smallest and largest eigenvalue of $\mathbf{W}_n$, respectively. This claim is correct for the case in which all eigenvalues of $\mathbf{W}_n$ are real and $\nu_{n,\min} < 0$ and $\nu_{n,\max} > 0$. Since, e.g., the eigenvalues of $\mathbf{W}_n$ need not be real, this claim does not hold in general.

The motivation for this normalization is that if $\mathbf{W}_n$ is row-normalized then $\mathbf{I}_n - \lambda\mathbf{W}_n$ is nonsingular for all values of λ in the interval $(-1, 1)$; this can be readily confirmed via Lemma 2. However, this normalization is quite different than those described above in that in row-normalizing a matrix one does not use a single normalization factor, but rather a different factor for the elements of each row. Therefore, in general, there exists no corresponding re-scaling factor for the autoregressive parameter that would lead to a specification that is equivalent to that corresponding to the un-normalized weights matrix. Consequently, unless theoretical issues suggest a row-normalized weights matrix, this approach will in general lead to a misspecified model.

The above discussion provides the motivation for our specification that the autoregressive parameters may depend on n. Furthermore, since some of the regressors may be spatial lags, we allow all of the model parameters to depend on the sample size.

3 GM Estimator for the Autoregressive Parameter ρ_n

In the following we introduce a class of GM estimators for ρ_n that can be easily computed, and prove their consistency and asymptotic normality under a set of general assumptions. We note that the discussion in this section only maintains model (2) for the disturbances $\mathbf{u}_n$, but not necessarily (1) for $\mathbf{y}_n$. Thus the results will also be useful in other settings such as cases where $\mathbf{y}_n$ is determined by a nonlinear model; see, e.g. Kelejian and Prucha (2001, p. 228). The estimators put forth below generalize the GM estimator for ρ_n introduced in Kelejian and Prucha (1999). In contrast to that earlier paper we now allow for heteroskedastic innovations $\varepsilon_{i,n}$, and optimal weighting of the moment conditions. We also do not confine the parameter space for ρ_n to be the interval $(-1, 1)$, and allow ρ_n to depend on n. In our earlier paper we only demonstrated the consistency of the estimator. In the following we also derive the asymptotic distribution of the considered estimators.

3.1 Definition of the GM Estimator for ρ_n

In the following let $\widetilde{\mathbf{u}}_n$ denote some predictor of $\mathbf{u}_n$. Furthermore, for notational convenience let $\overline{\mathbf{u}}_n = \mathbf{M}_n\mathbf{u}_n$ and $\overline{\overline{\mathbf{u}}}_n = \mathbf{M}_n\overline{\mathbf{u}}_n = \mathbf{M}_n^2\mathbf{u}_n$, and correspondingly, $\widetilde{\overline{\mathbf{u}}}_n = \mathbf{M}_n\widetilde{\mathbf{u}}_n$, and $\widetilde{\overline{\overline{\mathbf{u}}}}_n = \mathbf{M}_n^2\widetilde{\mathbf{u}}_n$. Similarly, let $\overline{\boldsymbol{\varepsilon}}_n = \mathbf{M}_n\boldsymbol{\varepsilon}_n$. It is readily seen that under Assumptions 1 and 2 we have the following moment conditions:

$$
\begin{aligned}
n^{-1}E\overline{\boldsymbol{\varepsilon}}_n'\overline{\boldsymbol{\varepsilon}}_n &= n^{-1}Tr\left\{\mathbf{M}_n\left[diag_{i=1}^n(E\varepsilon_{i,n}^2)\right]\mathbf{M}_n'\right\}, \qquad (4) \\
n^{-1}E\overline{\boldsymbol{\varepsilon}}_n'\boldsymbol{\varepsilon}_n &= 0.
\end{aligned}
$$

It proves convenient to rewrite these conditions as

$$
n^{-1}E\left[\begin{array}{c} \boldsymbol{\varepsilon}_n'\mathbf{A}_{1,n}\boldsymbol{\varepsilon}_n \\ \boldsymbol{\varepsilon}_n'\mathbf{A}_{2,n}\boldsymbol{\varepsilon}_n \end{array}\right] = 0 \qquad (5)
$$

with
$$\mathbf{A}_{1,n} = \mathbf{M}'_n\mathbf{M}_n - diag_{i=1}^n(\mathbf{m}'_{.i,n}\mathbf{m}_{.i,n}), \qquad \mathbf{A}_{2,n} = \mathbf{M}_n.$$

Under Assumptions 1 and 3 it is readily seen that the diagonal elements of $\mathbf{A}_{1,n}$ and $\mathbf{A}_{2,n}$ are zero and that the row and column sums of $\mathbf{A}_{1,n}$ and $\mathbf{A}_{2,n}$ are bounded uniformly in absolute value; see, e.g., Remark A.1 in Kelejian and Prucha (2004).

Our GM estimators for ρ_n are based on these moments. Specifically, note that in light of (2) $\boldsymbol{\varepsilon}_n = (\mathbf{I}_n - \rho_n\mathbf{M}_n)\mathbf{u}_n = \mathbf{u}_n - \rho_n\overline{\mathbf{u}}_n$ and so $\overline{\boldsymbol{\varepsilon}}_n = \overline{\mathbf{u}}_n - \rho_n\overline{\overline{\mathbf{u}}}_n$. Substituting these expressions into (4) or (5) yields the following two equation system:

$$\boldsymbol{\gamma}_n - \boldsymbol{\Gamma}_n\boldsymbol{\alpha}_n = 0 \tag{6}$$

where $\boldsymbol{\alpha}_n = [\rho_n, \rho_n^2]'$ and the elements of $\boldsymbol{\Gamma}_n = \left[\gamma_{rs,n}\right]_{r,s=1,2}$ and $\boldsymbol{\gamma}_n = \left[\gamma_{1,n}, \gamma_{2,n}\right]'$ are given by

$$
\begin{aligned}
\gamma_{11,n} &= 2n^{-1}E\left\{\overline{\overline{\mathbf{u}}}'_n\overline{\mathbf{u}}_n - Tr\left[\mathbf{M}_n\left[diag_{i=1}^n(\overline{u}_{i,n}u_{i,n})\right]\mathbf{M}'_n\right]\right\} \tag{7}\\
&= 2n^{-1}E\mathbf{u}'_n\mathbf{M}'_n\mathbf{A}_{1,n}\mathbf{u}_n,\\
\gamma_{12,n} &= -n^{-1}E\left\{\overline{\overline{\mathbf{u}}}'_n\overline{\overline{\mathbf{u}}}_n + Tr\left[\mathbf{M}_n\left[diag_{i=1}^n(\overline{u}_{i,n}^2)\right]\mathbf{M}'_n\right]\right\}\\
&= -n^{-1}E\mathbf{u}'_n\mathbf{M}'_n\mathbf{A}_{1,n}\mathbf{M}_n\mathbf{u}_n,\\
\gamma_{21,n} &= n^{-1}E(\mathbf{u}'_n\overline{\overline{\mathbf{u}}}_n + \overline{\mathbf{u}}'_n\overline{\mathbf{u}}_n)\\
&= n^{-1}E\mathbf{u}'_n\mathbf{M}'_n(\mathbf{A}_{2,n} + \mathbf{A}'_{2,n})\mathbf{u}_n,\\
\gamma_{22,n} &= -n^{-1}E\overline{\mathbf{u}}'_n\overline{\overline{\mathbf{u}}}_n\\
&= -n^{-1}E\mathbf{u}'_n\mathbf{M}'_n\mathbf{A}_{2,n}\mathbf{M}_n\mathbf{u}_n,\\
\gamma_{1,n} &= n^{-1}E\left\{\overline{\mathbf{u}}'_n\overline{\mathbf{u}}_n - Tr\left[\mathbf{M}_n\left[diag_{i=1}^n(u_{i,n}^2)\right]\mathbf{M}'_n\right]\right\}\\
&= n^{-1}E\mathbf{u}'_n\mathbf{A}_{1,n}\mathbf{u}_n,\\
\gamma_{2,n} &= n^{-1}E\mathbf{u}'_n\overline{\mathbf{u}}_n\\
&= n^{-1}E\mathbf{u}'_n\mathbf{A}_{2,n}\mathbf{u}_n.
\end{aligned}
$$

Now let $\widetilde{\boldsymbol{\Gamma}}_n = \left[\widetilde{\gamma}_{rs,n}\right]_{r,s=1,2}$ and $\widetilde{\boldsymbol{\gamma}}_n = \left[\widetilde{\gamma}_{1,n}, \widetilde{\gamma}_{2,n}\right]'$ denote corresponding estimators for the elements of $\boldsymbol{\Gamma}_n$ and $\boldsymbol{\gamma}_n$, which are obtained from the above expressions for the elements of $\boldsymbol{\Gamma}_n$ and $\boldsymbol{\gamma}_n$ by suppressing the expectations operator, and replacing the disturbances $\mathbf{u}_n$, $\overline{\mathbf{u}}_n$, and $\overline{\overline{\mathbf{u}}}_n$ by their predictors $\widetilde{\mathbf{u}}_n$, $\widetilde{\overline{\mathbf{u}}}_n$, and $\widetilde{\overline{\overline{\mathbf{u}}}}_n$, respectively. Then, the empirical analog of the relationship in (6) is

$$\widetilde{\boldsymbol{\gamma}}_n - \widetilde{\boldsymbol{\Gamma}}_n\boldsymbol{\alpha}_n = \boldsymbol{v}_n, \tag{8}$$

where $\boldsymbol{v}_n$ can be viewed as a vector of regression residuals. Our GM estimators of ρ_n, say $\widetilde{\rho}_n$, are now defined as weighted nonlinear least squares estimators based on (8). That is, let $\widetilde{\boldsymbol{\Upsilon}}_n$ be a 2×2 symmetric positive semidefinite (moments) weighting matrix; then $\widetilde{\rho}_n$ is defined as

$$\widetilde{\rho}_n = \widetilde{\rho}_n(\widetilde{\Upsilon}_n) = \underset{\rho \in [-a^\rho, a^\rho]}{argmin} \left\{ \left[\widetilde{\boldsymbol{\gamma}}_n - \widetilde{\boldsymbol{\Gamma}}_n \left[\begin{array}{c} \rho \\ \rho^2 \end{array} \right] \right]' \widetilde{\Upsilon}_n \left[\widetilde{\boldsymbol{\gamma}}_n - \widetilde{\boldsymbol{\Gamma}}_n \left[\begin{array}{c} \rho \\ \rho^2 \end{array} \right] \right] \right\}. \quad (9)$$

We note that the objective function for $\widetilde{\rho}_n$ remains well defined even for values of ρ_n for which $\mathbf{I}_n - \rho_n \mathbf{M}_n$ is singular, which allows us to take the optimization space for $\widetilde{\rho}_n$ to be any compact interval that contains the true parameter space. For computational efficiency it is best to use the formulae for the elements of $\widetilde{\boldsymbol{\Gamma}}_n$ and $\widetilde{\boldsymbol{\gamma}}_n$ corresponding to the first expression on the r.h.s. of (7) and to compute $\widetilde{\overline{\mathbf{u}}}_n$ and $\widetilde{\overline{\overline{\mathbf{u}}}}_n$ recursively as $\widetilde{\overline{\mathbf{u}}}_n = \mathbf{M}_n \widetilde{\mathbf{u}}_n$ and $\widetilde{\overline{\overline{\mathbf{u}}}}_n = \mathbf{M}_n \widetilde{\overline{\mathbf{u}}}_n$. In this fashion one can avoid the computation of $\mathbf{M}_n^2$, i.e., the computation of the product of two $n \times n$ matrices.

We now relate the above estimator for the autoregressive parameter to the GM estimator introduced in Kelejian and Prucha (1999). Under homoskedasticity, $\sigma^2 = \sigma_{i,n}^2$ and so $E[n^{-1} \boldsymbol{\varepsilon}_n' \boldsymbol{\varepsilon}_n] = \sigma^2$, $E[n^{-1} \overline{\boldsymbol{\varepsilon}}_n' \overline{\boldsymbol{\varepsilon}}_n] = \sigma^2 n^{-1} Tr \{\mathbf{M}_n \mathbf{M}_n'\}$, and $E[n^{-1} \overline{\boldsymbol{\varepsilon}}_n' \boldsymbol{\varepsilon}_n] = 0$. These three moment conditions underlie the GM estimator suggested in Kelejian and Prucha (1999). Substituting the first of these moment conditions into the second yields

$$E[\frac{1}{n} \overline{\boldsymbol{\varepsilon}}_n' \overline{\boldsymbol{\varepsilon}}_n] = E[n^{-1} \boldsymbol{\varepsilon}_n' \boldsymbol{\varepsilon}_n] n^{-1} Tr \{\mathbf{M}_n \mathbf{M}_n'\}, \quad (10)$$

$$E[\frac{1}{n} \overline{\boldsymbol{\varepsilon}}_n' \boldsymbol{\varepsilon}_n] = 0,$$

which is clearly a special case of (4) under homoskedasticity. It is not difficult to see that the GM estimator suggested in our previous paper can be viewed as being based on the two moment conditions in (10) with $\widetilde{\Upsilon}_n = diag(v_n, 1)$ and $v_n = 1/\left[1 + \left[n^{-1} Tr \{\mathbf{M}_n \mathbf{M}_n'\} \right]^2 \right]$. [8]

3.2   Consistency of the GM Estimator for ρ_n

To establish consistency of $\widetilde{\rho}_n$ we postulate the following additional assumptions.

Assumption 4 : *Let $\widetilde{u}_{i,n}$ denote the i-th element of $\widetilde{\mathbf{u}}_n$. We then assume that*

$$\widetilde{u}_{i,n} - u_{i,n} = \mathbf{d}_{i.,n} \boldsymbol{\Delta}_n$$

where $\mathbf{d}_{i.,n}$ and $\boldsymbol{\Delta}_n$ are $1 \times p$ and $p \times 1$ dimensional random vectors. Let $d_{ij,n}$ be the j-th element of $\mathbf{d}_{i.,n}$. Then, we assume that for some $\delta > 0$ $E|d_{ij,n}|^{2+\delta} \leq c_d < \infty$ where c_d does not depend on n, and that $n^{1/2} \|\boldsymbol{\Delta}_n\| = O_p(1)$.

---

[8] If we rewrite the moment conditions in (10) in the form corresponding to (5), then $\mathbf{A}_{1,n} = \mathbf{M}_n' \mathbf{M}_n - n^{-1} Tr (\mathbf{M}_n \mathbf{M}_n') \mathbf{I}_n$ and $\mathbf{A}_{2,n} = \mathbf{M}_n$.

Assumption 5 *(a)The smallest eigenvalue of $\boldsymbol{\Gamma}_n'\boldsymbol{\Gamma}_n$ is uniformly bounded away from zero.[9] (b) $\widetilde{\Upsilon}_n - \Upsilon_n = o_p(1)$, where Υ_n are 2×2 non-stochastic symmetric positive definite matrices. (c) The largest eigenvalues of Υ_n are bounded uniformly from above, and the smallest eigenvalues of Υ_n are uniformly bounded away from zero.*

Assumption 4 implies $n^{-1} \sum_{i=1}^n \|\mathbf{d}_{i,n}\|^{2+\delta} = O_p(1)$, which was maintained in Kelejian and Prucha (1999), and so is slightly stronger than their assumption. Assumption 4 should be satisfied for typical linear spatial models where $\widetilde{u}_i$ is based on $n^{1/2}$-consistent estimators of regression coefficients, $\mathbf{d}_{i,n}$ denotes the i-th row of the regressor matrix, and $\boldsymbol{\Delta}_n$ denotes the difference between the parameter estimator and the true parameter values. In the next section we will actually demonstrate that Assumption 4 holds for the estimated residuals of model (1) based on an instrumental variable procedure. Assumption 4 should also be satisfied for typical non-linear models provided the response function is differentiable in the parameters, and the derivatives are (uniformly over the parameter space) bounded by some random variable with bounded $2+\delta$ moments; compare Kelejian and Prucha (1999).

Assumption 5 ensures that the smallest eigenvalue of $\boldsymbol{\Gamma}_n'\Upsilon_n\boldsymbol{\Gamma}_n$ is uniformly bounded away from zero and will be sufficient to permit us to demonstrate that ρ_n is identifiably unique w.r.t. the nonstochastic analogue of the objective function of the GM estimator. This analogue is given by the function in curly brackets on the r.h.s. of (9) with $\widetilde{\boldsymbol{\gamma}}_n$, $\widetilde{\boldsymbol{\Gamma}}_n$ and $\widetilde{\Upsilon}_n$ replaced by $\boldsymbol{\gamma}_n$, $\boldsymbol{\Gamma}_n$ and Υ_n. Under homoskedasticity and $\widetilde{\Upsilon}_n = \Upsilon_n$ specified as at the end of the previous subsection this assumption is in essence equivalent to Assumption 5 in Kelejian and Prucha (1999).

Clearly Assumption 5 requires $\boldsymbol{\Gamma}_n$ to be nonsingular, or equivalently that $[tr\,(\mathbf{M}_n'\mathbf{A}_{1,n}\mathbf{M}_n\mathbf{S}_{u,n})\,, tr\,(\mathbf{M}_n'\mathbf{A}_{2,n}\mathbf{M}_n\mathbf{S}_{u,n})]'$ is linearly independent of $\left[tr\,\left(\mathbf{M}_n'(\mathbf{A}_{1,n} + \mathbf{A}_{1,n}')\mathbf{S}_{u,n}\right)\,, tr\,\left(\mathbf{M}_n'(\mathbf{A}_{2,n} + \mathbf{A}_{2,n}')\mathbf{S}_{u,n}\right)\right]'$, which is readily seen by observing that $E\mathbf{u}_n'\mathbf{M}_n'\mathbf{A}_{i,n}\mathbf{M}_n\mathbf{u}_n = tr\,(\mathbf{M}_n'\mathbf{A}_{i,n}\mathbf{M}_n\mathbf{S}_{u,n})$ and $n^{-1}E\mathbf{u}_n'\mathbf{M}_n'(\mathbf{A}_{i,n}+\mathbf{A}_{i,n}')\mathbf{u}_n = tr\,\left(\mathbf{M}_n'(\mathbf{A}_{i,n} + \mathbf{A}_{i,n}')\mathbf{S}_{u,n}\right)$ where $\mathbf{S}_{u,n} = (\mathbf{I}_n - \rho_n\mathbf{M}_n)^{-1}\,\boldsymbol{\Sigma}_n\,(\mathbf{I}_n - \rho_n\mathbf{M}_n)^{-1}$. It is not difficult to see that this linear independence condition is an analogue to identification conditions postulated in Lee (2007), Assumption 5(b), relating to quadratic forms under homoskedasticity.

We note that while Assumption 5 should be satisfied in many settings, it does not cover situations where all elements of the spatial weights matrix converge to zero uniformly as $n \rightarrow \infty$ - see Lee (2004) - since in this case the elements of $\boldsymbol{\Gamma}_n$ would tend to zero. On the other hand, Assumption 5 does not generally

---

[9]That is, $\lambda_{\min}(\boldsymbol{\Gamma}_n'\boldsymbol{\Gamma}_n) \geq \lambda_\gamma > 0$ where λ_γ does not depend on n. More specifically, in general Γ_n depends on M_n, ρ_n, $\sigma_{1,n}^2, \ldots \sigma_{n,n}^2$. Denoting this dependence as $\Gamma_n = \Gamma_n(M_n, \rho_n, \sigma_{1,n}^2, \ldots \sigma_{n,n}^2)$ the assumption should be understood as to postulate that

$$\inf_n \lambda_{\min}\left[\Gamma_n(M_n, \rho_n, \sigma_{1,n}^2, \ldots \sigma_{n,n}^2)'\Gamma_n(M_n, \rho_n, \sigma_{1,n}^2, \ldots \sigma_{n,n}^2)\right] > 0.$$

In this sense the assumption allows for λ_γ to depend on the sequence of spatial weights matrices M_n, and on the true values of the autoregressive parameters ρ_n and variances.

rule out settings where $\mathbf{M}_n$ is row normalized, there is an increasing number of nonzero elements in each row, and the row sums of the absolute values of the non-normalized elements are uniformly bounded.

The vector of derivatives (multiplied by minus one) of the moment conditions (6) w.r.t. ρ_n is given by $\mathbf{J}_n = \mathbf{\Gamma}_n[1, 2\rho_n]'$. As expected, the limiting distribution of the GM estimator $\widetilde{\rho}_n$ will be seen to depend on the inverse of $\mathbf{J}_n' \mathbf{\Upsilon}_n \mathbf{J}_n$. Assumption 5 also ensure that $\mathbf{J}_n' \mathbf{\Upsilon}_n \mathbf{J}_n$ is nonsingular.

Because of the equivalence of matrix norms it follows from Assumption 5 that the elements of Υ_n and Υ_n^{-1} are $O(1)$.

We can now give our basic consistency result for $\widetilde{\rho}_n$.

Theorem 1 *Let $\widetilde{\rho}_n = \widetilde{\rho}_n(\widetilde{\Upsilon}_n)$ denote the GM estimator defined by (9). Then, provided the optimization space contains the parameter space, and given Assumptions 1-5,*

$$\widetilde{\rho}_n - \rho_n \xrightarrow{p} 0 \ as \ n \to \infty.$$

Clearly the conditions of the theorem regarding $\widetilde{\Upsilon}_n$ and Υ_n are satisfied for $\widetilde{\Upsilon}_n = \Upsilon_n = \mathbf{I}_2$. In this case the estimator reduces to the nonlinear least squares estimator based on (8). This estimator can, e.g., be used to obtain initial consistent estimates of the autoregressive parameter. Choices of $\widetilde{\Upsilon}_n$ that lead to an efficient GM estimator for ρ_n (but require some consistent initial estimate of ρ_n) will be discussed below in conjunction with the asymptotic normality result.

3.3 Asymptotic Distribution of the GM Estimator for ρ_n

Let $\mathbf{D}_n = [\mathbf{d}'_{1.,n}, \ldots, \mathbf{d}'_{n.,n}]'$ where $\mathbf{d}_{i.,n}$ is defined in Assumption 4 so that $\widehat{\mathbf{u}}_n - \mathbf{u}_n = \mathbf{D}_n \boldsymbol{\Delta}_n$. To establish the asymptotic normality of $\widetilde{\rho}_n$ we need some additional assumptions.

Assumption 6 *For any $n \times n$ real matrix $\mathbf{A}_n$ whose row and column sums are bounded uniformly in absolute value*

$$n^{-1} \mathbf{D}_n' \mathbf{A}_n \mathbf{u}_n - n^{-1} E \mathbf{D}_n' \mathbf{A}_n \mathbf{u}_n = o_p(1).$$

A sufficient condition for Assumption 6 is, e.g., that the columns of $\mathbf{D}_n$ are of the form $\boldsymbol{\pi}_n + \mathbf{\Pi}_n \boldsymbol{\varepsilon}_n$, where the elements of $\boldsymbol{\pi}_n$ are bounded in absolute value and the row and column sums of $\mathbf{\Pi}_n$ are uniformly bounded in absolute value; see Lemma C.2. This will indeed be the case in many applications. In the next section we will verify that this assumption holds for the model given by (1) and (2), and where $\mathbf{D}_n$ equals the (negative of the) design matrix $\mathbf{Z}_n$.

Assumption 7 *Let $\boldsymbol{\Delta}_n$ be as defined in Assumption 4. Then*

$$n^{1/2} \boldsymbol{\Delta}_n = n^{-1/2} \mathbf{T}_n' \boldsymbol{\varepsilon}_n + o_p(1),$$

where $\mathbf{T}_n$ is a $n \times p$ dimensional real nonstochastic matrix whose elements are uniformly bounded in absolute value.

As remarked above, typically $\mathbf{\Delta}_n$ denotes the difference between the parameter estimator and the true parameter values. Assumption 7 will be satisfied by many estimators. In the next section we verify that this assumption indeed holds for the considered instrumental variable estimators for the parameters of model (1).

It may be helpful to provide some insight concerning the variance of the limiting distribution of the GM estimator $n^{1/2}(\widetilde{\rho}_n - \rho_n)$ given below. To that effect we note that an inspection of the derivation of this limiting distribution in Appendix C shows that it depends on the limiting distribution of the (properly normalized) vector of quadratic forms

$$\mathbf{v}_n = n^{-1/2} \left[\begin{array}{c} \frac{1}{2}\varepsilon_n'(\mathbf{A}_{1,n} + \mathbf{A}_{1,n}')\varepsilon_n + \mathbf{a}_{1,n}'\varepsilon_n \\ \frac{1}{2}\varepsilon_n'(\mathbf{A}_{2,n} + \mathbf{A}_{2,n}')\varepsilon_n + \mathbf{a}_{2,n}'\varepsilon_n \end{array} \right] \tag{11}$$

where for $r = 1, 2$ the $n \times n$ matrices $\mathbf{A}_{r,n}$ are defined in (5), and where the $n \times 1$ vectors $\mathbf{a}_{r,n}$ are defined as

$$\mathbf{a}_{r,n} = \mathbf{T}_n \boldsymbol{\alpha}_{r,n} \tag{12}$$

with

$$\boldsymbol{\alpha}_{r,n} = n^{-1} E\left[\mathbf{D}_n'\left(\mathbf{I}_n - \rho_n \mathbf{M}_n'\right)\left(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}'\right)\left(\mathbf{I}_n - \rho_n \mathbf{M}_n\right)\mathbf{u}_n\right].$$

From (11) and (12) we see that, in general, the limiting distribution of $n^{1/2}(\widetilde{\rho}_n - \rho_n)$ will depend on the limiting distribution of $n^{1/2}\mathbf{\Delta}_n$ via the matrix $\mathbf{T}_n$, unless $\boldsymbol{\alpha}_{r,n} = 0$. Clearly, if $\mathbf{D}_n$ is not stochastic, then $\boldsymbol{\alpha}_{r,n} = 0$. Within the the context of model (1) and with $\mathbf{D}_n$ equal to the (negative of the) design matrix $\mathbf{Z}_n$ this would be the case if the model does not contain a spatial lag of the endogenous variable.

Observing further that the diagonal elements of the matrices $\mathbf{A}_{r,n}$ are zero it follows from Lemma A.1 that the VC matrix of the vector of quadratic forms in (11) is given by $\mathbf{\Psi}_n = (\psi_{rs,n})$ where for $r, s = 1, 2$

$$\psi_{rs,n} = (2n)^{-1} tr\left[\left(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}'\right)\mathbf{\Sigma}_n\left(\mathbf{A}_{s,n} + \mathbf{A}_{s,n}'\right)\mathbf{\Sigma}_n\right] + n^{-1}\mathbf{a}_{r,n}'\mathbf{\Sigma}_n\mathbf{a}_{s,n}. \tag{13}$$

We now have the following result concerning the asymptotic distribution of $\widetilde{\rho}_n$. We note that the theorem does not assume convergence of the matrices involved.

Theorem 2 *(Asymptotic normality) Let $\widetilde{\rho}_n$ be the weighted nonlinear least squares estimators defined by (9). Then, provided the optimization space contains the parameter space, given Assumptions 1-7, and given that $\lambda_{\min}(\mathbf{\Psi}_n) \geq c_{\Psi}^* > 0$, we have*

$$n^{1/2}(\widetilde{\rho}_n - \rho_n) = (\mathbf{J}_n'\mathbf{\Upsilon}_n\mathbf{J}_n)^{-1}\mathbf{J}_n'\mathbf{\Upsilon}_n\mathbf{\Psi}_n^{1/2}\boldsymbol{\xi}_n + o_p(1) \tag{14}$$

where

$$\mathbf{J}_n = \mathbf{\Gamma}_n \left[\begin{array}{c} 1 \\ 2\rho_n \end{array} \right], \tag{15}$$

$$\boldsymbol{\xi}_n = \mathbf{\Psi}_n^{-1/2}\mathbf{v}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_2)$$

14

Furthermore $n^{1/2}(\widetilde{\rho}_n - \rho_n) = O_p(1)$ *and*

$$\mathbf{\Omega}_{\widetilde{\rho}_n}(\Upsilon_n) = (\mathbf{J}'_n \Upsilon_n \mathbf{J}_n)^{-1} \mathbf{J}'_n \Upsilon_n \mathbf{\Psi}_n \Upsilon_n \mathbf{J}_n (\mathbf{J}'_n \Upsilon_n \mathbf{J}_n)^{-1} \geq const > 0. \qquad (16)$$

The above theorem implies that the difference between the cumulative distribution function of $n^{1/2}(\widetilde{\rho}_n - \rho_n)$ and that of $N\left[0, \mathbf{\Omega}_{\widetilde{\rho}_n}\right]$ converges pointwise to zero, which justifies the use of the latter distribution as an approximation of the former.[10]

Remark 1. Clearly $\mathbf{\Omega}_{\widetilde{\rho}_n}(\mathbf{\Psi}_n^{-1}) = (\mathbf{J}'_n \mathbf{\Psi}_n^{-1} \mathbf{J}_n)^{-1}$ and $\mathbf{\Omega}_{\widetilde{\rho}_n}(\Upsilon_n) - \mathbf{\Omega}_{\widetilde{\rho}_n}(\mathbf{\Psi}_n^{-1})$ is positive semi-definite. Thus choosing $\widetilde{\Upsilon}_n$ as a consistent estimator for $\mathbf{\Psi}_n^{-1}$ leads to the efficient GM estimator. Such a consistent estimator will be developed in the next subsection. As discussed in the proof of the above theorem, the elements of $\mathbf{\Psi}_n$ are uniformly bounded in absolute value and hence $\lambda_{\max}(\mathbf{\Psi}_n) \leq c_\Psi^{**}$ for some $c_\Psi^{**} < \infty$. Since by assumption also $0 < c_\Psi^* \leq \lambda_{\min}(\mathbf{\Psi}_n)$ it follows that the conditions on the eigenvalues of Υ_n postulated in Assumption 5 are automatically satisfied by $\mathbf{\Psi}_n^{-1}$ We note that $\mathbf{\Psi}_n$ is, in general, only identical to the VC matrix of the moment vector in (5) if $\mathbf{a}_{r,n} = 0$. The terms involving $\mathbf{a}_{r,n}$ reflect the fact that the GM estimator is based on estimators of the disturbances $\mathbf{u}_n$ and not on the true disturbances. As noted above, $\mathbf{J}_n$ equals the vector of derivatives (multiplied by minus one) of the moment conditions (6) w.r.t. ρ_n, and thus $\mathbf{\Omega}_{\widetilde{\rho}_n}(\Upsilon_n)$ has the usual structure, except that here $\mathbf{\Psi}_n$ is not identical to the VC matrix of the moment vector.

Remark 2. From (11), (12), (14) and (15) we see that $n^{1/2}(\widetilde{\rho}_n - \rho_n)$ depends linearly on a vector of linear quadratic forms in the innovations $\boldsymbol{\varepsilon}_n$ plus a term of order $o_p(1)$. This result is helpful in establishing the joint distribution of $\widetilde{\rho}_n$ with that of estimators of some of the other model parameters of interest. In particular, it may be of interest to derive the joint limiting distribution of $n^{1/2}\mathbf{\Delta}_n$ and $n^{1/2}(\widetilde{\rho}_n - \rho_n)$. By Assumption 7, $n^{1/2}\mathbf{\Delta}_n$ is asymptotically linear in $\boldsymbol{\varepsilon}_n$ and hence the joint limiting distribution can be readily derived using the CLT for linear quadratic forms given in Appendix A. We will illustrate this below within the context of IV estimators for model (1).

We next introduce a consistent estimator for $\mathbf{\Omega}_{\widetilde{\rho}_n}$. For this purpose let

$$\widetilde{\mathbf{J}}_n = \widetilde{\mathbf{\Gamma}}_n \begin{bmatrix} 1 \\ 2\widetilde{\rho}_n \end{bmatrix} \qquad (17)$$

where $\widetilde{\mathbf{\Gamma}}_n$ is defined above by (7) and the discussion after that equation. We next define a HAC type estimator for $\mathbf{\Psi}_n$ whose elements are defined by (13). For this purpose let

$$\widetilde{\mathbf{\Sigma}}_n = diag_{i=1,\dots,n}(\widetilde{\varepsilon}_{i,n}^2)$$

---

[10] This follows from Corollary F4 in Pötscher and Prucha (1997). Compare also the discussion on pp. 86-87 in that reference.

with $\widetilde{\varepsilon}_n = (\mathbf{I}_n - \widetilde{\rho}_n \mathbf{M}_n) \widetilde{\mathbf{u}}_n$. We furthermore need to specify an estimator for $\mathbf{a}_{r,n} = \mathbf{T}_n \boldsymbol{\alpha}_{r,n}$. The matrix $\mathbf{T}_n$ introduced in Assumption 7 will in many applications be of the form

$$\mathbf{T}_n = \mathbf{F}_n \mathbf{P}_n \quad \text{with} \quad \mathbf{F}_n = \mathbf{H}_n \quad \text{or} \quad \mathbf{F}_n = (\mathbf{I}_n - \rho_n \mathbf{M}'_n)^{-1} \mathbf{H}_n, \quad (18)$$

where $\mathbf{H}_n$ is a real nonstochastic $n \times p_*$ matrix of instruments, and $\mathbf{P}_n$ is a real nonstochastic $p_* \times p$ matrix, with p as in Assumption 7. In the next section we will consider instrumental variable estimators for the parameters of model (1) and (2). In that section we will see that if $\boldsymbol{\Delta}_n$ corresponds to these instrumental variable estimators, then the matrix $\mathbf{T}_n$ will indeed have the above structure, and where $\mathbf{P}_n$ can be estimated consistently by some estimator $\widetilde{\mathbf{P}}_n$. We now define our estimator for $\mathbf{T}_n$ as[11]

$$\widetilde{\mathbf{T}}_n = \widetilde{\mathbf{F}}_n \widetilde{\mathbf{P}}_n \quad \text{with} \quad \widetilde{\mathbf{F}}_n = \mathbf{H}_n \quad \text{or} \quad \widetilde{\mathbf{F}}_n = (\mathbf{I}_n - \widetilde{\rho}_n \mathbf{M}'_n)^+ \mathbf{H}_n. \quad (19)$$

In light of (12) it now seems natural to estimate $\mathbf{a}_{r,n}$ by

$$\widetilde{\mathbf{a}}_{r,n} = \widetilde{\mathbf{T}}_n \widetilde{\boldsymbol{\alpha}}_{r,n}. \quad (20)$$

with

$$\widetilde{\boldsymbol{\alpha}}_{r,n} = n^{-1} \left[\mathbf{D}'_n (\mathbf{I}_n - \widetilde{\rho}_n \mathbf{M}'_n) (\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}) (\mathbf{I}_n - \widetilde{\rho}_n \mathbf{M}_n) \widetilde{\mathbf{u}}_n \right].$$

Given the above we now introduce the following HAC type estimator $\widetilde{\boldsymbol{\Psi}}_n = (\widetilde{\psi}_{rs,n})$ where for $r, s = 1, 2$

$$\widetilde{\psi}_{rs,n} = (2n)^{-1} tr \left[\left(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n} \right) \widetilde{\boldsymbol{\Sigma}}_n \left(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n} \right) \widetilde{\boldsymbol{\Sigma}}_n \right] + n^{-1} \widetilde{\mathbf{a}}'_{r,n} \widetilde{\boldsymbol{\Sigma}}_n \widetilde{\mathbf{a}}_{s,n}. \quad (21)$$

Furthermore, based on $\widetilde{\boldsymbol{\Psi}}_n$ we define the following estimator for $\boldsymbol{\Omega}_{\widetilde{\rho}_n}$:

$$\widetilde{\boldsymbol{\Omega}}_{\widetilde{\rho}_n} = (\widetilde{\mathbf{J}}'_n \widetilde{\boldsymbol{\Upsilon}}_n \widetilde{\mathbf{J}}_n)^+ \widetilde{\mathbf{J}}'_n \widetilde{\boldsymbol{\Upsilon}}_n \widetilde{\boldsymbol{\Psi}}_n \widetilde{\boldsymbol{\Upsilon}}_n \widetilde{\mathbf{J}}_n (\widetilde{\mathbf{J}}'_n \widetilde{\boldsymbol{\Upsilon}}_n \widetilde{\mathbf{J}}_n)^+. \quad (22)$$

The next theorem establishes the consistency of $\widetilde{\boldsymbol{\Psi}}_n$ and $\widetilde{\boldsymbol{\Omega}}_{\widetilde{\rho}_n}$.

Theorem 3 : *(VC matrix estimation) Suppose all of the assumptions of Theorem 2, apart from Assumption 5, hold and that additionally all of the fourth moments of the elements of $\mathbf{D}_n$ are uniformly bounded. Suppose furthermore that (a) that the elements of the nonstochastic matrices $\mathbf{H}_n$ are uniformly bounded in absolute value,(b) $\sup_n |\rho_n| < 1$ and the row and column sums of $\mathbf{M}_n$ are bounded uniformly in absolute value by, respectively, one and some finite constant (possibly after a renormalization of the weights matrix and parameter space as discussed in Section 2.2), and (c) $\widetilde{\mathbf{P}}_n - \mathbf{P}_n = o_p(1)$ with $\mathbf{P}_n = O(1)$. Then*

$$\widetilde{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}_n = o_p(1), \qquad \widetilde{\boldsymbol{\Psi}}_n^{-1} - \boldsymbol{\Psi}_n^{-1} = o_p(1).$$

---

[11] The reason for using the generalized inverse is that $\widetilde{\rho}_n$ defined by (9) is not forced to lie in the parameter space, and thus $\mathbf{I}_n - \widetilde{\rho}_n \mathbf{M}_n$ may be singular (where the probability of this event goes to zero as the sample size increases).

If furthermore Assumption 5 holds, then also

$$\widetilde{\boldsymbol{\Omega}}_{\widetilde{\rho}_n} - \boldsymbol{\Omega}_{\widetilde{\rho}_n} = o_p(1).$$

The hypothesis of zero spatial correlation in the disturbances, i.e., $H_0 : \rho_n = 0$, can now be tested in terms of $N\left[0, \widetilde{\boldsymbol{\Omega}}_{\widetilde{\rho}_n}\right]$.

Remark 3. We note that the above theorem also holds if $\widetilde{\rho}_n$ is replaced by any other estimator $\widetilde{\widetilde{\rho}}_n$ with $n^{1/2}(\widetilde{\widetilde{\rho}}_n - \rho_n) = O_p(1)$. In case $\mathbf{F}_n = \mathbf{H}_n$ condition (b) can be dropped. The consistency result for $\widetilde{\boldsymbol{\Psi}}_n^{-1}$ verifies that this estimator for $\boldsymbol{\Psi}_n^{-1}$ can indeed be used in the formulation of an efficient GM estimator, as discussed after Theorem 2.

Remark 4. The modules underlying the derivation of Theorems 2 and 3 can be readily extended to cover a wider class of estimators. A crucial underlying ingredient is the CLT for vectors of linear quadratic forms given in Appendix A, which was used to establish the limiting distribution of the vector of linear quadratic forms (11); that CLT is based on Kelejian and Prucha (2001). We emphasize that while in (5) we consider two moment conditions, all of the above results generalize trivially to the case where the GM estimator for ρ_n corresponds to m moment conditions

$$n^{-1}E[E\boldsymbol{\varepsilon}_n'\mathbf{A}_{1,n}\boldsymbol{\varepsilon}_n, \ldots, \boldsymbol{\varepsilon}_n'\mathbf{A}_{m,n}\boldsymbol{\varepsilon}_n]' = 0, \tag{23}$$

where the diagonal elements of $\mathbf{A}_{r,n}$ are zero and the row and column sums of $\mathbf{A}_{r,n}$ are bounded uniformly in absolute value. The focus of this paper is on two-step estimation procedures, which is motivated by their computational simplicity, generality of the first step (where residuals may come from nonlinear models) and, since at least under homoskedasticity, Monte Carlo experiments suggest that very little efficiency is lost, see, e.g., Das, Kelejian and Prucha (2003). Given instruments $\mathbf{H}_n$, one-step GMM estimators for all parameters, i.e., $\rho_n, \lambda_n, \beta_n$, of the SARAR$(1,1)$ model could be defined by augmenting those moment conditions (23) by the conditions

$$n^{-1}E\mathbf{H}_n'\boldsymbol{\varepsilon}_n = n^{-1}E[\mathbf{h}_{.1,n}'\boldsymbol{\varepsilon}_n, \ldots, \mathbf{h}_{.p_*,n}'\boldsymbol{\varepsilon}_n]' = 0. \tag{24}$$

with

$$\boldsymbol{\varepsilon}_n = (\mathbf{I}_n - \rho_n\mathbf{M}_n)(\mathbf{y}_n - \mathbf{X}_n\beta_n - \lambda_n\mathbf{W}_n\mathbf{y}_n).$$

The limiting distribution of the stacked moment vector follows immediately from the CLT for vectors of linear quadratic forms. Theorem 3 establishes consistent estimation of the VC matrix of the vector of (normalized) linear quadratic forms (11). Estimation of the VC matrix of the vector of (normalized) linear quadratic forms corresponding to the stacked moment conditions (23) and (24), is analogous. In fact, in this case estimation simplifies in that the

components of the vector are either quadratic or linear, and the elements of the linear terms $\mathbf{h}_{r,n}$ are observed.[12]

3.4 Joint Asymptotic Distribution of GM Estimator for ρ_n and Estimators of Other Model Parameters

In the following we discuss how the above results can be extended to obtain the joint asymptotic distribution of the GM estimator for ρ_n and of other estimators that are asymptotically linear in the innovations ε_n, i.e., that are of the form considered in Assumption 7. As remarked above, the IV estimators for the regression parameters of model (1) and (2) considered in the next section will be of this form. Based on the joint distribution it will then be possible to test joint hypothesis concerning ρ_n and other model parameters.

In the following we will give results concerning the joint asymptotic distribution of $\widetilde{\rho}_n - \rho_n$ and $\mathbf{\Delta}_n$ as considered in Assumptions 4 and 7 in conjunction with the estimation of the disturbances $\mathbf{u}_n$. Clearly, in general it will be of interest to have available results not only concerning the joint asymptotic distribution of $\widetilde{\rho}_n - \rho_n$ and $\mathbf{\Delta}_n$, but also concerning other estimators, say, $\mathbf{\Delta}_n^*$ that are of the general form considered in Assumption 7. To avoid having to introduce further notation we give our result in terms of $\widetilde{\rho}_n - \rho_n$ and $\mathbf{\Delta}_n$, but then comment on what changes would be needed in the formulae to accommodate other estimators $\mathbf{\Delta}_n^*$ in place of $\mathbf{\Delta}_n$. The discussion assumes that $\mathbf{T}_n = \mathbf{F}_n\mathbf{P}_n$ and $\widetilde{\mathbf{T}}_n = \widetilde{\mathbf{F}}_n\widetilde{\mathbf{P}}_n$ as defined in the previous subsection.

In light of the above discussion we expect that the joint limiting distribution of $n^{1/2}(\widetilde{\rho}_n - \rho_n)$ and $n^{1/2}\mathbf{\Delta}_n$ will depend on the limiting distribution of $\left[n^{-1/2}\mathbf{F}_n'\varepsilon_n, \mathbf{v}_n'\right]'$. Observing again that the diagonal elements of the matrices $\mathbf{A}_{r,n}$ are zero it follows from Lemma A.1 that the VC matrix of this vector of linear and linear quadratic form is given by

$$\mathbf{\Psi}_{\circ,n} = \left[\begin{array}{cc} \mathbf{\Psi}_{\Delta\Delta,n} & \mathbf{\Psi}_{\Delta\rho,n} \\ \mathbf{\Psi}_{\Delta\rho,n}' & \mathbf{\Psi}_n \end{array} \right] \tag{25}$$

with $\mathbf{\Psi}_{\Delta\Delta,n} = n^{-1}\mathbf{F}_n'\mathbf{\Sigma}_n\mathbf{F}_n$, $\mathbf{\Psi}_{\Delta\rho,n} = n^{-1}\mathbf{F}_n'\mathbf{\Sigma}_n\left[\mathbf{a}_{1,n}, \mathbf{a}_{2,n}\right]$ and where $\mathbf{\Psi}_n$ is defined by (13). We shall also employ the following estimator for $\mathbf{\Psi}_{\circ,n}$:

$$\widetilde{\mathbf{\Psi}}_{\circ,n} = \left[\begin{array}{cc} \widetilde{\mathbf{\Psi}}_{\Delta\Delta,n} & \widetilde{\mathbf{\Psi}}_{\Delta\rho,n} \\ \widetilde{\mathbf{\Psi}}_{\Delta\rho,n}' & \widetilde{\mathbf{\Psi}}_n \end{array} \right] \tag{26}$$

---

[12]Several months after completing this paper we became aware of a paper by Lin and Lee (2005) that considers an SAR(1) model with unknown heteroskedasticity. That paper is complementary to this one, with a focus and extensive discussion of one-step estimation procedures for that model among other things. That paper does not discuss specification issues regarding the parameter space of the autoregressive parameters, which are considered in this paper.

with $\widetilde{\boldsymbol{\Psi}}_{\Delta\Delta,n} = n^{-1}\widetilde{\mathbf{F}}'_n\widetilde{\boldsymbol{\Sigma}}_n\widetilde{\mathbf{F}}_n$, $\widetilde{\boldsymbol{\Psi}}_{\Delta\rho,n} = n^{-1}\widetilde{\mathbf{F}}'_n\widetilde{\boldsymbol{\Sigma}}_n[\widetilde{\mathbf{a}}_{1,n}, \widetilde{\mathbf{a}}_{2,n}]$ and where $\widetilde{\boldsymbol{\Psi}}_n$ is defined by (21).

We now have the following result concerning the joint limiting distribution of $n^{1/2}(\widetilde{\rho}_n - \rho_n)$ and $n^{1/2}\boldsymbol{\Delta}_n$.

Theorem 4 *Suppose all of the assumptions of Theorem 3 hold, and $\lambda_{\min}(\boldsymbol{\Psi}_{\circ,n}) \geq c^*_{\Psi_\circ} > 0$. Then*

$$\left[\begin{array}{c} n^{1/2}\boldsymbol{\Delta}_n \\ n^{1/2}(\widetilde{\rho}_n - \rho_n) \end{array}\right] = \left[\begin{array}{cc} \mathbf{P}'_n & 0 \\ 0 & (\mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1}\mathbf{J}'_n\boldsymbol{\Upsilon}_n \end{array}\right]\boldsymbol{\Psi}^{1/2}_{\circ,n}\boldsymbol{\xi}_{\circ,n} + o_p(1), (27)$$

$$\boldsymbol{\xi}_{\circ,n} = \boldsymbol{\Psi}^{-1/2}_{\circ,n}\left[n^{-1/2}\mathbf{F}'_n\boldsymbol{\varepsilon}_n, \mathbf{v}'_n\right]' \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{p*+2}).$$

Furthermore let

$$\boldsymbol{\Omega}_{\circ,n} = \left[\begin{array}{cc} \mathbf{P}'_n & 0 \\ 0 & (\mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1}\mathbf{J}'_n\boldsymbol{\Upsilon}_n \end{array}\right]\boldsymbol{\Psi}_{\circ,n}\left[\begin{array}{cc} \mathbf{P}_n & 0 \\ 0 & \boldsymbol{\Upsilon}_n\mathbf{J}_n(\mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1} \end{array}\right](28)$$

$$\widetilde{\boldsymbol{\Omega}}_{\circ,n} = \left[\begin{array}{cc} \widetilde{\mathbf{P}}'_n & 0 \\ 0 & (\widetilde{\mathbf{J}}'_n\widetilde{\boldsymbol{\Upsilon}}_n\widetilde{\mathbf{J}}_n)^+\widetilde{\mathbf{J}}'_n\widetilde{\boldsymbol{\Upsilon}}_n \end{array}\right]\widetilde{\boldsymbol{\Psi}}_{\circ,n}\left[\begin{array}{cc} \widetilde{\mathbf{P}}_n & 0 \\ 0 & \widetilde{\boldsymbol{\Upsilon}}_n\widetilde{\mathbf{J}}_n(\widetilde{\mathbf{J}}'_n\widetilde{\boldsymbol{\Upsilon}}_n\widetilde{\mathbf{J}}_n)^+ \end{array}\right] (29)$$

then $\widetilde{\boldsymbol{\Psi}}_{\circ,n} - \boldsymbol{\Psi}_{\circ,n} = o_p(1)$, $\widetilde{\boldsymbol{\Omega}}_{\circ,n} - \boldsymbol{\Omega}_{\circ,n} = o_p(1)$, and $\boldsymbol{\Psi}_{\circ,n} = O(1)$, $\boldsymbol{\Omega}_{\circ,n} = O(1)$.

The above theorem implies that the difference between the joint cumulative distribution function of $n^{1/2}[\boldsymbol{\Delta}'_n, (\widetilde{\rho}_n - \rho_n)]'$ and that of $N[0, \boldsymbol{\Omega}_{\circ,n}]$ converges pointwise to zero, which justifies the use of the latter distribution as an approximation of the former. The theorem also states that $\widetilde{\boldsymbol{\Omega}}_{\circ,n}$ is a consistent estimator for $\boldsymbol{\Omega}_{\circ,n}$.

Remark 5. The above result generalizes readily to cases where we are interested in the joint distribution between $\widetilde{\rho}_n - \rho_n$ and some other estimator, say, $\boldsymbol{\Delta}^*_n$, where $n^{1/2}\boldsymbol{\Delta}^*_n = n^{-1/2}\mathbf{T}^{*\prime}_n\boldsymbol{\varepsilon}_n + o_p(1)$, $\mathbf{T}^*_n = \mathbf{F}^*_n\mathbf{P}^*_n$ and $\widetilde{\mathbf{T}}_n = \widetilde{\mathbf{F}}^*_n\widetilde{\mathbf{P}}^*_n$, assuming that analogous assumptions are maintained for this estimator. In particular, the results remain valid, but with $\boldsymbol{\Psi}_{\Delta\Delta,n} = n^{-1}\mathbf{F}^{*\prime}_n\boldsymbol{\Sigma}_n\mathbf{F}^*_n$, $\boldsymbol{\Psi}_{\Delta\rho,n} = n^{-1}\mathbf{F}^{*\prime}_n\boldsymbol{\Sigma}_n[\mathbf{a}_{1,n}, \mathbf{a}_{2,n}]$, $\widetilde{\boldsymbol{\Psi}}_{\Delta\Delta,n} = n^{-1}\widetilde{\mathbf{F}}^{*\prime}_n\widetilde{\boldsymbol{\Sigma}}_n\widetilde{\mathbf{F}}^*_n$, $\widetilde{\boldsymbol{\Psi}}_{\Delta\rho,n} = n^{-1}\widetilde{\mathbf{F}}^{*\prime}_n\widetilde{\boldsymbol{\Sigma}}_n[\widetilde{\mathbf{a}}_{1,n}, \widetilde{\mathbf{a}}_{2,n}]$, and $\mathbf{P}_n$, $\widetilde{\mathbf{P}}_n$ replaced by $\mathbf{P}^*_n$, $\widetilde{\mathbf{P}}^*_n$.

4   Instrumental Variable Estimator for $\boldsymbol{\delta}_n$

As remarked, the consistency and asymptotic normality results developed in an important paper by Lee (2004) for the quasi-ML estimator for the SARAR(1,1) model defined by (1) and (2) under the assumption of homoskedastic innovations do not carry over to the case where the innovations are heteroskedastic. In fact, under heteroskedasticity the limiting objective function of the quasi-ML

estimator would generally not be maximized at the true parameter values, and therefore the quasi-ML estimator would be inconsistent. Also, the asymptotic normality results developed by Kelejian and Prucha (1998), Kelejian, Prucha and Yuzefovich (2004) and Lee (2003) for instrumental variable (IV) estimators of the SARAR(1,1) model do not carry over to the case where the innovations are heteroskedastic. In the following we provide results concerning the asymptotic distribution of IV estimators allowing the innovations to be heteroskedastic. More specifically, we will show that the considered IV estimators satisfy certain conditions such that their asymptotic distribution can be readily obtained via Theorem 4. We also allow for a more general definition of the parameter space of the spatial autoregressive parameters to avoid certain pitfalls discussed in Section 2.

4.1   Instruments

It is evident from (3) that in general $\mathbf{W}_n\mathbf{y}_n$ will be correlated with the disturbances $\mathbf{u}_n$, which motivates the use of IV estimation procedures. We maintain the following assumptions w.r.t. the $n \times k$ regressor matrices $\mathbf{X}_n$, and the $n \times p_*$ instrument matrices $\mathbf{H}_n$.

Assumption 8 : *The regressor matrices $\mathbf{X}_n$ have full column rank (for n large enough). Furthermore, the elements of the matrices $\mathbf{X}_n$ are uniformly bounded in absolute value.*

Assumption 9 : *The instrument matrices $\mathbf{H}_n$ have full column rank $p_* \geq k+1$ (for all n large enough). Furthermore, the elements of the matrices $\mathbf{H}_n$ are uniformly bounded in absolute value. Additionally $\mathbf{H}_n$ is assumed to, at least, contain the linearly independent columns of $(\mathbf{X}_n, \mathbf{M}_n\mathbf{X}_n)$.*

Assumption 10 : *The instruments $\mathbf{H}_n$ satisfy furthermore:*

(a) $\mathbf{Q_{HH}} = \lim_{n\to\infty} n^{-1}\mathbf{H}'_n\mathbf{H}_n$ *is finite, and nonsingular.*

(b) $\mathbf{Q_{HZ}} = p\lim_{n\to\infty} n^{-1}\mathbf{H}'_n\mathbf{Z}_n$ *and* $\mathbf{Q_{HMZ}} = p\lim_{n\to\infty} n^{-1}\mathbf{H}'_n\mathbf{M}_n\mathbf{Z}_n$ *are finite and have full column rank. Furthermore, let* $\mathbf{Q_{HZ*}}(\rho_n) = \mathbf{Q_{HZ}} - \rho_n\mathbf{Q_{HMZ}}$, *then the smallest eigenvalue of* $\mathbf{Q'_{HZ*}}(\rho_n)\mathbf{Q_{HH}^{-1}}\mathbf{Q_{HZ*}}(\rho_n)$ *is bounded away from zero uniformly in n.*

(c) $\mathbf{Q_{H\Sigma H}} = \lim_{n\to\infty} n^{-1}\mathbf{H}'_n\mathbf{\Sigma}_n\mathbf{H}_n$ *is finite and nonsingular.*

The above assumptions are similar to those maintained in Kelejian and Prucha (1998, 2004), and Lee (2003), and so a discussion which is quite similar to those given in those papers also applies here. Regarding the specification of the instruments $\mathbf{H}_n$ observe first that

$$E(\mathbf{W}_n\mathbf{y}_n) = \mathbf{W}_n\left(\mathbf{I}_n - \lambda_n\mathbf{W}_n\right)^{-1}\mathbf{X}_n\boldsymbol{\beta}_n = \sum_{i=0}^{\infty} \lambda_n^i\mathbf{W}_n^{i+1}\mathbf{X}_n\boldsymbol{\beta}_n \qquad (30)$$

provided that the characteristic roots of $\lambda_n \mathbf{W}_n$ are less than one in absolute value; compare Lemma 1 and 2 concerning the choice of the parameter space for λ_n. The instrument matrices $\mathbf{H}_n$ will be used to instrument $\mathbf{Z}_n = (\mathbf{X}_n, \mathbf{W}_n \mathbf{y}_n)$ and $\mathbf{M}_n \mathbf{Z}_n = (\mathbf{M}_n \mathbf{X}_n, \mathbf{M}_n \mathbf{W}_n \mathbf{y}_n)$ in terms of their predicted values from a least squares regression on $\mathbf{H}_n$, i.e., $\widehat{\mathbf{Z}}_n = \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_n$ and $\widehat{\mathbf{M}_n \mathbf{Z}_n} = \mathbf{P}_{\mathbf{H}_n} \mathbf{M}_n \mathbf{Z}_n$ with $\mathbf{P}_{\mathbf{H}_n} = \mathbf{H}_n (\mathbf{H}'_n \mathbf{H}_n)^{-1} \mathbf{H}'_n$. Towards approximating the ideal instruments $E(\mathbf{Z}_n) = (\mathbf{X}_n, \mathbf{W}_n E(\mathbf{y}_n))$ and $E(\mathbf{M}_n \mathbf{Z}_n) = (\mathbf{M}_n X_n, \mathbf{M}_n \mathbf{W}_n E(\mathbf{y}_n))$ it seems reasonable, in light of (30), to take $\mathbf{H}_n$ to be a subset of the linearly independent columns of

$$(\mathbf{X}_n, \mathbf{W}_n \mathbf{X}_n, \mathbf{W}_n^2 \mathbf{X}_n, \ldots, \mathbf{W}_n^q \mathbf{X}_n, \mathbf{M}_n \mathbf{X}_n, \mathbf{M}_n \mathbf{W}_n \mathbf{X}_n, , \ldots, \mathbf{M}_n \mathbf{W}_n^q \mathbf{X}_n) \quad (31)$$

where q is a pre-selected finite constant.[13] We note that if $\mathbf{H}_n$ is selected as in (31) it follows from Assumptions 3 and 8 that its elements will be bounded in absolute value as postulated in Assumption 9. Assumption 9 ensures that $\mathbf{X}_n$ and $\mathbf{M}_n \mathbf{X}_n$ are instrumented by themselves. Finally we note that the assumption that $\mathbf{H}_n$ has full column rank could be relaxed at the expense of working with generalized inverses.

4.2 Definition, Consistency and Asymptotic Normality

Towards estimating the model (1) and (2) we propose a three step procedure. In the first step the model is estimated by two stage least squares (2SLS) using the instruments $\mathbf{H}_n$. In the second step the autoregressive parameter, ρ_n, is estimated using the generalized moments estimation approach from Section 3 based on the 2SLS residuals obtained via the first step. In the third step, the regression model in (1) is re-estimated by 2SLS after transforming the model via a Cochrane-Orcutt-type transformation to account for the spatial correlation.

More specifically, the first step 2SLS estimator is defined as:

$$\widetilde{\boldsymbol{\delta}}_n = (\widehat{\mathbf{Z}}'_n \mathbf{Z}_n)^{-1} \widehat{\mathbf{Z}}'_n \mathbf{y}_n, \quad (32)$$

where $\widehat{\mathbf{Z}}_n = \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_n = (\mathbf{X}_n, \widehat{\mathbf{W}_n \mathbf{y}_n})$ and $\widehat{\mathbf{W}_n \mathbf{y}_n} = \mathbf{P}_{\mathbf{H}_n} \mathbf{W}_n \mathbf{y}_n$. In the second step we estimate ρ_n by the GM procedure defined by (9) based on the 2SLS residuals $\widetilde{\mathbf{u}}_n = \mathbf{y}_n - \mathbf{Z}_n \widetilde{\boldsymbol{\delta}}_n$. We denote the GM estimator again as $\widetilde{\rho}_n$.

The next lemma shows that various assumptions maintained in Section 3 w.r.t. the estimator of the regression parameters and estimated residuals are automatically satisfied by the 2SLS estimator $\widetilde{\boldsymbol{\delta}}_n$ and the corresponding residuals.

Lemma 3 : *Suppose Assumptions 1-3 and 8-10 hold, and $\sup_n \|\boldsymbol{\beta}_n\| < \infty$. Let $\mathbf{D}_n = -\mathbf{Z}_n$, then the fourth moments of the elements of $\mathbf{D}_n$ are uniformly*

---

[13] In Kelejian, Prucha, and Yuzefovich (2004), who considered the case of homoskedastic innovations, the instruments were determined more generally by taking q as a function of the sample size n, i.e., q_n, such that $q_n \to \infty$ as $n \to \infty$. Their Monte Carlo results suggest that $q = 2$ may be sufficient for many applications.

bounded, Assumption 6 holds, and:

(a) $n^{1/2}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n) = n^{-1/2}\mathbf{T}'_n\boldsymbol{\varepsilon}_n + o_p(1)$ *with* $\mathbf{T}_n = \mathbf{F}_n\mathbf{P}_n$ *and where*

$$\mathbf{P}_n = \mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}}[\mathbf{Q}'_{\mathbf{HZ}}\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}}]^{-1},$$

$$\mathbf{F}_n = (\mathbf{I}_n - \rho_n\mathbf{M}'_n)^{-1}\mathbf{H}_n.$$

(b) $n^{-1/2}\mathbf{T}'_n\boldsymbol{\varepsilon}_n = O_p(1)$.

(c) $\mathbf{P}_n = O_p(1)$ *and* $\widetilde{\mathbf{P}}_n - \mathbf{P}_n = o_p(1)$ *for*

$$\widetilde{\mathbf{P}}_n = (n^{-1}\mathbf{H}'_n\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}'_n\mathbf{Z}_n) \times$$
$$[(n^{-1}\mathbf{Z}'_n\mathbf{H}_n)(n^{-1}\mathbf{H}'_n\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}'_n\mathbf{Z}_n)]^{-1},$$

The condition $\sup_n \|\boldsymbol{\beta}_n\| < \infty$ is trivially satisfied if $\boldsymbol{\beta}_n = \boldsymbol{\beta}$. Of course, parts (a) and (b) together imply that $\widetilde{\boldsymbol{\delta}}_n$ is $n^{1/2}$-consistent for $\boldsymbol{\delta}_n$.

Clearly $\widetilde{\mathbf{u}}_n = \mathbf{u}_n + \mathbf{D}_n\boldsymbol{\Delta}_n$ with $\mathbf{D}_n = -\mathbf{Z}_n$ and $\boldsymbol{\Delta}_n = \widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n$. Lemma 3 shows that in essence under Assumptions 1-3 and 8-10 the 2SLS residuals automatically satisfy the conditions postulated in Assumptions 4, 6 and 7 with $\mathbf{D}_n = -\mathbf{Z}_n$, $\boldsymbol{\Delta}_n = \widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n$ and $\mathbf{T}_n$ as specified in the lemma. Consequently the results concerning consistency and asymptotic normality of the GM estimator for ρ_n in Theorems 1 and 2 apply in particular to the GM estimator $\widetilde{\rho}_n$ based on 2SLS residuals. Lemma 3 also establishes that the fourth moments of the elements of $\mathbf{D}_n = -\mathbf{Z}_n$ are uniformly bounded. The lemma also gives explicit expression for $\mathbf{P}_n$ and $\widetilde{\mathbf{P}}_n$ and verifies the conditions postulated w.r.t. to those matrices in Theorems 3 and 4. Hence the results of those two theorems also cover the GM estimator $\widetilde{\rho}_n$ and the 2SLS estimator $\widetilde{\boldsymbol{\delta}}_n$. In particular, Theorem 4 gives the joint limiting distribution of $n^{1/2}(\widetilde{\rho}_n - \rho_n)$ and $n^{1/2}\boldsymbol{\Delta}_n = n^{1/2}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n)$ where $\mathbf{D}_n = -\mathbf{Z}_n$, the matrices $\mathbf{P}_n$, $\mathbf{F}_n$, $\widetilde{\mathbf{P}}_n$ are as in Lemma 3, and $\widetilde{\mathbf{F}}_n = (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}'_n)^+ \mathbf{H}_n$.[14]

We now turn to the third step. Applying a Cochrane-Orcutt-type transformation to (1) yields

$$\mathbf{y}_{n*}(\rho_n) = \mathbf{Z}_{n*}(\rho_n)\boldsymbol{\delta}_n + \boldsymbol{\varepsilon}_n, \tag{33}$$

where $\mathbf{y}_{n*}(\rho_n) = \mathbf{y}_n - \rho_n\mathbf{M}_n\mathbf{y}_n$ and $\mathbf{Z}_{n*}(\rho_n) = \mathbf{Z}_n - \rho_n\mathbf{M}_n\mathbf{Z}_n$. Our estimator $\widehat{\boldsymbol{\delta}}_n$ for $\boldsymbol{\delta}_n$ in this third step is now defined as the 2SLS procedure applied to the transformed model (33) after replacing ρ_n by $\widetilde{\rho}_n$. That is,

$$\widehat{\boldsymbol{\delta}}_n = [\widehat{\mathbf{Z}}_{n*}(\widetilde{\rho}_n)'\mathbf{Z}_{n*}(\widetilde{\rho}_n)]^{-1}\widehat{\mathbf{Z}}_{n*}(\widetilde{\rho}_n)'\mathbf{y}_{n*}(\widetilde{\rho}_n) \tag{34}$$

where $\widehat{\mathbf{Z}}_{n*}(\widetilde{\rho}_n) = \mathbf{P}_{\mathbf{H}_n}\mathbf{Z}_{n*}(\widetilde{\rho}_n)$. To express the dependence of $\widehat{\boldsymbol{\delta}}_n$ on $\widetilde{\rho}_n$ we also write $\widehat{\boldsymbol{\delta}}_n = \widehat{\boldsymbol{\delta}}_n(\widetilde{\rho}_n)$.

---

[14] An alternative estimation approach is to use a HAC procedure to estimate the variance covariance matrix of the 2SLS estimator. Such an approach was considered in Pinske, Slade, and Brett (2002) and Kelejian and Prucha (2005). While such an approach is more robust, it does not yield a simple testing strategy for a joint test of spatial dependencies in the endogenous, exogenous and disturbances.

The next lemma shows again that various assumptions maintained in Section 3 w.r.t. the estimator of the regression parameters and estimated residuals are automatically satisfied by the generalized spatial 2SLS estimator $\widehat{\delta}_n$ and corresponding residuals.

Lemma 4 : *Suppose the assumptions of Lemma 3, and let $\widehat{\boldsymbol{\delta}}_n(\widehat{\rho}_n)$ be as defined by (34), where $\widehat{\rho}_n$ is any $n^{1/2}$-consistent estimator for ρ_n (such as the GM estimator $\widetilde{\rho}_n$ based on 2SLS residuals). Then*
(a) $n^{1/2}[\widehat{\boldsymbol{\delta}}_n(\widehat{\rho}_n) - \boldsymbol{\delta}_n] = n^{-1/2}\mathbf{T}_n^{*\prime}\boldsymbol{\varepsilon}_n + o_p(1)$ with $\mathbf{T}_n^* = \mathbf{F}_n^*\mathbf{P}_n^*$ and where

$$
\begin{aligned}
\mathbf{P}_n^* &= \mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_n)[\mathbf{Q}_{\mathbf{HZ}*}'(\rho_n)\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_n)]^{-1} \\
\mathbf{F}_n^* &= \mathbf{H}_n.
\end{aligned}
$$

(b) $n^{-1/2}\mathbf{T}_n^{*\prime}\boldsymbol{\varepsilon}_n = O_p(1)$.
(c) $\mathbf{P}_n^* = O_p(1)$ and $\widetilde{\mathbf{P}}_n^* - \mathbf{P}_n^* = o_p(1)$ for

$$
\begin{aligned}
\widetilde{\mathbf{P}}_n^* &= (n^{-1}\mathbf{H}_n'\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}_n'\mathbf{Z}_{n*}(\widehat{\rho}_n)) \times \\
&\quad \left[(n^{-1}\mathbf{Z}_{n*}'(\widehat{\rho}_n)\mathbf{H}_n)(n^{-1}\mathbf{H}_n'\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}_n'\mathbf{Z}_{n*}(\widehat{\rho}_n))\right]^{-1}.
\end{aligned}
$$

Frequently we will be interested in the joint distribution of the generalized spatial 2SLS estimator $\widehat{\boldsymbol{\delta}}_n(\widehat{\rho}_n)$ and the GM estimator $\widetilde{\rho}_n$. In light of Lemma 3 and 4 the limiting distribution of $n^{1/2}\left[(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n)', (\widetilde{\rho}_n - \rho_n)\right]'$ now follows immediately from Theorem 4 and Remark 5 after that theorem, with $\boldsymbol{\Delta}_n^* = \widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n$. The asymptotic variance covariance matrix and its corresponding estimator are given by (28) and (29) with modifications as described in Remark 5 and $\mathbf{D}_n = -\mathbf{Z}_n$. The expressions for the matrices $\mathbf{P}_n, \mathbf{F}_n, \widetilde{\mathbf{P}}_n$ are as in Lemma 3, and $\widetilde{\mathbf{F}}_n = (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n')^+ \mathbf{H}_n$. The expressions for the matrices $\mathbf{P}_n^*, \mathbf{F}_n^*, \widetilde{\mathbf{P}}_n^*$ are as in Lemma 4, and $\widetilde{\mathbf{F}}_n^* = \mathbf{H}_n$. The joint distribution can then be used to test in particular the joint hypothesis $H_0 : \lambda_n = \rho_n = 0$ in the usual manner.

Finally, consider the estimated residuals corresponding to $\widehat{\boldsymbol{\delta}}_n$, i.e., $\widetilde{\mathbf{u}}_n^* = \mathbf{y}_n - \mathbf{Z}_n\widehat{\boldsymbol{\delta}}_n = \mathbf{u}_n + \mathbf{D}_n\boldsymbol{\Delta}_n^*$. Clearly, in light of Lemma 4 we could use those residuals to define a corresponding GM estimator based on those residuals, and a discussion analogous to that after Lemma 3 would also apply applies here. Of course, further iterative procedures are possible, and their asymptotic properties would again be covered by the asymptotic theory developed above.

Remark 6. The estimation theory of Section 3 was developed under a set of fairly general assumptions. The results given in this section are illustrative as to how Assumptions 4, 6, 7 can be implied for the 2SLS and GS2SLS estimator of the SARAR(1,1) model from the more primitive Assumptions 8 - 10. The assumption that $\mathbf{X}_n$ is nonstochastic rules out the presence of endogenous regressors (apart from the spatial lag $\mathbf{W}_n\mathbf{y}_n$). Now suppose that $\mathbf{X}_n = \left[\overline{\mathbf{X}}_n, \mathbf{Y}_n\right]$ and correspondingly $\mathbf{D}_n = -\mathbf{Z}_n = -\left[\overline{\mathbf{X}}_n, \mathbf{Y}_n, \mathbf{W}_n\mathbf{y}_n\right]$, where $\overline{\mathbf{X}}_n$ satisfies Assumptions 8-10, with $\mathbf{X}_n$ replaced by $\overline{\mathbf{X}}_n$ (including in the formulation of the

instruments), and where $\mathbf{Y}_n$ is a matrix of endogenous variables. Then given the fourth moments of the elements of $\mathbf{D}_n$ are uniformly bounded, and Assumption 6 holds, parts (a),(b) and (c) of Lemmata 3 and 4 still hold but with $\widehat{\mathbf{Z}}_n = \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_n = \left[\overline{\mathbf{X}}_n, \mathbf{P}_{\mathbf{H}_n} \mathbf{Y}_n, \mathbf{P}_{\mathbf{H}_n} \mathbf{W}_n \mathbf{y}_n \right]$ and

$$
\begin{aligned}
\widehat{\mathbf{Z}}_{n*}(\widetilde{\rho}_n) &= \mathbf{P}_{\mathbf{H}_n} \mathbf{Z}_{n*}(\widetilde{\rho}_n) \\
&= \left[(\mathbf{I}_n - \widetilde{\rho}_n \mathbf{M}_n) \overline{\mathbf{X}}_n, \mathbf{P}_{\mathbf{H}_n} (\mathbf{I}_n - \widetilde{\rho}_n \mathbf{M}'_n) \mathbf{Y}_n, \mathbf{P}_{\mathbf{H}_n} (\mathbf{I}_n - \widetilde{\rho}_n \mathbf{M}_n) \mathbf{W}_n \mathbf{y}_n \right] .
\end{aligned}
$$

In specifying a full system of equations analogous to Kelejian and Prucha (2004) one could furthermore develop more primitive conditions that ensure the moment condition for the elements of $\mathbf{D}_n$ as well as Assumption 6.

5 Some Monte Carlo Results

In order to obtain some insights relating to the small sample properties of our suggested estimators, as well as the quasi-maximum likelihood estimator, we undertook a limited Monte Carlo study. Our study is limited in that we only considered one set of parameter values. On the other hand, we consider four different sample sizes, as well as a case in which the innovations are heteroskedastic, and one in which they are homoskedastic. Although our results are quite suggestive, clearly a more extensive Monte Carlo study is of interest.[15]

Our Monte Carlo study relates to the model in (1) and (2) with $\mathbf{M} = \mathbf{W}$.[16] We also took $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2]$ to be an $n \times 2$ matrix of observations on two exogenous variables. The four values of the sample size, n, we considered are $250, 500, 1000$, and 2000. In all the simulations, the values of $\mathbf{x}_1$ and $\mathbf{x}_2$ are based on data described in Kelejian and Robinson (1992) which relate to income per capita and the percent of rental housing in 1980 in 760 counties in the US mid-western states. The 760 observations on the income and rental variables were normalized to have zero mean and unit variance. We repeated the sample of 760 observations to generate samples of larger sizes. For sample sizes $n = 250$ and 500 we used the first 250 or 500 observations. In formulating the $\mathbf{W}$ matrix we took guidance from the specification considered in Kelejian and Prucha (1999). They considered a weights matrix, to which they referred to as "5 ahead and 5 behind". The reason for their designation is that the non-zero elements in the i-th row of $\mathbf{W}$ are in positions $i + 1, \ldots, i + 5, i + 3$, and $i - 1, \ldots, i - 5$, for $i = 6, \ldots, n - 5$. Thus, in these rows the i-th element of $\mathbf{u}$ is directly related to the five elements of $\mathbf{u}$ directly after it and to the five directly before. The matrix $\mathbf{W}$ is defined in a circular world so that, e.g., in the first row the non-zero elements are in positions $2, \ldots, 6$ (i.e., five ahead) and $n, \ldots, n - 5$, (i.e., five

---

[15] We note that such a study is currently being conducted by Irani Arraiz, David Drukker and the authors, but given its dimensions such a study is outside of the scope of the present paper.

[16] For ease of notation, in this section we do not indicate the possible dependence of various quantities on the sample size.

behind). The positioning of the non-zero elements in rows $2, ..., 6, n, ..., n-5$ are determined analogously. Our specification follows Kelejian and Prucha (1999), but modifies the weights matrix such that the middle third of the units only have two neighbors, one immediately ahead and one immediately behind, and so the number of neighbors d_i varies in our specification. The matrix was row normalized in that in each row the nonzero elements were set equal to $1/d_i$.

In all of our experiments the true model parameter values were: $\lambda = 0.3$, $\beta' = (1, 1)$ and $\rho = -0.8$. Let ξ_i denote independent draws from the standard normal distribution. In the homoskedastic case the innovations were generated as $\varepsilon_i = \sigma \xi_i$ with $\sigma^2 = 2$, and in the heteroskedastic case as $\varepsilon_i = \sigma_i \xi_i$ with $\sigma_i^2 = d_i/4$, which yields an average variance of approximately 2. In all cases the dependent vector $\mathbf{y}$ was generated in term of the reduced form (3). For each Monte Carlo experiment we performed 2000 Monte Carlo iterations.

Our Monte Carlo results are given in Table 1. They relate to the generalized spatial 2SLS estimator $\widehat{\boldsymbol{\delta}}(\widetilde{\rho})$ defined in (34) based on the instrument matrix $\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}]$ and with $\widetilde{\rho}$ the efficient GM estimator based on 2SLS residuals. We denote the estimator as $\widehat{\boldsymbol{\delta}}_{GS} = (\widehat{\beta}'_{GS}, \widehat{\lambda}_{GS})'$. Furthermore, we also report on the efficient GM estimator for ρ based on generalized spatial 2SLS residuals, and denote the estimator as $\widehat{\rho}_{GS}$. We also compute an estimate of the asymptotic variance covariance matrix of $(\widehat{\boldsymbol{\delta}}'_{GS}, \widehat{\rho}_{GS})'$ based on (29) using the results collected in Lemma 4, and compute rejection rates based on corresponding t-test statistics or the respective hypotheses that the parameters are equal to the true values. The test are performed at a nominal 5 percent significance, using the normal limiting distribution as an approximation for the critical values. For comparison we also report on the quasi-normal ML estimator for δ and ρ, which we denote as $\widehat{\boldsymbol{\delta}}_{ML} = (\widehat{\beta}'_{ML}, \widehat{\lambda}_{ML})'$ and $\widehat{\rho}_{ML}$. The rejection rates are computed based on t-test statistics, using (as would frequently be the case) the inverse of the negative Hessian of the log-likelihood function as the estimator for the asymptotic variance covariance matrix of the quasi ML estimator.

In Table 1 we report on the Monte Carlo estimates of the mean and standard deviations of $\widehat{\rho}_{GS}, \widehat{\lambda}_{GS}, \widehat{\beta}_{1,GS}, \widehat{\beta}_{2,GS}$ and $\widehat{\rho}_{ML}, \widehat{\lambda}_{ML}, \widehat{\beta}_{1,ML}, \widehat{\beta}_{2,ML}$. Focusing on the heteroskedastic case we see that the generalized spatial 2SLS/GM estimators shows very little bias even in small samples. In contrast, the ML estimators are significantly biased. For example, for $n = 1000$ the ML estimates for ρ and λ are -.584 and .210, while the true values are -.8 and .3, respectively. The biases do not decline with the sample size. To further explore the cause of the biases we first note that we expect the ML estimator to converge to the maximizers of the expected log-likelihood function. Hence we have also computed the maximizers of the expected log-likelihood function, which are reported in the column "Max" in Table 1. It turns out that under heteroskedasticity, for the experiment at hand, those maximizers are different from the true parameter values, and that indeed the ML estimators seem to converge to those maximizers. The rejection rates of the tests corresponding to the generalized spatial 2SLS/GM estimators are reasonably close to the nominal value of 0.05. For example, for $n = 1000$, the rejection rates range from 0.046 to 0.054. This suggests that the derived large

sample distribution and the developed estimators of the asymptotic variance covariance matrix are indeed useful towards approximating the small sample distribution. The rejection rates for the ML estimators range, as is expected due to the serious biases, from 0.176 to 1.000.

Under homoskedasticity the expected log-likelihood function is maximized at the true paramter values, and both the generalized spatial 2SLS/GM estimators and ML estimators show very little bias. We also note that the loss of efficiency of the generalized spatial 2SLS/GM estimators relative to the ML estimators seems generally modest.

Table 1: Monte Carlo Estimates

Estimator		Heteroskedasticity				Homoskedasticity			
	True	Max	Mean	Std.Dev.	Rejct.Rate	Max	Mean	Std.Dev.	Rejct.Rate
n=250									
$\widehat{\rho}_{GS}$	-0.8		-.795	.0968	.041		-.788	.0660	.070
$\widehat{\lambda}_{GS}$	0.3		.297	.0781	.053		.288	.0816	.073
$\widehat{\beta}_{1,GS}$	1.0		.999	.1039	.062		1.006	.0999	.071
$\widehat{\beta}_{2,GS}$	1.0		.999	.1046	.069		1.004	.0998	.067
$\widehat{\rho}_{ML}$	-0.8	-0.585	-.578	.0715	.859	-0.8	-.789	.0518	.046
$\widehat{\lambda}_{ML}$	0.3	0.211	.203	.0670	.294	0.3	.290	.0602	.048
$\widehat{\beta}_{1,ML}$	1.0	1.067	1.069	.0961	.144	1.0	1.006	.0902	.050
$\widehat{\beta}_{2,ML}$	1.0	1.040	1.041	.1026	.098	1.0	1.003	.0959	.058
n=500									
$\widehat{\rho}_{GS}$	-0.8		-.795	.0557	.046		-.795	.0406	.055
$\widehat{\lambda}_{GS}$	0.3		.299	.0504	.057		.296	.0506	.051
$\widehat{\beta}_{1,GS}$	1.0		.998	.0713	.061		1.001	.0724	.060
$\widehat{\beta}_{2,GS}$	1.0		1.001	.0641	.044		1.004	.0666	.049
$\widehat{\rho}_{ML}$	-0.8	-.590	-.586	.0495	.999	-0.8	-.795	0352	.058
$\widehat{\lambda}_{ML}$	0.3	.218	.216	.0425	.458	0.3	.296	.0414	.049
$\widehat{\beta}_{1,ML}$	1.0	1.073	1.072	.0647	.177	1.0	1.001	.0662	.052
$\widehat{\beta}_{2,ML}$	1.0	1.040	1.042	.0618	.092	1.0	1.004	.0643	.049
n=1000									
$\widehat{\rho}_{GS}$	-0.8		-.798	.0377	.047		-.798	.0275	.049
$\widehat{\lambda}_{GS}$	0.3		.299	.0361	.048		.298	.0365	.050
$\widehat{\beta}_{1,GS}$	1.0		1.001	.0509	.054		1.001	.0501	.054
$\widehat{\beta}_{2,GS}$	1.0		.999	.0472	.046		1.000	.0470	.043
$\widehat{\rho}_{ML}$	-0.8	-.586	-.584	.0346	1.000	-0.8	-.798	.0239	.056
$\widehat{\lambda}_{ML}$	0.3	.212	.210	0304	.819	0.3	.297	.0297	.050
$\widehat{\beta}_{1,ML}$	1.0	1.071	1.073	.0465	.348	1.0	1.002	.0457	.053
$\widehat{\beta}_{2,ML}$	1.0	1.050	1.050	.0449	.176	1.0	1.001	.0448	.037
n=2000									
$\widehat{\rho}_{GS}$	-0.8		-.799	.0267	.055		-.798	.0195	.052
$\widehat{\lambda}_{GS}$	0.3		.299	.0254	.053		.299	.0258	.056
$\widehat{\beta}_{1,GS}$	1.0		1.000	.0351	.048		1.001	.0354	.053
$\widehat{\beta}_{2,GS}$	1.0		.999	.0324	.043		1.000	.0325	.044
$\widehat{\rho}_{ML}$	-0.8	-.587	-.585	.0247	1.000	-0.8	-.798	.0171	.050
$\widehat{\lambda}_{ML}$	0.3	.214	.213	.0217	.980	0.3	.298	.0215	.055
$\widehat{\beta}_{1,ML}$	1.0	1.076	1.076	.0320	.651	1.0	1.002	.0324	.050
$\widehat{\beta}_{2,ML}$	1.0	1.047	1.046	.0306	.295	1.0	1.000	.0311	.047

6   Summary and Suggestions for Further Research

In this paper we introduce a new class of GM estimators for the autoregressive parameter of a spatially autoregressive disturbance process allowing for innovations with unknown heteroskedasticity. The estimation theory for the GM estimators is developed in a modular fashion under a fairly general set of assumptions, and should cover many (linear and nonlinear) models. The general theory is then utilized to establish the consistency and asymptotic normality of IV estimators for the regression parameters for an important class of spatial models, frequently referred to as $SARAR(1,1)$ models. The paper provides results concerning the joint asymptotic distribution of the GM estimator for the autoregressive parameter of the disturbance process and IV estimators for the model regression parameters. Among other things, the results allow for a joint test that the autoregressive parameters corresponding to the spatial lags of the dependent variable and disturbance term are both zero We also provide a discussion of the specification of the parameter space for $SARAR(1,1)$ models. We demonstrate that a computationally simple re-scaling of the weights matrix leads to an equivalent model containing a (re-scaled) autoregressive parameter which has a user-friendly parameter space. Unfortunately, previous studies in the literature have often re-scaled their weights matrix in such a way that the "before and after" scaled models are not equivalent.

One suggestion for further research is to extend the results to panel data. A further suggestion for future research would be a comprehensive Monte Carlo study which focuses on the small sample properties of the considered estimators. Such a study may also consider comparisons with the quasi-ML estimator (specified under homoskedasticity).

A   Appendix: CLT for Vectors of Linear Quadratic Forms

In the following we state, for the convenience of the reader, a central limit theorem (CLT) for vectors of linear quadratic forms with heteroskedastic innovations. This CLT is based on a CLT given in Kelejian and Prucha (2001) for the scalar case. We first state a lemma that collects useful results on the mean and VC matrix between (and as a special case the variance of) linear quadratic forms.

Lemma A.1 : *Let $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_n)'$ be a random vector with zero mean and positive definite variance covariance matrix $\boldsymbol{\Xi}$, let $A = (a_{ij})$ and $B = (b_{ij})$ be $n \times n$ nonstochastic symmetric matrices, and let $\mathbf{a}$ and $\mathbf{b}$ be $n \times 1$ nonstochastic vectors. Consider the decomposition $\boldsymbol{\Xi} = \mathbf{SS}'$, let $\mathbf{A}_* = (a_{ij,*}) = \mathbf{S}'\mathbf{AS}$ and $\mathbf{B}_* = (b_{ij,*}) = \mathbf{S}'\mathbf{BS}$, and let $\mathbf{a}_* = \mathbf{S}'\mathbf{a}$ and let $\mathbf{b}_* = \mathbf{S}'\mathbf{b}$. Furthermore let $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_n)' = \mathbf{S}^{-1}\boldsymbol{\zeta}$. Then assuming that the elements of $\boldsymbol{\eta}$ are independently distributed with zero mean, unit variance, and finite third and fourth moments $E(\eta_i^3) = \mu_{\eta_i}^{(3)}$ and $E(\eta_i^4) = \mu_{\eta_i}^{(4)}$ we have*

$$E(\boldsymbol{\zeta}'\mathbf{A}\boldsymbol{\zeta} + \mathbf{a}'\boldsymbol{\zeta}) = tr(\mathbf{A}_*) = tr(\mathbf{A}\boldsymbol{\Xi}),$$
$$cov(\boldsymbol{\zeta}'\mathbf{A}\boldsymbol{\zeta} + \mathbf{a}'\boldsymbol{\zeta}, \boldsymbol{\zeta}'\mathbf{B}\boldsymbol{\zeta} + \mathbf{b}'\boldsymbol{\zeta}) = 2tr(\mathbf{A}\boldsymbol{\Xi}\mathbf{B}\boldsymbol{\Xi}) + \mathbf{a}'\boldsymbol{\Xi}\mathbf{b}$$
$$+ \sum_{i=1}^{n} a_{ii,*}b_{ii,*}\left[\mu_{\eta_i}^{(4)} - 3\right] + \sum_{i=1}^{n}(a_{i,*}b_{ii,*} + b_{i,*}a_{ii,*})\mu_{\eta_i}^{(3)}.$$

Remark A.1. The above expression for the covariance exploits the independence of the elements of $\boldsymbol{\eta}$. A convenient way to obtain these expressions is to re-write the linear quadratic form as a sum of martingale differences as, e.g., in Kelejian and Prucha (2001, Appendix A). If the diagonal elements of $\mathbf{A}_*$ and $\mathbf{B}_*$ are zero the last two terms drop out from the expression for the covariance, and so one need only assume the existence of second moments. The last two terms also drop out from the expression for the covariance in the case where $\boldsymbol{\zeta}$ – or equivalently $\boldsymbol{\eta}$ – is normally distributed, since in this case $\mu_{\eta_i}^{(3)} = 0$ and $\mu_{\eta_i}^{(4)} = 3$. Of course, the variance of a linear quadratic form is obtained as the special case where $\mathbf{A} = \mathbf{B}$ and $\mathbf{a} = \mathbf{b}$. Obviously, in case $\mathbf{A}$ and $\mathbf{B}$ are not symmetric the above formulae apply with $\mathbf{A}$ and $\mathbf{B}$ replaced by $(\mathbf{A} + \mathbf{A}')/2$ and $(\mathbf{B} + \mathbf{B}')/2$.

Consider the linear quadratic forms $(r = 1, \ldots, m)$

$$\mathbf{Q}_{r,n} = \boldsymbol{\xi}_n'\mathbf{A}_{r,n}\boldsymbol{\xi}_n + \mathbf{a}_{r,n}'\boldsymbol{\xi}_n$$

where $\boldsymbol{\xi}_n = (\xi_{1,n}, \ldots, \xi_{n,n})'$ is an $n \times 1$ random vector, $\mathbf{A}_{r,n} = (a_{ij,r,n})_{i,j=1,\ldots,n}$ is an $n \times n$ nonstochastic real matrix, and $a_{r,n} = (a_{1,r,n}, \ldots, a_{n,r,n})'$ is an $n \times 1$ nonstochastic real vector. We maintain the following assumptions:

Assumption A.1 *The real valued random variables of the array $\{\xi_{i,n} : 1 \leq i \leq n,\ n \geq 1\}$ satisfy $E\xi_{i,n} = 0$. Furthermore, for each $n \geq 1$ the random variables $\xi_{1,n}, \ldots, \xi_{n,n}$ are totally independent.*

Assumption A.2 *For $r = 1, \ldots, m$ the elements of the array of real numbers $\{a_{ij,r,n} : 1 \leq i, j \leq n,\ n \geq 1\}$ satisfy $a_{ij,r,n} = a_{ji,r,n}$ and $\sup_{1 \leq j \leq n, n \geq 1} \sum_{i=1}^{n} |a_{ij,r,n}| < \infty$. The elements of the array of real numbers $\{a_{i,r,n} : 1 \leq i \leq n,\ n \geq 1\}$ satisfy $\sup_n n^{-1} \sum_{i=1}^{n} |a_{i,r,n}|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$.*

Assumption A.3 *For $r = 1, \ldots, m$ one of the following two conditions holds:*

(a) $\sup_{1 \leq i \leq n, n \geq 1} E \left|\xi_{i,n}\right|^{2+\eta_2} < \infty$ *for some $\eta_2 > 0$ and $a_{ii,r,n} = 0$.*

(b) $\sup_{1 \leq i \leq n, n \geq 1} E \left|\xi_{i,n}\right|^{4+\eta_2} < \infty$ *for some $\eta_2 > 0$ (but possibly $a_{ii,r,n} \neq 0$).*

Let $\mu_{\mathbf{Q}_{r,n}}$ and $\sigma_{\mathbf{Q}_{rs,n}}$ denote the mean of $\mathbf{Q}_{r,n}$ and the covariance between $\mathbf{Q}_{r,n}$ and $\mathbf{Q}_{s,n}$, respectively, for $r, s = 1, \ldots, m$. Then it follows immediately from Lemma A.1 that under Assumptions A.1 and A.3

$$\mu_{\mathbf{Q}_{r,n}} = \sum_{i=1}^{n} a_{ii,r,n} \sigma_{i,n}^2,$$

$$\sigma_{\mathbf{Q}_{rs,n}} = 2 \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij,r,n} a_{ij,s,n} \sigma_{i,n}^2 \sigma_{j,n}^2 + \sum_{i=1}^{n} a_{i,r,n} a_{i,s,n} \sigma_{i,n}^2$$

$$+ \sum_{i=1}^{n} a_{ii,r,n} a_{ii,s,n} \left[\mu_{i,n}^{(4)} - 3\sigma_{i,n}^4\right] + \sum_{i=1}^{n} (a_{i,r,n} a_{ii,s,n} + a_{i,s,n} a_{ii,r,n}) \mu_{i,n}^{(3)}$$

where $\sigma_{i,n}^2 = E\xi_{i,n}^2$, $\mu_{i,n}^{(3)} = E\xi_{i,n}^3$ and $\mu_{i,n}^{(4)} = E\xi_{i,n}^4$. In case Assumption A.3(a) holds, the mean of $\mathbf{Q}_{r,n}$ is zero and the last two terms drop out from the expression for the variance.

In the following we give a CLT for the $m \times 1$ vector of linear quadratic forms

$$\mathbf{V}_n = \begin{bmatrix} \mathbf{Q}_{1,n} & , \ldots, & \mathbf{Q}_{m,n} \end{bmatrix}'.$$

Let $\boldsymbol{\mu}_{V_n} = E\mathbf{V}_n = [\mu_{\mathbf{Q}_{1,n}}, \ldots, \mu_{\mathbf{Q}_{m,n}}]'$ and $\Sigma_{V_n} = [\sigma_{\mathbf{Q}_{rs,n}}]_{r,s=1,\ldots,m}$ denote the mean and VC matrix of $\mathbf{V}_n$, respectively. We then have the following theorem.

Theorem A.1 *Suppose Assumptions A.1-A.3 hold and $n^{-1}\lambda_{\min}(\Sigma_{V_n}) \geq c$ for some $c > 0$. Let $\Sigma_{V_n} = \left(\Sigma_{V_n}^{1/2}\right)\left(\Sigma_{V_n}^{1/2}\right)'$, then*

$$\Sigma_{V_n}^{-1/2}(\mathbf{V}_n - \boldsymbol{\mu}_{V_n}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_m).$$

Remark A.2. Since the diagonal elements of a positive definite matrix are greater than or equal to the smallest eigenvalue of that matrix it follows that $n^{-1}\lambda_{\min}(\boldsymbol{\Sigma}_{V_n}) \geq c$ implies that $n^{-1}\sigma_{\mathbf{Q}_{n,ii}} \geq c$. Therefore this assumption automatically implies the assumption maintained in Theorem 1 of Kelejian and Prucha (2001) w.r.t. the variance of a linear quadratic form for each of the linear quadratic forms $\mathbf{Q}_{r,n}$. Of course, the theorem remains valid, if all assumptions are assumed to hold for $n > n_0$ where n_0 is finite.

Proof of Theorem A.1 : Let α be some arbitrary $m \times 1$ vector with $\|\alpha\| = 1$, let

$$\boldsymbol{\pi}_n = (\pi_{1,n}, \ldots, \pi_{m,n}) = n^{1/2}\alpha'\boldsymbol{\Sigma}_{V_n}^{-1/2},$$

and define

$$\mathbf{Q}_n = n^{1/2}\alpha'\boldsymbol{\Sigma}_{V_n}^{-1/2}\mathbf{V}_n = \sum_{r=1}^{m}\pi_{r,n}\boldsymbol{\xi}_n'\mathbf{A}_{r,n}\boldsymbol{\xi}_n + \sum_{r=1}^{m}\pi_{r,n}\mathbf{a}_{r,n}'\boldsymbol{\xi}_n = \boldsymbol{\xi}_n'\mathbf{C}_n\boldsymbol{\xi}_n + \mathbf{d}_n'\boldsymbol{\xi}_n,$$

where $\mathbf{C}_n = \sum_{r=1}^{m}\pi_{r,n}\mathbf{A}_{r,n}$ and $\mathbf{d}_n = \sum_{r=1}^{m}\pi_{r,n}\mathbf{a}_{r,n}$. Observe that

$$\mu_{\mathbf{Q}_n} = E\mathbf{Q}_n = n^{1/2}\alpha'\boldsymbol{\Sigma}_{V_n}^{-1/2}\boldsymbol{\mu}_{V_n} = \sum_{r=1}^{m}\pi_{r,n}\mu_{\mathbf{Q}_{r,n}},$$

$$\sigma_{\mathbf{Q}_n}^2 = var(\mathbf{Q}_n) = n\alpha'\boldsymbol{\Sigma}_{V_n}^{-1/2}\boldsymbol{\Sigma}_{V_n}\boldsymbol{\Sigma}_{V_n}^{-1/2'}\alpha = n.$$

To prove the theorem, using the Cramer Wold device, it suffices to show that

$$\alpha'\boldsymbol{\Sigma}_{V_n}^{-1/2}(\mathbf{V}_n - \boldsymbol{\mu}_{V,n}) = \frac{\mathbf{Q}_n - \mu_{\mathbf{Q}_n}}{\sigma_{\mathbf{Q}_n}} \xrightarrow{d} N(0,1). \tag{A.1}$$

To show that (A.1) holds we verify that $\boldsymbol{\xi}_n$ and $\mathbf{C}_n$ and $\mathbf{d}_n$ satisfy the conditions of Theorem 1 in Kelejian and Prucha (2001). Assumptions A.1 and A.3 are identical to Assumptions 1 and 3 in Kelejian and Prucha (2001). Furthermore, $n^{-1}\sigma_{\mathbf{Q}_n}^2 = 1$. Thus it suffices to verify that Assumption 2 in Kelejian and Prucha (2001) holds for the elements of $\mathbf{C}_n$ and $\mathbf{d}_n$.

Clearly $\lambda_{\max}(n\boldsymbol{\Sigma}_{V_n}^{-1}) = 1/\lambda_{\min}(n^{-1}\boldsymbol{\Sigma}_{V_n}) \leq 1/c$. By Proposition 43 in Dhrymes (1978, p. 470) the matrices $n\boldsymbol{\Sigma}_{V_n}^{-1} = n\boldsymbol{\Sigma}_{V_n}^{-1/2'}\boldsymbol{\Sigma}_{V_n}^{-1/2}$ and $n\boldsymbol{\Sigma}_{V_n}^{-1/2}\boldsymbol{\Sigma}_{V_n}^{-1/2'}$ have the same characteristic roots. Thus $\|\boldsymbol{\pi}_n\|^2 = \alpha'(n\boldsymbol{\Sigma}_{V_n}^{-1/2}\boldsymbol{\Sigma}_{V_n}^{-1/2'})\alpha \leq \lambda_{\max}(n\boldsymbol{\Sigma}_{V_n}^{-1/2}\boldsymbol{\Sigma}_{V_n}^{-1/2'})\|\alpha\|^2 = \lambda_{\max}(n\boldsymbol{\Sigma}_{V_n}^{-1}) \leq 1/c < \infty$, and hence $|\pi_{r,n}| \leq c_\pi$ where $c_\pi = 1/\sqrt{c}$.

Observe that

$$|c_{ij,n}| = \sum_{r=1}^{m}|\pi_{r,n}||a_{ij,r,n}| \leq c_\pi\sum_{r=1}^{m}|a_{ij,r,n}|$$

and hence

$$\sup_{1\leq j\leq n, n\geq 1}\sum_{i=1}^{n}|c_{ij,n}| \leq c_\pi\sum_{r=1}^{m}\sup_{1\leq j\leq n, n\geq 1}\sum_{i=1}^{n}|a_{ij,r,n}| < \infty$$

31

in light of Assumption A.2. This shows that the elements of $\mathbf{C}_n$ satisfy Assumption 2 in Kelejian and Prucha (2001).

Next observe that

$$
\begin{aligned}
|d_{i,n}|^{2+\eta_1} &\leq \left[\sum_{r=1}^m |\pi_{r,n}|\,|a_{i,r,n}|\right]^{2+\eta_1} \leq c_\pi^{2+\eta_1}\left[\sum_{r=1}^m |a_{i,r,n}|\right]^{2+\eta_1} \\
&\leq c_\pi^{2+\eta_1} m^{2+\eta_1} \sum_{r=1}^m |a_{i,r,n}|^{2+\eta_1}
\end{aligned}
$$

and hence

$$
\sup_n n^{-1}\sum_{i=1}^n |d_{i,n}|^{2+\eta_1} \leq c_\pi^{2+\eta_1} m^{2+\eta_1}\sum_{r=1}^m \sup_n n^{-1}\sum_{i=1}^n |a_{i,r,n}|^{2+\eta_1} < \infty
$$

in light of Assumption A.2. Therefore the elements of $\mathbf{d}_n$ satisfy Assumption 2 in Kelejian and Prucha (2001). ∎

B   Appendix: Proofs for Section 2

Proof of Lemma 1: Clearly $\mathbf{I}_n - \lambda W_n$ is non-singular for $\lambda = 0$. For $\lambda \neq 0$ we have $\det(\mathbf{I}_n - \lambda W_n) = (-\lambda)^n \det(W_n - \lambda^{-1}\mathbf{I}_n)$. Consequently $\mathbf{I}_n - \lambda W_n$ is non-singular for values of $\lambda^{-1} \notin \{\nu_{1,n},\ldots,\nu_{n,n}\}$. In particular, $\mathbf{I}_n - \lambda\mathbf{W}_n$ is non-singular for $|\lambda^{-1}| > \tau_n$. Rewriting this last inequality as $|\lambda| < \tau_n^{-1}$ completes the proof. ∎

Proof of Lemma 2 : As an immediate consequence of Geršgorin's Theorem – see, e.g., Horn and Johnson (1985), pp. 344-346 – we have $\tau_n = \max\{|\nu_{1,n}|,\ldots,|\nu_{n,n}|\} \leq \tau_n^*$. The claim now follows from Lemma 1. ∎

C   Appendix: Proofs for Section 3

Remark C.1. Suppose the row and column sums of the $np \times np$ matrices $\mathbf{A}_n = (a_{ij,n})$ are bounded uniformly in absolute value by some finite constants c_A, then $\sum_{i=1}^{np} |a_{ij,n}|^q \leq c_A^q$ for $q > 1$ This result is trivially seen to hold since $\sum_{i=1}^{np} |a_{ij,n}|^q = c_A^{q-1}\sum_{i=1}^{np} |a_{ij,n}|\,|a_{ij,n}/c_A|^{q-1} \leq c_A^{q-1}\sum_{i=1}^{np} |a_{ij,n}| \leq c_A^q$.

We shall make use of the following lemma.

Lemma C.1 : *Suppose the row and column sums of the real nonstochastic $n \times n$ matrices $\mathbf{A}_n$ are uniformly bounded in absolute value. Let $\mathbf{u}_n$ be defined by (2) and let $\widetilde{\mathbf{u}}_n$ denote a predictor for $\mathbf{u}_n$. Suppose Assumptions 1 - 4 hold then:*

(a) $n^{-1}E\left|\mathbf{u}_n'\mathbf{A}_n\mathbf{u}_n\right| = O(1)$, $var(n^{-1}\mathbf{u}_n'\mathbf{A}_n\mathbf{u}_n) = o(1)$ *and*

$$n^{-1}\widetilde{\mathbf{u}}_n'\mathbf{A}_n\widetilde{\mathbf{u}}_n - n^{-1}E\mathbf{u}_n'\mathbf{A}_n\mathbf{u}_n = o_p(1).$$

(b) $n^{-1}E\left|\mathbf{d}_{.s,n}'\mathbf{A}_n\mathbf{u}_n\right| = O(1)$, $s = 1,\ldots,p$, *where* $\mathbf{d}_{.s,n}$ *denotes the s-th column of* $\mathbf{D}_n$, *and*

$$n^{-1}\mathbf{D}_n'\mathbf{A}_n\widetilde{\mathbf{u}}_n - n^{-1}E\mathbf{D}_n'\mathbf{A}_n\mathbf{u}_n = o_p(1).$$

(c) *If furthermore Assumption 6 holds, then*

$$n^{-1/2}\widetilde{\mathbf{u}}_n'\mathbf{A}_n\widetilde{\mathbf{u}}_n = n^{-1/2}\mathbf{u}_n'\mathbf{A}_n\mathbf{u}_n + \boldsymbol{\alpha}_n'n^{1/2}\boldsymbol{\Delta}_n + o_p(1)$$

with $\boldsymbol{\alpha}_n = n^{-1}E\mathbf{D}_n'(\mathbf{A}_n + \mathbf{A}_n')\mathbf{u}_n$. *(Of course, in light of (b) we have* $\boldsymbol{\alpha}_n = O(1)$ *and* $n^{-1}\mathbf{D}_n'(\mathbf{A}_n + \mathbf{A}_n')\widetilde{\mathbf{u}}_n - \boldsymbol{\alpha}_n = o_p(1)$.*)*

Proof. We first prove part (a) of the lemma. Let $\vartheta_n = n^{-1}\mathbf{u}_n'\mathbf{A}_n\mathbf{u}_n$ and $\widetilde{\vartheta}_n = n^{-1}\widetilde{\mathbf{u}}_n'\mathbf{A}_n\widetilde{\mathbf{u}}_n$, then given (2) we have $\vartheta_n = n^{-1}\boldsymbol{\varepsilon}_n'\mathbf{B}_n\boldsymbol{\varepsilon}_n$ with

$$\mathbf{B}_n = (1/2)\left(\mathbf{I}_n - \rho_n\mathbf{M}_n'\right)^{-1}\left(\mathbf{A}_n + \mathbf{A}_n'\right)\left(\mathbf{I}_n - \rho_n\mathbf{M}_n\right)^{-1}.$$

By Assumption 3 the row and column sums of the matrices $(\mathbf{I}_n - \rho_n\mathbf{M}_n)^{-1}$ are uniformly bounded in absolute value. Since this property is preserved under matrix addition and multiplication - see, e.g., Remark A.1 in Kelejian and Prucha (2004) - it follows that also the row and column sums of the matrices $\mathbf{B}_n$ are uniformly bounded in absolute value. Let $\boldsymbol{\Sigma}_n = diag(\sigma_{1,n}^2,\ldots,\sigma_{n,n}^2)$, then given Assumption 2 it follows further that also the row and column sums of the matrices $\mathbf{B}_n\boldsymbol{\Sigma}_n\mathbf{B}_n\boldsymbol{\Sigma}_n$ are uniformly bounded in absolute value. In the following let $K < \infty$ be a common bound for the row and column sums of the absolute elements of $\mathbf{B}_n$, $\boldsymbol{\Sigma}_n$ and $\mathbf{B}_n\boldsymbol{\Sigma}_n\mathbf{B}_n\boldsymbol{\Sigma}_n$ and of their respective elements. Then

$$
\begin{aligned}
E\left|\vartheta_n\right| &= n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij,n}|\,E\left|\varepsilon_i\right|\left|\varepsilon_j\right| \leq n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij,n}|\,\sigma_{i,n}\sigma_{j,n} \leq K^3,\\
var(\vartheta_n) &= n^{-2}2tr(\mathbf{B}_n\boldsymbol{\Sigma}_n\mathbf{B}_n\boldsymbol{\Sigma}_n) + n^{-2}\sum_{i=1}^{n}b_{ii,n}^2\left[E\varepsilon_{i,n}^4 - 3\sigma_{i,n}^4\right]\\
&\leq n^{-1}2K + n^{-1}K^2\sup_i\left[E\varepsilon_{i,n}^4 - 3\sigma_{i,n}^4\right],
\end{aligned}
$$

where we utilized Lemma A.1. Given that the fourth moments of the $\varepsilon_{i,n}$ are uniformly bounded in light of Assumption 2 this establishes the first two claims of part (a) of the lemma.

We next proof the last claim of part (a) of the lemma. The above discussion implies that $\vartheta_n - E\vartheta_n = o_p(1)$. Hence it suffices to show that $\widetilde{\vartheta}_n - \vartheta_n = o_p(1)$. By Assumption 4 we have $\widetilde{\mathbf{u}}_n - \mathbf{u}_n = \mathbf{D}_n\boldsymbol{\Delta}_n$ where $\mathbf{D}_n = [\mathbf{d}_{1.,n}',\ldots,\mathbf{d}_{n.,n}']'$. Utilizing this expression yields

$$\widetilde{\vartheta}_n - \vartheta_n = \phi_n + \psi_n$$

33

with

$$\begin{aligned}
\phi_n &= n^{-1}\boldsymbol{\Delta}_n'\mathbf{D}_n'\left(\mathbf{A}_n + \mathbf{A}_n'\right)\mathbf{u}_n = n^{-1}\boldsymbol{\Delta}_n'\mathbf{D}_n'\mathbf{C}_n\boldsymbol{\varepsilon}_n, \\
\psi_n &= n^{-1}\boldsymbol{\Delta}_n'\mathbf{D}_n'\mathbf{A}_n\mathbf{D}_n\boldsymbol{\Delta}_n,
\end{aligned}$$

and $\mathbf{C}_n = [\mathbf{c}_{1.,n}', \dots, \mathbf{c}_{n.,n}']' = \left(\mathbf{A}_n + \mathbf{A}_n'\right)\left(\mathbf{I}_n - \rho_n\mathbf{M}_n\right)^{-1}$. The row and column sums of the matrices $\mathbf{C}_n$ are again seen to be uniformly bounded in absolute value. Let $\overline{K} < \infty$ denote a uniform bound for the row and column sums of the absolute elements of the matrices $\mathbf{A}_n$ and $\mathbf{C}_n$.

To prove the claim we now show that both ϕ_n and ψ_n are $o_p(1)$. Using the triangle and Hölder inequality we get

$$\begin{aligned}
|\phi_n| &= \left| n^{-1}\sum_{i=1}^{n}\boldsymbol{\Delta}_n'\mathbf{d}_{i.,n}'\mathbf{c}_{i.,n}\boldsymbol{\varepsilon}_n \right| \tag{C.1} \\
&\leq n^{-1}\|\boldsymbol{\Delta}_n\|\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|\sum_{j=1}^{n}|c_{ij,n}|\,|\varepsilon_{j,n}| \leq n^{-1}\|\boldsymbol{\Delta}_n\|\sum_{j=1}^{n}|\varepsilon_{j,n}|\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|\,|c_{ij,n}| \\
&\leq n^{-1}\|\boldsymbol{\Delta}_n\|\sum_{j=1}^{n}|\varepsilon_{j,n}|\left(\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|^p\right)^{1/p}\left(\sum_{i=1}^{n}|c_{ij,n}|^q\right)^{1/q} \\
&\leq \overline{K}n^{1/p-1/2}\left(n^{1/2}\|\boldsymbol{\Delta}_n\|\right)\left(n^{-1}\sum_{j=1}^{n}|\varepsilon_{j,n}|\right)\left(n^{-1}\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|^p\right)^{1/p}
\end{aligned}$$

for $p = 2+\delta$ and $1/p + 1/q = 1$, and where $\delta > 0$ is as in Assumption 4. The last inequality utilizes the observation of Remark C.1. Since the $\varepsilon_{j,n}$ are independent with bounded second moments it follows that $n^{-1}\sum_{j=1}^{n}|\varepsilon_{j,n}| = O_p(1)$. The terms $n^{1/2}\|\boldsymbol{\Delta}_n\|$ and $n^{-1}\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|^p$ are $O_p(1)$ by Assumption 4. Since $n^{1/p-1/2} \to 0$ as $n \to \infty$ it follows that $\phi_n = o_p(1)$.

Again, using the triangle and Hölder inequality yields

$$\begin{aligned}
|\psi_n| &= \left| n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}\boldsymbol{\Delta}_n'\mathbf{d}_{i.,n}'a_{ij,n}\mathbf{d}_{j.,n}\boldsymbol{\Delta}_n \right| \tag{C.2} \\
&\leq n^{-1}\|\boldsymbol{\Delta}_n\|^2\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|\sum_{j=1}^{n}\|\mathbf{d}_{j.,n}\|\,|a_{ij,n}| \\
&\leq n^{-1}\|\boldsymbol{\Delta}_n\|^2\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|\left(\sum_{j=1}^{n}\|\mathbf{d}_{j.,n}\|^p\right)^{1/p}\left(\sum_{j=1}^{n}|a_{ij,n}|^q\right)^{1/q} \\
&\leq \overline{K}n^{1/p}\|\boldsymbol{\Delta}_n\|^2\left(n^{-1}\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|\right)\left(n^{-1}\sum_{j=1}^{n}\|\mathbf{d}_{j.,n}\|^p\right)^{1/p} \\
&\leq \overline{K}n^{1/p-1/2}n^{-1/2}(n^{1/2}\|\boldsymbol{\Delta}_n\|)^2\left(n^{-1}\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|^p\right)^{2/p}
\end{aligned}$$

34

with p and q as before. By Assumption 4 both $n^{-1}\sum_{i=1}^{n}\|\mathbf{d}_{i.,n}\|^{p}$ and $n^{1/2}\|\boldsymbol{\Delta}_{n}\|$ are $O_{p}(1)$. Since $n^{1/p-1/2}\to 0$ as $n\to\infty$ it follows that $\psi_{n}=o_{p}(1)$. From the last inequality we see also that $n^{1/2}\psi_{n}=o_{p}(1)$.

We next prove part (b) of the lemma. In the following let $\vartheta_{s,n}^{*}$ denote the s-th element of $n^{-1}\mathbf{D}_{n}'\mathbf{A}_{n}\mathbf{u}_{n}$. In light of the discussion after Assumption 3 and given Assumption 4 there exists a constant $\overline{\overline{K}}<\infty$ such that $Eu_{i,n}^{2}\leq\overline{\overline{K}}$ and $E\,|d_{js,n}|^{p}\leq\overline{\overline{K}}$. W.o.l.o.g. assume that the row and column sums of the matrices $\mathbf{A}_{n}$ are uniformly bounded by $\overline{\overline{K}}$. Utilizing the Cauchy-Schwarz and Lyapunov inequalities we then have $E\,|u_{i,n}|\,|d_{js,n}|\leq\left[Eu_{i,n}^{2}\right]^{1/2}\left[Ed_{js,n}^{2}\right]^{1/2}\leq\left[Eu_{i,n}^{2}\right]^{1/2}\left(E\,|d_{js,n}|^{p}\right)^{1/p}\leq\overline{\overline{K}}^{1/2+1/p}$ with p as before, and hence

$$E\left|\vartheta_{s,n}^{*}\right|=n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij,n}|\,E\left[|u_{i,n}|\,|d_{js,n}|\right]\leq\overline{\overline{K}}^{1/2+1/p}n^{-1}\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij,n}|\leq\overline{\overline{K}}^{3/2+1/p}<\infty,$$

which shows that indeed $E\left|n^{-1}\mathbf{d}_{.s,n}'\mathbf{A}_{n}\mathbf{u}_{n}\right|=O(1)$. Of course, the argument also shows that $\boldsymbol{\alpha}_{n}=n^{-1}E\mathbf{D}_{n}'(\mathbf{A}_{n}+\mathbf{A}_{n}')\mathbf{u}_{n}=O(1)$. Next observe that

$$n^{-1}\mathbf{D}_{n}'\mathbf{A}_{n}\widetilde{\mathbf{u}}_{n}=n^{-1}\mathbf{D}_{n}'\mathbf{A}_{n}\mathbf{u}_{n}+\phi_{n}^{*}$$

where $\phi_{n}^{*}=n^{-1}\mathbf{D}_{n}'\mathbf{A}_{n}\mathbf{D}_{n}\boldsymbol{\Delta}_{n}$. It now follows as a special case of our demonstration that $\phi_{n}=o_{p}(1)$ that also $\phi_{n}^{*}=o_{p}(1)$, which completes the proof of part (b).

We next prove part (c). In light of the proof of part (a) we have

$$n^{-1/2}\widetilde{\mathbf{u}}_{n}'\mathbf{A}_{n}\widetilde{\mathbf{u}}_{n}=n^{-1/2}\mathbf{u}_{n}'\mathbf{A}_{n}\mathbf{u}_{n}+[n^{-1}\mathbf{u}_{n}'(\mathbf{A}_{n}+\mathbf{A}_{n}')\mathbf{D}_{n}]n^{1/2}\boldsymbol{\Delta}_{n}+n^{1/2}\psi_{n}$$

with $n^{1/2}\psi_{n}=o_{p}(1)$. In light of part (b) we have $n^{-1}\mathbf{u}_{n}'(\mathbf{A}_{n}+\mathbf{A}_{n}')\mathbf{D}_{n}-\boldsymbol{\alpha}_{n}'=o_{p}(1)$. The claim follows since $n^{1/2}\boldsymbol{\Delta}_{n}=O_{p}(1)$ by Assumption 4. ∎

Remark C.2: For future reference it proves helpful to note that in light of Remark A.1 in Kelejian and Prucha (2004) the constant $\overline{K}$ used in proving the last claim of part (a) of the above lemma can be chosen as $\overline{K}=2c_{P}c_{A}$ where c_{P} and c_{A} denote a bound for the row and column sums of the absolute elements of $(\mathbf{I}_{n}-\rho_{n}\mathbf{M}_{n})^{-1}$ and $\mathbf{A}_{n}$. Furthermore it proves helpful to observe that in light of (C.1) and (C.2)

$$\left|\widetilde{\vartheta}_{n}-\vartheta_{n}\right|\leq 2c_{P}c_{A}\varsigma_{n}$$

where $\varsigma_{n}=o_{p}(1)$ does not depend on $\mathbf{A}_{n}$.

Proof of Theorem 1: The existence and measurability of $\widetilde{\rho}_{n}$ is assured by, e.g., Lemma 3.4 in Pötscher and Prucha (1997). The objective function of the weighted nonlinear least squares estimator and its corresponding non-stochastic counterpart are given by, respectively,

$$R_n(\omega, \rho) = \left[\widetilde{\boldsymbol{\Gamma}}_n(\rho, \rho^2)' - \widetilde{\boldsymbol{\gamma}}_n\right]' \widetilde{\Upsilon}_n \left[\widetilde{\boldsymbol{\Gamma}}_n(\rho, \rho^2)' - \widetilde{\boldsymbol{\gamma}}_n\right]$$

$$\overline{R}_n(\rho) = \left[\boldsymbol{\Gamma}_n(\rho, \rho^2)' - \boldsymbol{\gamma}_n\right]' \Upsilon_n \left[\boldsymbol{\Gamma}_n(\rho, \rho^2)' - \boldsymbol{\gamma}_n\right].$$

To prove the consistency of $\widetilde{\rho}_n$ we show that the conditions of, e.g., Lemma 3.1 in Pötscher and Prucha (1997) are satisfied for the problem at hand. We first show that ρ_n is an identifiably unique sequence of minimizers of $\overline{R}_n$. Observe that $\overline{R}_n(\rho) \geq 0$ and that $\overline{R}_n(\rho_n) = 0$. From (6) we have $\boldsymbol{\gamma}_n = \boldsymbol{\Gamma}_n[\rho_n, \rho_n^2]'$. Utilizing Assumption 5 we then get

$$
\begin{aligned}
\overline{R}_n(\rho) - \overline{R}_n(\rho_n) &= \overline{R}_n(\rho) \\
&= \left[\rho - \rho_n, \rho^2 - \rho_n^2\right] \boldsymbol{\Gamma}_n' \Upsilon_n \boldsymbol{\Gamma}_n \left[\rho - \rho_n, \rho^2 - \rho_n^2\right]' \\
&\geq \lambda_{\min}(\Upsilon_n) \lambda_{\min}(\boldsymbol{\Gamma}_n' \boldsymbol{\Gamma}_n) \left[\rho - \rho_n, \rho^2 - \rho_n^2\right] \left[\rho - \rho_n, \rho^2 - \rho_n^2\right]' \\
&\geq \lambda_* \left[\rho - \rho_n\right]^2
\end{aligned}
$$

for some $\lambda_* > 0$. Hence for every $\varepsilon > 0$ and n we have:

$$
\begin{aligned}
& \inf_{\{\rho \in [-a^\rho, a^\rho] : \|\rho - \rho_n\| \geq \varepsilon\}} [\overline{R}_n(\rho) - \overline{R}_n(\rho_n)] \\
&\geq \inf_{\{\rho \in [-a^\rho, a^\rho] : \|\rho - \rho_n\| \geq \varepsilon\}} \lambda_*^2 \left[\rho - \rho_n\right]^2 = \lambda_* \varepsilon^2 > 0,
\end{aligned}
$$

which proves that ρ_n is identifiably unique. Next let $\boldsymbol{\Phi}_n = [\boldsymbol{\Gamma}_n, -\boldsymbol{\gamma}_n]$ and $\widetilde{\boldsymbol{\Phi}}_n = [\widetilde{\boldsymbol{\Gamma}}_n, -\widetilde{\boldsymbol{\gamma}}_n]$, then

$$
\begin{aligned}
\left|R_n(\omega, \rho) - \overline{R}_n(\rho)\right| &= \left|[\rho, \rho^2, 1]\left[\widetilde{\boldsymbol{\Phi}}_n' \widetilde{\Upsilon}_n \widetilde{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_n' \Upsilon_n \boldsymbol{\Phi}_n\right][\rho, \rho^2, 1]'\right| \\
&\leq \left\|\widetilde{\boldsymbol{\Phi}}_n' \widetilde{\Upsilon}_n \widetilde{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_n' \Upsilon_n \boldsymbol{\Phi}_n\right\| \left\|\rho, \rho^2, 1\right\|^2 \\
&\leq \left\|\widetilde{\boldsymbol{\Phi}}_n' \widetilde{\Upsilon}_n \widetilde{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_n' \Upsilon_n \boldsymbol{\Phi}_n\right\| [1 + (a^\rho)^2 + (a^\rho)^4].
\end{aligned}
$$

As is readily seen from the respective second expressions on the r.h.s. of (7), the elements of $\boldsymbol{\Phi}_n$ and $\widetilde{\boldsymbol{\Phi}}_n$ are all of the form $n^{-1} E\mathbf{u}_n' \mathbf{A}_n \mathbf{u}_n$ and $n^{-1} \widetilde{\mathbf{u}}_n' \mathbf{A}_n \widetilde{\mathbf{u}}_n$ where the row and column sums of $\mathbf{A}_n$ are bounded uniformly in absolute value. It now follows immediately from Lemma C.1 that $\widetilde{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_n \xrightarrow{p} 0$, and that the elements of $\widetilde{\boldsymbol{\Phi}}_n$ and $\boldsymbol{\Phi}_n$ are, respectively, $O_p(1)$ and $O(1)$. The analogous properties are seen to hold for the elements of $\widetilde{\Upsilon}_n$ and Υ_n in light of Assumption 5. Given this it follows from the above inequality that $R_n(\omega, \rho) - \overline{R}_n(\rho)$ converges to zero uniformly over the optimization space $[-a^\rho, a^\rho]$, i.e.,

$$
\sup_{\rho \in [-a^\rho, a^\rho]} \left|R_n(\omega, \rho) - \overline{R}_n(\rho)\right| \leq \left\|\widetilde{\boldsymbol{\Phi}}_n' \widetilde{\Upsilon}_n \widetilde{\boldsymbol{\Phi}}_n - \boldsymbol{\Phi}_n' \Upsilon_n \boldsymbol{\Phi}_n\right\| [1 + (a^\rho)^2 + (a^\rho)^4] \xrightarrow{p} 0
$$

as $n \to \infty$. The consistency of $\widetilde{\rho}_n$ now follows directly from Lemma 3.1 in Pötscher and Prucha (1997). ∎

We make use of the following lemma.

Lemma C.2 : *Let $\mathbf{u}_n$ be defined by (2) and let $\mathbf{D}_n = [\mathbf{d}'_{1.,n}, \ldots, \mathbf{d}'_{n.,n}]'$, where $\mathbf{d}_{i,n}$ is defined in Assumption 4. Suppose Assumptions 1 - 3 hold, and suppose furthermore that the columns of $\mathbf{D}_n$ are of the form $\boldsymbol{\pi}_n + \boldsymbol{\Pi}_n \boldsymbol{\varepsilon}_n$, where the elements of $\boldsymbol{\pi}_n$ are bounded in absolute value and the row and column sums of $\boldsymbol{\Pi}_n$ are uniformly bounded in absolute value. Then (a) $E d_{ij,n}^4 \leq const < \infty$, and thus the moment condition in Assumption 4 is automatically implied, and (b) Assumption 6 is automatically implied.*

Proof. Observe that for the fourth moment of the i-th element of $\boldsymbol{\pi}_n + \boldsymbol{\Pi}_n \boldsymbol{\varepsilon}_n$ we have

$$
E\big(\pi_{i,n} + \sum_{j=1}^{n} \pi_{ij,n}\varepsilon_{j,n}\big)^4 \leq 2^4 \left[\pi_i^4 + E\left(\sum_{j=1}^{n} \pi_{ij,n}\varepsilon_{j,n} \right)^4 \right] \leq
$$

$$
2^4 \left[\pi_i^4 + \sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n} |\pi_{ij,n}|\,|\pi_{ik,n}|\,|\pi_{il,n}|\,|\pi_{im,n}|\, E\,|\varepsilon_{j,n}|\,|\varepsilon_{k,n}|\,|\varepsilon_{l.n}|\,|\varepsilon_{m,n}| \right] \leq const < \infty
$$

since by assumption the elements of $\boldsymbol{\pi}_n$ are bounded in absolute value and the row and column sums of $\boldsymbol{\Pi}_n$ are uniformly bounded in absolute value, and since $E\,|\varepsilon_{j,n}|\,|\varepsilon_{k,n}|\,|\varepsilon_{l.n}|\,|\varepsilon_{m,n}|$ is uniformly bounded in light of Assumption 2. This proves claim (a).

Given the maintained assumptions of the lemma and utilizing (3) the elements of $n^{-1}\mathbf{D}'_n \mathbf{A}_n \mathbf{u}_n$ are seen to be of the form

$$
\chi_n = n^{-1}\mathbf{b}_n^{*\prime}\boldsymbol{\varepsilon}_n + n^{-1}\boldsymbol{\varepsilon}'_n \mathbf{B}_n^{*}\boldsymbol{\varepsilon}_n
$$

where $\mathbf{b}_n^{*\prime} = \boldsymbol{\pi}'_n \mathbf{A}_n \left(\mathbf{I}_n - \rho_n \mathbf{M}_n\right)^{-1}$ and $\mathbf{B}_n^{*} = \boldsymbol{\Pi}'_n \mathbf{A}_n \left(\mathbf{I}_n - \rho_n \mathbf{M}_n\right)^{-1}$. Given Remark A.1 in Kelejian and Prucha (2004) the elements of $\mathbf{b}_n^{*}$ are seen to be bounded and the rows and column sums of $\mathbf{B}_n^{*}$ are seen to be uniformly bounded in absolute value. By arguments analogous to those in the proof of Lemma C.1 it then readily seen that $E\chi_n = O(1)$ and $var(\chi_n) = o(1)$. Claim (b) now follows from Chebychev's inequality. ∎

Proof of Theorem 2: Define

$$
q_n(\rho_n, \boldsymbol{\Delta}_n) = \widetilde{\boldsymbol{\gamma}}_n - \widetilde{\boldsymbol{\Gamma}}_n \left[\begin{array}{c} \rho_n \\ \rho_n^2 \end{array} \right] = \left[\begin{array}{c} n^{-1}\widetilde{\mathbf{u}}'_n \mathbf{C}_{1,n} \widetilde{\mathbf{u}}_n \\ n^{-1}\widetilde{\mathbf{u}}'_n \mathbf{C}_{2,n} \widetilde{\mathbf{u}}_n \end{array} \right] \tag{C.3}
$$

with $\mathbf{C}_{r,n} = (1/2)\left(\mathbf{I}_n - \rho_n \mathbf{M}'_n\right)\left(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}\right)\left(\mathbf{I}_n - \rho_n \mathbf{M}_n\right)$, and where the matrices $\mathbf{A}_{r,n}$ are defined by (5), $r = 1, 2$. The second equality in (C.3) is seen to hold in light of the discussion surrounding equations (4)-(9). For later use

37

we observe that the rows and column sums of $\mathbf{C}_{r,n}$ are uniformly bounded in absolute value; see, e.g., Remark A.1 in Kelejian and Prucha (2004).

We have shown in Theorem 1 that the GM estimator $\widetilde{\rho}_n$ defined in (9) is consistent. Apart on a set whose probability tends to zero the estimator satisfies the following first order condition

$$q_n(\widetilde{\rho}_n, \boldsymbol{\Delta}_n)' \widetilde{\Upsilon}_n \frac{\partial q_n(\widetilde{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho} = 0.$$

Substituting the mean value theorem expression

$$q_n(\widetilde{\rho}_n, \boldsymbol{\Delta}_n) = q_n(\rho_n, \boldsymbol{\Delta}_n) + \frac{\partial q_n(\overline{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho}(\widetilde{\rho}_n - \rho_n)$$

into the first order condition yields

$$\frac{\partial q_n(\widetilde{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho'} \widetilde{\Upsilon}_n \frac{\partial q_n(\overline{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho} n^{1/2}(\widetilde{\rho}_n - \rho_n) = -\frac{\partial q_n(\widetilde{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho'} \widetilde{\Upsilon}_n n^{1/2} q_n(\rho_n, \boldsymbol{\Delta}_n)$$

(C.4)

where $\overline{\rho}_n$ is some between value. Observe that

$$\frac{\partial q_n(\rho, \boldsymbol{\Delta}_n)}{\partial \rho} = -\widetilde{\boldsymbol{\Gamma}}_n \begin{bmatrix} 1 \\ 2\rho \end{bmatrix}$$

(C.5)

and consider the nonnegative scalars

$$
\begin{aligned}
\widetilde{\boldsymbol{\Xi}}_n &= \frac{\partial q_n(\widetilde{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho'} \widetilde{\Upsilon}_n \frac{\partial q_n(\overline{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho} = \begin{bmatrix} 1 \\ 2\widetilde{\rho}_n \end{bmatrix}' \widetilde{\boldsymbol{\Gamma}}_n' \widetilde{\Upsilon}_n \widetilde{\boldsymbol{\Gamma}}_n \begin{bmatrix} 1 \\ 2\overline{\rho}_n \end{bmatrix}, & \text{(C.6)} \\
\boldsymbol{\Xi}_n &= \begin{bmatrix} 1 \\ 2\rho_n \end{bmatrix}' \boldsymbol{\Gamma}_n' \Upsilon_n \boldsymbol{\Gamma}_n \begin{bmatrix} 1 \\ 2\rho_n \end{bmatrix}.
\end{aligned}
$$

In proving Theorem 1 we have demonstrated that $\widetilde{\boldsymbol{\Gamma}}_n - \boldsymbol{\Gamma}_n \overset{p}{\to} 0$ and that the elements of $\widetilde{\boldsymbol{\Gamma}}_n$ and $\boldsymbol{\Gamma}_n$ are $O_p(1)$ and $O(1)$, respectively. By Assumption 5 we have $\widetilde{\Upsilon}_n - \Upsilon_n = o_p(1)$ and also that the elements of $\widetilde{\Upsilon}_n$ and Υ_n are $O_p(1)$ and $O(1)$, respectively. Since $\widetilde{\rho}_n$ and $\overline{\rho}_n$ are consistent and bounded in absolute value, clearly

$$\widetilde{\boldsymbol{\Xi}}_n - \boldsymbol{\Xi}_n \overset{p}{\to} 0 \tag{C.7}$$

as $n \to \infty$, and furthermore $\widetilde{\boldsymbol{\Xi}}_n = O_p(1)$ and $\boldsymbol{\Xi}_n = O(1)$. In particular $\boldsymbol{\Xi}_n \leq \lambda_{\Xi}^{**}$ where λ_{Ξ}^{**} is some finite constant. In light of Assumption 5 we have $\boldsymbol{\Xi}_n \geq \lambda_{\min}(\Upsilon_n)\lambda_{\min}(\boldsymbol{\Gamma}_n'\boldsymbol{\Gamma}_n)(1 + 4\rho_n^2) \geq \lambda_{\Xi}^*$ for some $\lambda_{\Xi}^* > 0$. Hence $0 < \boldsymbol{\Xi}_n^{-1} \leq 1/\lambda_{\Xi}^* < \infty$, and thus we also have $\boldsymbol{\Xi}_n^{-1} = O(1)$. Let $\widetilde{\boldsymbol{\Xi}}_n^+$ denote the generalized inverse of $\widetilde{\boldsymbol{\Xi}}_n$. It then follows as a special case of Lemma F1 in Pötscher and Prucha (1997) that $\widetilde{\boldsymbol{\Xi}}_n$ is nonsingular eventually with probability tending to one, that $\widetilde{\boldsymbol{\Xi}}_n^+ = O_p(1)$, and that

$$\widetilde{\boldsymbol{\Xi}}_n^+ - \boldsymbol{\Xi}_n^{-1} \overset{p}{\to} 0 \tag{C.8}$$

as $n \to \infty$.

Premultiplying (C.4) by $\widetilde{\boldsymbol{\Xi}}_n^+$ and rearranging terms yields

$$n^{1/2}(\widetilde{\rho}_n - \rho_n) = \left[1 - \widetilde{\boldsymbol{\Xi}}_n^+ \widetilde{\boldsymbol{\Xi}}_n\right] n^{1/2}(\widetilde{\rho}_n - \rho_n) - \widetilde{\boldsymbol{\Xi}}_n^+ \frac{\partial q_n(\widetilde{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho'} \widetilde{\Upsilon}_n n^{1/2} q_n(\rho_n, \boldsymbol{\Delta}_n).$$

In light of the above discussion the the first term on the r.h.s. is zero on ω-sets of probability tending to one. This yields

$$n^{1/2}(\widetilde{\rho}_n - \rho_n) = -\widetilde{\boldsymbol{\Xi}}_n^+ \frac{\partial q_n(\widetilde{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho'} \widetilde{\Upsilon}_n n^{1/2} q_n(\rho_n, \boldsymbol{\Delta}_n) + o_p(1). \qquad \text{(C.9)}$$

Observe that

$$\widetilde{\boldsymbol{\Xi}}_n^+ \frac{\partial q_n(\widetilde{\rho}_n, \boldsymbol{\Delta}_n)}{\partial \rho'} \widetilde{\Upsilon}_n - \boldsymbol{\Xi}_n^{-1} \left[\begin{array}{c} 1 \\ 2\rho_n \end{array}\right]' \boldsymbol{\Gamma}_n' \Upsilon_n = o_p(1). \qquad \text{(C.10)}$$

In light of (C.3) and Lemma C.1 the elements of $n^{1/2} q_n(\rho_n, \boldsymbol{\Delta}_n)$ can be expressed as $(r = 1, 2)$

$$n^{-1/2} \widetilde{\mathbf{u}}_n' \mathbf{C}_{r,n} \widetilde{\mathbf{u}}_n = n^{-1/2} \mathbf{u}_n' \mathbf{C}_{r,n} \mathbf{u}_n + \boldsymbol{\alpha}_{r,n}' n^{1/2} \boldsymbol{\Delta}_n + o_p(1)$$

where

$$\boldsymbol{\alpha}_{r,n} = 2n^{-1} E \mathbf{D}_n' \mathbf{C}_{r,n} \mathbf{u}_n.$$

Furthermore, the lemma implies that the elements of $\boldsymbol{\alpha}_{r,n}$ are uniformly bounded in absolute value. Utilizing $\mathbf{u}_n = (\mathbf{I}_n - \rho_n \mathbf{M}_n)^{-1} \boldsymbol{\varepsilon}_n$ and Assumption 7 we have

$$n^{1/2} q_n(\rho_n, \boldsymbol{\Delta}_n) = n^{-1/2} \left[\begin{array}{c} \frac{1}{2}\boldsymbol{\varepsilon}_n'(\mathbf{A}_{1,n} + \mathbf{A}_{1,n}')\boldsymbol{\varepsilon}_n + \mathbf{a}_{1,n}'\boldsymbol{\varepsilon}_n \\ \frac{1}{2}\boldsymbol{\varepsilon}_n'(\mathbf{A}_{2,n} + \mathbf{A}_{2,n}')\boldsymbol{\varepsilon}_n + \mathbf{a}_{2,n}'\boldsymbol{\varepsilon}_n \end{array}\right] + o_p(1) \qquad \text{(C.11)}$$

where $\mathbf{a}_{r,n} = \mathbf{T}_n \boldsymbol{\alpha}_{r,n}$, $r = 1, 2$. Observe that the elements of $\mathbf{a}_{r,n}$ are uniformly bounded in absolute value. As discussed before Theorem 2 in the text, the VC matrix of the vector of quadratic forms on the r.h.s. of (C.11) is given by $\boldsymbol{\Psi}_n$ where the elements of that matrix are given in (13). Those elements can be written more explicitly as

$$\psi_{rs,n} = \frac{1}{2n} \sum_{i=1}^{n} \sum_{j=1}^{n} (a_{ij,r,n} + a_{ji,r,n})(a_{ij,s,n} + a_{ji,s,n})\sigma_{i,n}^2 \sigma_{j,n}^2 + \frac{1}{n} \sum_{i=1}^{n} a_{i,r,n} a_{i,s,n} \sigma_{i,n}^2$$

$$\text{(C.12)}$$

By assumption $\lambda_{\min}(\boldsymbol{\Psi}_n) \geq const > 0$. Since the matrices $\mathbf{A}_{r,n}$, the vectors $\mathbf{a}_{r,n}$, and the innovations $\boldsymbol{\varepsilon}_n$ satisfy all of the remaining assumptions of the central limit theorem for vectors of linear quadratic forms given above as Theorem A.1 it follows that

$$\boldsymbol{\xi}_n = -\boldsymbol{\Psi}_n^{-1/2} n^{-1/2} \left[\begin{array}{c} \frac{1}{2}\boldsymbol{\varepsilon}_n'(\mathbf{A}_{1,n} + \mathbf{A}_{1,n}')\boldsymbol{\varepsilon}_n + \mathbf{a}_{1,n}'\boldsymbol{\varepsilon}_n \\ \frac{1}{2}\boldsymbol{\varepsilon}_n'(\mathbf{A}_{2,n} + \mathbf{A}_{2,n}')\boldsymbol{\varepsilon}_n + \mathbf{a}_{2,n}'\boldsymbol{\varepsilon}_n \end{array}\right] \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_2). \qquad \text{(C.13)}$$

Since the row and column sums of the matrices $\mathbf{A}_{r,n}$ are uniformly bounded in absolute value, and since the elements of $\mathbf{a}_{r,n}$ and the variances are uniformly

bounded by finite constants it is readily seen from (C.12) that the elements of $\boldsymbol{\Psi}_n$, and hence those of $\boldsymbol{\Psi}_n^{1/2}$ are uniformly bounded. It now follows from (C.9), (C.10) and (C.13) that

$$n^{1/2}(\widetilde{\rho}_n - \rho_n) = \boldsymbol{\Xi}_n^{-1} \begin{bmatrix} 1 \\ 2\rho_n \end{bmatrix}' \boldsymbol{\Gamma}_n' \Upsilon_n \boldsymbol{\Psi}_n^{1/2} \boldsymbol{\xi}_n + o_p(1). \qquad (C.14)$$

Observing that $\boldsymbol{\Xi}_n = \mathbf{J}_n' \Upsilon_n \mathbf{J}_n$, where $\mathbf{J}_n = \boldsymbol{\Gamma}_n[1, 2\rho_n]'$, this establishes (14). Since all of the nonstochastic terms on the r.h.s. of (C.14) are $O(1)$ it follows that $n^{1/2}(\widetilde{\rho}_n - \rho_n) = O_p(1)$. Next recall that $0 < \lambda_{\Xi}^* \le \boldsymbol{\Xi}_n \le \lambda_{\Xi}^{**} < \infty$. Hence

$$\boldsymbol{\Xi}_n^{-1} \mathbf{J}_n' \Upsilon_n \boldsymbol{\Psi}_n \Upsilon_n \mathbf{J}_n \boldsymbol{\Xi}_n^{-1} \ge \lambda_{\min}(\boldsymbol{\Psi}_n) \left[\lambda_{\min}(\Upsilon_n)\right]^2 \lambda_{\min}(\boldsymbol{\Gamma}_n' \boldsymbol{\Gamma}_n)(1+4\rho_n^2)/(\lambda_{\Xi}^{**})^2 \ge const > 0.$$

This establishes the last claim of the theorem. ∎

As part of proving Theorem 3 will be to show that $\widetilde{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}_n = o_p(1)$. Observe that the elements can be written as

$$\widetilde{\psi}_{rs,n} = \widetilde{\psi}_{rs,n}^* + \widetilde{\psi}_{rs,n}^{**}, \quad \text{and} \quad \psi_{rs,n} = \psi_{rs,n}^* + \psi_{rs,n}^{**} \qquad (C.15)$$

with

$$\begin{aligned}
\widetilde{\psi}_{rs,n}^* &= (2n)^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij,n} \widetilde{\varepsilon}_{i,n}^2 \widetilde{\varepsilon}_{j,n}^2, \quad &\psi_{rs,n}^* &= (2n)^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij,n} \sigma_{i,n}^2 \sigma_{j,n}^2, \\
\widetilde{\psi}_{rs,n}^{**} &= n^{-1} \widetilde{\boldsymbol{\alpha}}_{r,n}' \widetilde{\mathbf{P}}_n' \widetilde{\mathbf{F}}_n' \widetilde{\boldsymbol{\Sigma}}_n \widetilde{\mathbf{F}}_n \widetilde{\mathbf{P}}_n \widetilde{\boldsymbol{\alpha}}_{s,n}, \quad &\psi_{rs,n}^{**} &= n^{-1} \boldsymbol{\alpha}_{r,n}' \mathbf{P}_n' \mathbf{F}_n' \boldsymbol{\Sigma}_n \mathbf{F}_n \mathbf{P}_n \boldsymbol{\alpha}_{s,n},
\end{aligned} \qquad (C.16)$$

where $a_{ij,n} = (a_{ij,r,n} + a_{ji,r,n})(a_{ij,s,n} + a_{ji,s,n})$.

The next two lemmata will be used to show that $\widetilde{\psi}_{rs,n}^* - \psi_{rs,n}^* = o_p(1)$.

Lemma C.3 : *Suppose Assumptions 1 - 3 hold. Let $\Lambda_n = n^{-1}(\boldsymbol{\sigma}_n^2)' \mathbf{A}_n \boldsymbol{\sigma}_n^2$ and $\overline{\Lambda}_n = n^{-1}(\boldsymbol{\varepsilon}_n^2)' \mathbf{A}_n \boldsymbol{\varepsilon}_n^2$ with $\boldsymbol{\sigma}_n^2 = (\sigma_{1,n}^2, \ldots, \sigma_{n,n}^2)'$ and $\boldsymbol{\varepsilon}_n^2 = (\varepsilon_{1,n}^2, \ldots, \varepsilon_{n,n}^2)'$, and where the $n \times n$ matrices $\mathbf{A}_n$ are real nonstochastic and symmetric. Suppose further that the diagonal elements of the matrices $\mathbf{A}_n$ are zero and that their row and column sums are uniformly bounded in absolute value. Then $E\overline{\Lambda}_n = \Lambda_n = O(1)$ and $var(\overline{\Lambda}_n) = o(1)$, and hence $\overline{\Lambda}_n - \Lambda_n \xrightarrow{p} 0$ as $n \to \infty$, and $\overline{\Lambda}_n = O_p(1)$.*

Proof. Observe that

$$\begin{aligned}
\overline{\Lambda}_n &= n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij,n} \varepsilon_{i,n}^2 \varepsilon_{j,n}^2, & (C.17) \\
\Lambda_n &= n^{-1} \sum_{i=1}^n \sum_{j=1}^n a_{ij,n} \sigma_{i,n}^2 \sigma_{j,n}^2.
\end{aligned}$$

Since $a_{ii,n} = 0$ and the innovations $\varepsilon_{i,n}$ are assumed to be independent, clearly $E\overline{\Lambda}_n = \Lambda_n$. The claim that $\Lambda_n = O(1)$ follows from the triangle inequality, observing that the variances $\sigma_{i,n}^2$ are uniformly bounded and that furthermore also the row and column sums of the matrices $\mathbf{A}_n$ are uniformly

40

bounded in absolute value. Define $\boldsymbol{\zeta}_n = \boldsymbol{\varepsilon}_n^2 - \boldsymbol{\sigma}_n^2$, then $\boldsymbol{\zeta}_n \sim (\mathbf{0}, \boldsymbol{\Xi}_n)$ where $\boldsymbol{\Xi}_n = diag_{i=1}^n \left[E(\varepsilon_{i,n}^4) - \sigma_{i,n}^4 \right]$. Clearly

$$\overline{\Lambda}_n = \Lambda_n + n^{-1}\boldsymbol{\zeta}_n'\mathbf{A}_n\boldsymbol{\zeta}_n + 2n^{-1}(\boldsymbol{\sigma}_n^2)'\mathbf{A}_n\boldsymbol{\zeta}_n.$$

It now follows from Lemma A.1 that

$$var(\overline{\Lambda}_n) = 2n^{-2}tr\left[\mathbf{A}_n\boldsymbol{\Xi}_n\mathbf{A}_n\boldsymbol{\Xi}_n\right] + 4n^{-2}(\boldsymbol{\sigma}_n^2)'\mathbf{A}_n\boldsymbol{\Xi}_n\mathbf{A}_n'\boldsymbol{\sigma}_n^2.$$

Given the maintained assumptions it follows, e.g., from Remark A.1 in Kelejian and Prucha (2004) that the row and column sums of $\mathbf{A}_n\boldsymbol{\Xi}_n\mathbf{A}_n\boldsymbol{\Xi}_n$ and $\mathbf{A}_n\boldsymbol{\Xi}_n\mathbf{A}_n$ are uniformly bounded in absolute value. Given that the variances $\sigma_{i,n}^2$ are uniformly bounded it follows that $var(\overline{\Lambda}_n) = o(1)$. Hence by Chebychev's inequality $\overline{\Lambda}_n - \Lambda_n = o_p(1)$. The last claim is now obvious. $\blacksquare$

Lemma C.4 : *Suppose Assumptions 1 - 4 hold. Let $\boldsymbol{\varepsilon}_n = (\mathbf{I}_n - \rho_n\mathbf{M}_n)\,\mathbf{u}_n$, and let $\widetilde{\boldsymbol{\varepsilon}}_n = (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n)\,\widetilde{\mathbf{u}}_n$ with $\widetilde{\mathbf{u}}_n = \mathbf{u}_n + \mathbf{D}_n\boldsymbol{\Delta}_n$ and $\mathbf{D}_n = [\mathbf{d}_{1.,n}', \ldots, \mathbf{d}_{n.,n}']'$, and where $\widetilde{\rho}_n$ can be any estimator that satisfies $n^{1/2}(\widetilde{\rho}_n - \rho_n) = O_p(1)$. Define $\widetilde{\Lambda}_n = n^{-1}(\widetilde{\boldsymbol{\varepsilon}}_n^2)'\mathbf{A}_n(\widetilde{\boldsymbol{\varepsilon}}_n^2)$, $\overline{\Lambda}_n = n^{-1}(\boldsymbol{\varepsilon}_n^2)'\mathbf{A}_n\boldsymbol{\varepsilon}_n^2$ with $\widetilde{\boldsymbol{\varepsilon}}_n^2 = (\widetilde{\varepsilon}_{1,n}^2, \ldots, \widetilde{\varepsilon}_{n,n}^2)'$, $\boldsymbol{\varepsilon}_n^2 = (\varepsilon_{1,n}^2, \ldots, \varepsilon_{n,n}^2)'$, and where the $n \times n$ matrices $\mathbf{A}_n$ are real nonstochastic and symmetric. Suppose further that the diagonal elements of the matrices $\mathbf{A}_n$ are zero and that their row and column sums are uniformly bounded in absolute value, and that $Ed_{ij,n}^4 \leq K_d < \infty$. Then, $\widetilde{\Lambda}_n - \overline{\Lambda}_n \overset{p}{\to} 0$ as $n \to \infty$, and $\widetilde{\Lambda}_n = O_p(1)$.*

Proof. Observe that

$$\begin{aligned}
\widetilde{\Lambda}_n - \overline{\Lambda}_n &=& n^{-1}\sum_{i=1}^n\sum_{j=1}^n a_{ij,n}\left[\widetilde{\varepsilon}_{i,n}^2\widetilde{\varepsilon}_{j,n}^2 - \varepsilon_{i,n}^2\varepsilon_{j,n}^2\right] \qquad (\text{C.18})\\
&=& \psi_{1n} + \psi_{2n} + \psi_{3n}
\end{aligned}$$

with

$$\psi_{1n} = n^{-1}\sum_{i=1}^n\sum_{j=1}^n a_{ij,n}\left[\widetilde{\varepsilon}_{i,n}^2 - \varepsilon_{i,n}^2\right]\varepsilon_{j,n}^2, \qquad (\text{C.19})$$

$$\psi_{2n} = n^{-1}\sum_{i=1}^n\sum_{j=1}^n a_{ij,n}\varepsilon_{i,n}^2\left[\widetilde{\varepsilon}_{j,n}^2 - \varepsilon_{j,n}^2\right],$$

$$\psi_{3n} = n^{-1}\sum_{i=1}^n\sum_{j=1}^n a_{ij,n}\left[\widetilde{\varepsilon}_{i,n}^2 - \varepsilon_{i,n}^2\right]\left[\widetilde{\varepsilon}_{j,n}^2 - \varepsilon_{j,n}^2\right].$$

Observe that

$$\begin{aligned}
\widetilde{\boldsymbol{\varepsilon}}_n &=& (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n)\,\widetilde{\mathbf{u}}_n = (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n)\,(\mathbf{u}_n + \mathbf{D}_n\boldsymbol{\Delta}_n) \qquad (\text{C.20})\\
&=& \boldsymbol{\varepsilon}_n + (\mathbf{I}_n - \rho_n\mathbf{M}_n)\,\mathbf{D}_n\boldsymbol{\Delta}_n - (\widetilde{\rho}_n - \rho_n)\mathbf{M}_n\mathbf{u}_n - (\widetilde{\rho}_n - \rho_n)\mathbf{M}_n\mathbf{D}_n\boldsymbol{\Delta}_n\\
&=& \boldsymbol{\varepsilon}_n + \boldsymbol{\eta}_n
\end{aligned}$$

with
$$\boldsymbol{\eta}_n = \mathbf{R}_n \left[\boldsymbol{\Delta}'_n, (\rho_n - \widetilde{\rho}_n), (\rho_n - \widetilde{\rho}_n)\boldsymbol{\Delta}'_n \right]'$$

where
$$\mathbf{R}_n = [(\mathbf{I}_n - \rho_n \mathbf{M}_n)\mathbf{D}_n, \mathbf{M}_n (\mathbf{I}_n - \rho_n \mathbf{M}_n)^{-1}\boldsymbol{\varepsilon}_n, \mathbf{M}_n \mathbf{D}_n].$$

In light of Assumptions 3 and since the elements of $\mathbf{D}_n$ and $\boldsymbol{\varepsilon}_n$ have bounded fourth moments it follows that the columns of $\mathbf{R}_n$ are all of the form $\underline{\boldsymbol{\pi}}_n + \underline{\boldsymbol{\Pi}}_n \boldsymbol{\xi}_n$, where the elements of the $n \times 1$ vector $\underline{\boldsymbol{\pi}}_n$ are uniformly bounded in absolute value by some finite constant, where the row and column sums of the $n \times n$ matrix $\underline{\boldsymbol{\Pi}}_n$ are uniformly bounded in absolute value by some finite constant, and where the fourth moments of the elements of $\boldsymbol{\xi}_n$ are also bounded by some finite constant. In light of this it is easily seen that the fourth moments of the elements of $\mathbf{R}_n$ are bounded by some finite constant. Consequently

$$\left| \eta_{i,n} \right| \le \alpha_n \beta_{i,n} \tag{C.21}$$

where $\alpha_n = \|[\boldsymbol{\Delta}'_n, (\rho_n - \widetilde{\rho}_n), (\rho_n - \widetilde{\rho}_n)\boldsymbol{\Delta}'_n]\|$ and $\beta_{i,n} = \|\mathbf{r}_{i\cdot,n}\|$ with $E\beta_{i,n}^4 \le K_\beta$ where K_β is some finite constant. Of course, w.o.l.g., we can select K_β such that $E\beta_{i,n}^\gamma \le K_\beta$ for $\gamma \le 4$. Observe furthermore that in light of the maintained assumptions $n^{1/2}\alpha_n = O_p(1)$. Given Assumption 2 we have: $E\left|\varepsilon_{i,n}\right|^\gamma \le K_\varepsilon$, $\gamma \le 4+\eta$, for some $K_\varepsilon < \infty$. By the assumptions of the lemma $\sum_{l=1}^n |a_{il,n}| < K_a$ for some $K_a < \infty$. In the following let $K = \max\{1, K_a, K_\beta, K_\varepsilon\}$. In light of Remark C.1 it then follows that $\sum_{l=1}^n |a_{il,n}|^\gamma < K^\gamma$ and $\sum_{l=1}^n |b_{il,n}|^\gamma < K^\gamma$ for $\gamma \ge 1$.

From (C.20) and (C.21) we have

$$\left| \widetilde{\varepsilon}_{i,n}^2 - \varepsilon_{i,n}^2 \right| = \left| 2\varepsilon_{i,n}^2 \eta_{i,n} + \eta_{i,n}^2 \right| \le 2\alpha_n \beta_{i,n} \varepsilon_{i,n}^2 + \alpha_n^2 \beta_{i,n}^2. \tag{C.22}$$

Applying the triangle inequality to the first equation of (C.19) and making use of (C.22) yields

$$
\begin{aligned}
|\psi_{1n}| &\le 2\alpha_n n^{-1} \sum_{i=1}^n \beta_{i,n} \varepsilon_{i,n}^2 \sum_{j=1}^n |a_{ij,n}| \varepsilon_{j,n}^2 \\
&+ \alpha_n^2 n^{-1} \sum_{i=1}^n \beta_{i,n}^2 \sum_{j=1}^n |a_{ij,n}| \varepsilon_{j,n}^2 \le \delta_{1n}^* + \delta_{1n}^{**}, \tag{C.23} \\
\delta_{1n}^* &= 2K n^{1/r - 1/2}(n^{1/2}\alpha_n)\zeta_n \left[n^{-1}\sum_{i=1}^n \beta_{i,n}\varepsilon_{i,n}^2 \right], \\
\delta_{1n}^{**} &= K n^{1/r - 1}(n^{1/2}\alpha_n)^2 \zeta_n \left[n^{-1}\sum_{i=1}^n \beta_{i,n}^2 \right].
\end{aligned}
$$

where $\zeta_n = \left[n^{-1}\sum_{j=1}^n |\varepsilon_{j,n}|^{4+\eta} \right]^{1/r}$ with $r = 2 + \eta/2$. In obtaining the above result we also utilized that by Hölder's inequality $\sum_{j=1}^n |a_{ij,n}| \varepsilon_{j,n}^2 \le n^{1/r} K \zeta_n$.

42

Since $E\left|\varepsilon_{j,n}\right|^{4+\eta} \leq K$ we have $\zeta_n = O_p(1)$. As remarked $n^{1/2}\alpha_n = O_p(1)$. Since $E\beta_{i,n}\varepsilon_{i,n}^2 \leq \left[E\beta_{i,n}^2\right]^{1/2}\left[E\varepsilon_{i,n}^4\right]^{1/2} \leq K$ and $E\beta_{i,n}^2 \leq K$ it follows further that also the terms in square brackets in (C.23) are $O_p(1)$. Since $n^{1/r-1/2}$ and $n^{1/r-1}$ tend to zero as n tends to infinity we have $\delta_{1n}^* = o_p(1)$ and $\delta_{1n}^{**} = o_p(1)$ and hence $\psi_{1n} = o_p(1)$. Because of symmetry it follows that also $\psi_{2n} = o_p(1)$.

Applying the triangle inequality to the last equation of (C.19) and making use of (C.22) yields

$$|\psi_{3n}| \leq 4\alpha_n^2 n^{-1}\sum_{i=1}^n \beta_{i,n}\left|\varepsilon_{i,n}\right|\sum_{j=1}^n \left|a_{ij,n}\right|\beta_{j,n}\left|\varepsilon_{j,n}\right| \tag{C.24}$$

$$+\quad 2\alpha_n^3 n^{-1}\sum_{i=1}^n \beta_{i,n}\left|\varepsilon_{i,n}\right|\sum_{j=1}^n \left|a_{ij,n}\right|\beta_{j,n}^2 + 2\alpha_n^3 n^{-1}\sum_{j=1}^n \beta_{j,n}\left|\varepsilon_{j,n}\right|\sum_{i=1}^n \left|a_{ij,n}\right|\beta_{i,n}^2$$

$$+\quad \alpha_n^4 n^{-1}\sum_{i=1}^n \beta_{i,n}^2 \sum_{j=1}^n \left|a_{ij,n}\right|\beta_{j,n}^2 \leq \delta_{3n}^* + \delta_{3n}^{**} + \delta_{3n}^{***}$$

$$\delta_{3n}^* = 4Kn^{1/2-1}(n^{1/2}\alpha_n)^2\overline{\zeta}_n\left[n^{-1}\sum_{i=1}^n \beta_{i,n}\left|\varepsilon_{i,n}\right|\right]$$

$$\delta_{3n}^{**} = 4Kn^{-1}(n^{1/2}\alpha_n)^3\overline{\overline{\zeta}}_n\left[n^{-1}\sum_{i=1}^n \beta_{i,n}\left|\varepsilon_{i,n}\right|\right]$$

$$\delta_{3n}^{***} = Kn^{-3/2}(n^{1/2}\alpha_n)^4\overline{\overline{\zeta}}_n\left[n^{-1}\sum_{i=1}^n \beta_{i,n}^2\right]$$

where $\overline{\zeta}_n = \left[n^{-1}\sum_{j=1}^n \beta_{j,n}^2\varepsilon_{j,n}^2\right]^{1/2}$ and $\overline{\overline{\zeta}}_n = \left[n^{-1}\sum_{j=1}^n \beta_{j,n}^4\right]^{1/2}$. In obtaining the above result we also utilized that by Hölder's inequality $\sum_{j=1}^n \left|a_{ij,n}\right|\beta_{j,n}\left|\varepsilon_{j,n}\right| \leq n^{1/2}K\overline{\zeta}_n$, $\sum_{j=1}^n \left|a_{ij,n}\right|\beta_{j,n}^2 \leq n^{1/2}K\overline{\overline{\zeta}}_n$ and $\sum_{i=1}^n \left|a_{ij,n}\right|\beta_{i,,n}^2 \leq n^{1/2}K\overline{\overline{\zeta}}_n$. Since $E\beta_{j,n}^2\varepsilon_{j,n}^2 \leq \left[E\beta_{j,n}^4 E\varepsilon_{j,n}^4\right]^{1/2} \leq K$ and $E\beta_{j,n}^4 \leq K$ it follows that $\overline{\zeta}_n = O_p(1)$ and $\overline{\overline{\zeta}}_n = O_p(1)$. Since $E\beta_{i,n}\left|\varepsilon_{i,n}\right| \leq \left[E\beta_{i,n}^2\right]^{1/2}\left[E\varepsilon_{i,n}^2\right]^{1/2} \leq K$ all terms terms in square brackets in (C.24) are seen to be $O_p(1)$. Observing that $n^{1/2}\alpha_n = O_p(1)$ it follows that $\delta_{3n}^* = o_p(1)$, $\delta_{3n}^{**} = o_p(1)$, $\delta_{3n}^{***} = o_p(1)$ and hence $\psi_{3n} = o_p(1)$. This proves that $\widetilde{\Lambda}_n - \overline{\Lambda}_n \overset{p}{\rightarrow} 0$. The last claim follows trivially since by Lemma C.3 $\overline{\Lambda}_n = O_p(1)$. ∎

The next two lemmata will be used to show that $\widetilde{\psi}_{rs,n}^{**} - \psi_{rs,n}^{**} = o_p(1)$, where $\widetilde{\psi}_{rs,n}^{**}$ and $\psi_{rs,n}^{**}$ are defined in (C.15) and (C.16).

Lemma C.5 : *Suppose Assumptions 1 - 4 hold. Let $\boldsymbol{\varepsilon}_n = \left(\mathbf{I}_n - \rho_n\mathbf{M}_n\right)\mathbf{u}_n$, and let $\widetilde{\boldsymbol{\varepsilon}}_n = \left(\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n\right)\widetilde{\mathbf{u}}_n$ with $\widetilde{\mathbf{u}}_n = \mathbf{u}_n + \mathbf{D}_n\boldsymbol{\Delta}_n$ and $\mathbf{D}_n = [\mathbf{d}'_{1.,n},\ldots,\mathbf{d}'_{n.,n}]'$, and where $\widetilde{\rho}_n$ can be any estimator that satisfies $\widetilde{\rho}_n - \rho_n = o_p(1)$. Let $\mathbf{a}_n$ and*

$\mathbf{b}_n$ be $n \times 1$ vectors whose elements are uniformly bounded in absolute value by $c < \infty$ and let $\mathbf{\Sigma}_n = diag_{i=1,\ldots,n}(\sigma_i^2)$ and $\widetilde{\mathbf{\Sigma}}_n = diag_{i=1,\ldots,n}(\widetilde{\varepsilon}_{i,n}^2)$. Then:

(a) $n^{-1}\mathbf{a}_n' \widetilde{\mathbf{\Sigma}}_n \mathbf{b}_n - n^{-1}\mathbf{a}_n' \mathbf{\Sigma}_n \mathbf{b}_n = o_p(1)$ and $n^{-1}\mathbf{a}_n' \mathbf{\Sigma}_n \mathbf{b}_n = O(1)$.

(b) There exist random variables ς_n that do not depend on $\mathbf{a}_n$ and $\mathbf{b}_n$ such that

$$\left| n^{-1}\mathbf{a}_n' \widetilde{\mathbf{\Sigma}}_n \mathbf{b}_n - n^{-1}\mathbf{a}_n' \mathbf{\Sigma}_n \mathbf{b}_n \right| \leq K(c)(1 + \varsigma_n)$$

with $\varsigma_n = o_p(1)$ and where $K(c) < \infty$ is a constant that depends monotonically on c (as well as on some other bounds maintained in the assumptions).

Proof. Let $\widetilde{\tau}_n = n^{-1}\mathbf{a}_n' \widetilde{\mathbf{\Sigma}}_n \mathbf{b}_n$ and $\tau_n = n^{-1}\mathbf{a}_n' \mathbf{\Sigma}_n \mathbf{b}_n$. We furthermore define $\overline{\tau}_n = n^{-1}\mathbf{a}_n' \overline{\mathbf{\Sigma}}_n \mathbf{b}_n$ where $\overline{\mathbf{\Sigma}}_n = diag_{i=1,\ldots,n}(\varepsilon_{i,n}^2)$. Clearly

$$|\widetilde{\tau}_n - \tau_n| \leq |\overline{\tau}_n - \tau_n| + |\widetilde{\tau}_n - \overline{\tau}_n|. \tag{C.25}$$

By the weak law of large numbers for i.d. random variables, observing that the fourth moments of $\varepsilon_{i,n}$ are uniformly bounded, we have

$$\overline{\tau}_n - \tau_n = n^{-1}\sum_{i=1}^{n} a_{i,n}b_{i,n}\left(\varepsilon_{i,n}^2 - \sigma_{i,n}^2\right) = o_p(1) \tag{C.26}$$

and thus $|\overline{\tau}_n - \tau_n| = o_p(1)$. Next let

$$\zeta_{1,n} = n^{-1}\sum_{i=1}^{n}\left[\left|\varepsilon_{i,n}^2 - \sigma_{i,n}^2\right| - E\left|\varepsilon_{i,n}^2 - \sigma_{i,n}^2\right|\right]$$

and c_* be such that $\sigma_{i,n}^2 \leq c_*$, then using the triangle inequality several times, it is readily seen

$$|\overline{\tau}_n - \tau_n| \leq c^2\zeta_{1,n} + 2c^2c_*. \tag{C.27}$$

Again by the weak law of large numbers for i.d. random variables we have $\zeta_{1,n} = o_p(1)$.

It proves convenient to rewrite $\widetilde{\tau}_n$ and $\overline{\tau}_n$ as

$$\begin{aligned} \widetilde{\tau}_n &= n^{-1}\widetilde{\boldsymbol{\varepsilon}}_n' \mathbf{C}_n \widetilde{\boldsymbol{\varepsilon}}_n = n^{-1}\widetilde{\mathbf{u}}_n' \mathbf{C}_n \widetilde{\mathbf{u}}_n - 2\widetilde{\rho}_n n^{-1}\widetilde{\mathbf{u}}_n' \mathbf{M}_n' \mathbf{C}_n \widetilde{\mathbf{u}}_n + \widetilde{\rho}_n^2 n^{-1}\widetilde{\mathbf{u}}_n' \mathbf{M}_n' \mathbf{C}_n \mathbf{M}_n \widetilde{\mathbf{u}}_n, \\ \overline{\tau}_n &= n^{-1}\boldsymbol{\varepsilon}_n' \mathbf{C}_n \boldsymbol{\varepsilon}_n = n^{-1}\mathbf{u}_n' \mathbf{C}_n \mathbf{u}_n - 2\rho_n n^{-1}\mathbf{u}_n' \mathbf{M}_n' \mathbf{C}_n \mathbf{u}_n + \rho_n^2 n^{-1}\mathbf{u}_n' \mathbf{M}_n' \mathbf{C}_n \mathbf{M}_n \mathbf{u}_n, \end{aligned}$$

where $\mathbf{C}_n = diag_{i=1}^{n}(a_{i,n}b_{i,n})$. Given the maintained assumptions the row and column sums of the respective matrices forming the quadratic forms on the r.h.s. of the above equations are uniformly bounded in absolute value by some constants that depend monotonically on c as well as other bounds maintained in the assumptions; compare again Remark A.1 in Kelejian and Prucha (2004). It now follows from Remark C.2 after the proof of Lemma C.1, the triangle inequality

and observing that $\widetilde{\rho}_n - \rho_n = o_p(1)$, that there exist random variables $\varsigma_{2,n}$ that do not depend on $\mathbf{a}_n$ and $\mathbf{b}_n$ as well as a constant $k(c)$ that depends monotonically on c (as well as on some other bounds maintained in the assumptions) such that

$$|\widetilde{\tau}_n - \overline{\tau}_n| = k(c)\varsigma_{2,n} \tag{C.28}$$

and $\varsigma_{2,n} = o_p(1)$.

The first claim of part (a) of the lemma follows immediately form (C.25), (C.26)and (C.28). The second claim is obvious observing that $\left|n^{-1}\mathbf{a}_n'\mathbf{\Sigma}_n\mathbf{b}_n\right| \leq n^{-1}\sum_{i=1}^n |a_{i,n}|\,|b_{i,n}|\,\sigma_{i,n}^2 \leq c^2 c_* < \infty$. Part (b) of the lemma follows form (C.25), (C.27) and (C.28) with $K(c) = c^2 + 2c^2 c_* + k(c)$ and $\varsigma_n = \varsigma_{1,n} + \varsigma_{2,n}$, observing that $\varsigma_{1,n}$ and $\varsigma_{2,n}$ are nonnegative. ∎

Lemma C.6 : *Suppose Assumptions 1 - 4 hold. Furthermore assume that* $\sup_n |\rho_n| < 1$, *and the row sums and column sums of* $\mathbf{M}_n$ *are bounded uniformly in absolute value by, respectively, one and some finite constant. Let* $\boldsymbol{\varepsilon}_n = (\mathbf{I}_n - \rho_n\mathbf{M}_n)\,\mathbf{u}_n$, *and let* $\widetilde{\boldsymbol{\varepsilon}}_n = (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n)\,\widetilde{\mathbf{u}}_n$ *with* $\widetilde{\mathbf{u}}_n = \mathbf{u}_n + \mathbf{D}_n\boldsymbol{\Delta}_n$ *and* $\mathbf{D}_n = [\mathbf{d}_{1.,n}', \ldots, \mathbf{d}_{n.,n}']'$, *and where* $\widetilde{\rho}_n$ *can be any estimator that satisfies* $\widetilde{\rho}_n - \rho_n = o_p(1)$. *Let*

$$\mathbf{F}_n = \widetilde{\mathbf{F}}_n = \mathbf{H}_n$$

or

$$\mathbf{F}_n = (\mathbf{I}_n - \rho_n\mathbf{M}_n)^{-1}\mathbf{H}_n \qquad and \qquad \widetilde{\mathbf{F}}_n = (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n)^+\mathbf{H}_n,$$

where $\mathbf{H}_n$ *is an* $n \times p_*$ *matrix whose elements are uniformly bounded in absolute value, let* $\mathbf{\Sigma}_n = diag_{i=1,\ldots,n}(\sigma_i^2)$ *and* $\widetilde{\mathbf{\Sigma}}_n = diag_{i=1,\ldots,n}(\widetilde{\varepsilon}_{i,n}^2)$, *then* $n^{-1}\widetilde{\mathbf{F}}_n'\widetilde{\mathbf{\Sigma}}_n\widetilde{\mathbf{F}}_n - n^{-1}\mathbf{F}_n'\mathbf{\Sigma}_n\mathbf{F}_n = o_p(1)$ *and* $n^{-1}\mathbf{F}_n'\mathbf{\Sigma}_n\mathbf{F}_n = O(1)$.[17]

Proof. The proof will focus on the case where $\mathbf{F}_n = (\mathbf{I}_n - \rho_n\mathbf{M}_n)^{-1}\mathbf{H}_n$ and $\widetilde{\mathbf{F}}_n = (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n)^+\mathbf{H}_n$ in that it will be obvious that the arguments implicitly also cover the case where $\mathbf{F}_n = \widetilde{\mathbf{F}}_n = \mathbf{H}_n$.

Clearly under the maintained assumptions there exists a ρ_* with $\sup_n |\rho_n| < \rho_* < 1$ and so the row sums and column sums of $\rho_*\mathbf{M}_n$ are bounded uniformly in absolute value by, respectively, one and some finite constant.

Now let ν_n denote the difference in the (r, s)-th elements of $n^{-1}\widetilde{\mathbf{F}}_n'\widetilde{\mathbf{\Sigma}}_n\widetilde{\mathbf{F}}_n$ and $n^{-1}\mathbf{F}_n'\mathbf{\Sigma}_n\mathbf{F}_n$. Then

$$\nu_n = \sum_{i=1}^7 \nu_{i,n} \tag{C.29}$$

---

[17]We would like to thank Benedikt Pötscher for very helpful comments on parts of the proof of this lemma.

where

$$\nu_{1,n} = n^{-1}(\widetilde{\mathbf{f}}_{.r,n} - \mathbf{f}_{.r,n})'(\widetilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)(\widetilde{\mathbf{f}}_{.s,n} - \mathbf{f}_{.s,n}) \qquad (\text{C.30})$$

$$\nu_{2,n} = n^{-1}(\widetilde{\mathbf{f}}_{.r,n} - \mathbf{f}_{.r,n})'(\widetilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)\mathbf{f}_{.s,n}$$

$$\nu_{3,n} = n^{-1}\mathbf{f}_{.r,n}'(\widetilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)(\widetilde{\mathbf{f}}_{.s,n} - \mathbf{f}_{.s,n})$$

$$\nu_{4,n} = n^{-1}\mathbf{f}_{.r,n}'(\widetilde{\boldsymbol{\Sigma}}_n - \boldsymbol{\Sigma}_n)\mathbf{f}_{.s,n}$$

$$\nu_{5,n} = n^{-1}(\widetilde{\mathbf{f}}_{.r,n} - \mathbf{f}_{.r,n})'\boldsymbol{\Sigma}_n(\widetilde{\mathbf{f}}_{.s,n} - \mathbf{f}_{.s,n})$$

$$\nu_{6,n} = n^{-1}(\widetilde{\mathbf{f}}_{.r,n} - \mathbf{f}_{.r,n})'\boldsymbol{\Sigma}_n\mathbf{f}_{.s,n}$$

$$\nu_{7,n} = n^{-1}\mathbf{f}_{.r,n}'\boldsymbol{\Sigma}_n(\widetilde{\mathbf{f}}_{.s,n} - \mathbf{f}_{.s,n})$$

and

$$\widetilde{\mathbf{f}}_{.s,n} - \mathbf{f}_{.s,n} = \left[(\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n)^+ - (\mathbf{I}_n - \rho_n\mathbf{M}_n)^{-1}\right]\mathbf{h}_{.s,n}.$$

In the following we will be confronted with terms of the form

$$\varkappa_n^{(k,l)} = n^{-1}\rho_*^{l+k}\mathbf{h}_{.r,n}'\mathbf{M}_n'^l\widetilde{\boldsymbol{\Sigma}}_n\mathbf{M}_n^k\mathbf{h}_{.s,n} - n^{-1}\rho_*^{l+k}\mathbf{h}_{.r,n}'\mathbf{M}_n'^l\boldsymbol{\Sigma}_n\mathbf{M}_n^k\mathbf{h}_{.s,n}. \quad (\text{C.31})$$

Under the maintained assumptions the row sums of $\rho_*^k\mathbf{M}_n^k$ are bounded uniformly in absolute value by one; this is readily seen by arguments like those in Kelejian and Prucha (1999, p. 526). Let $c < \infty$ be a bound for the absolute elements of $\mathbf{H}_n$, then it follows that c is also a bound for the absolute elements of $\rho_*^k\mathbf{M}_n^k\mathbf{h}_{.s,n}$. Given this it follows immediately form Lemma C.5 that $\varkappa_n^{(k,l)} = o_p(1)$ and that there exists a random variable $\varsigma_n = o_p(1)$ and a constant $K(c)$ such that

$$\left|\varkappa_n^{(k,l)}\right| \le K(c)(1 + \varsigma_n). \qquad (\text{C.32})$$

In the following let c be such that also $\sigma_{i,n}^2 \le c$.

In proving the lemma we will use the usual subsequence argument that utilizes that a sequence of random vectors taking their values in $\mathbf{R}^k$, $1 \le k \le \infty$, converges in probability iff for every subsequence there exists a subsequence such that the random vectors converge almost surely; see, e.g., Gänsler and Stute (1977), pp. 61-62 and p. 332. Also observe that convergence in probability [almost sure convergence] of random vectors taking their values in $\mathbf{R}^k$, $1 \le k \le \infty$, is equivalent with convergence in probability [almost sure convergence] for each of the coordinates of the random vector. Now let (n_m) denote some subsequence, then there exists a subsequence (n'_m) such that for $\omega \in A$, $P(A) = 1$,

$$\left|\widetilde{\rho}_{n'_m}(\omega) - \rho_{n'_m}\right| \to 0, \quad \varsigma_{n'_m}(\omega) \to 0, \qquad (\text{C.33})$$

$$\varkappa_{n'_m}^{(k,l)}(\omega) \to 0, \quad , k, l = 1, \ldots, \infty,$$

as $n_{m'} \to \infty$. Consequently there exists an index N_ω such that for all $n'_m \ge N_\omega$:

$$\varsigma_{n'_m}(\omega) \le K(c) \text{ and } \left|\widetilde{\rho}_{n'_m}(\omega)\right| \le \rho_{**}, \qquad (\text{C.34})$$

46

where $\rho_{**} = (\sup_n |\rho_n| + \rho_*)/2$.

In the following assume that $n'_m \geq N_\omega$. Since $\left|\widetilde{\rho}_{n'_m}(\omega)\right| < \rho_*$ the row sums of $\widetilde{\rho}_{n'_m}(\omega)\mathbf{M}_{n'_m}$ are less than unity in absolute value. In light of Horn and Johnson (1985), p. 301, it then follows that $\mathbf{I}_{n'_m} - \widetilde{\rho}_{n'_m}(\omega)\mathbf{M}_{n'_m}$ is invertible and that

$$\left(\mathbf{I}_{n'_m} - \widetilde{\rho}_{n'_m}(\omega)\mathbf{M}_{n'_m}\right)^+ - \left(\mathbf{I}_{n'_m} - \rho_{n'_m}\mathbf{M}_{n'_m}\right)^{-1} = \sum_{l=1}^{\infty} \left[\widetilde{\rho}_{n'_m}^l(\omega) - \rho_{n'_m}^l\right]\mathbf{M}_{n'_m}^l.$$

Hence in light of the definition of $\nu_{1,n}$ in (C.30) we have

$$\nu_{1,n'_m}(\omega) = \sum_{l=1}^{\infty}\sum_{k=1}^{\infty}\chi_{n'_m}^{(k,l)}(\omega)$$

with

$$\chi_{n'_m}^{(k,l)}(\omega) = \left[\left(\widetilde{\rho}_{n'_m}^l(\omega) - \rho_{n'_m}^l\right)/\rho_*^l\right]\left[\left(\widetilde{\rho}_{n'_m}^k(\omega) - \rho_{n'_m}^k\right)/\rho_*^k\right]\varkappa_{n'_m}^{(k,l)}(\omega).$$

Clearly $\chi_{n'_m}^{(k,l)}(\omega) \to 0$ in light of (C.33) and (C.34). Furthermore, observing that $\left|\widetilde{\rho}_{n'_m}^l(\omega) - \rho_{n'_m}^l\right|/\rho_*^l \leq 2(\rho_{**}/\rho_*)^l$ and $\left|\varkappa_{n'_m}^{(k,l)}(\omega)\right| \leq 2K(c)$ we have $\left|\chi_{n'_m}^{(k,l)}(\omega)\right| \leq B^{(l,k)}$ with $B^{(l,k)} = 8K(c)(\rho_{**}/\rho_*)^{l+k}$. Since $\rho_{**}/\rho_* < 1$ by construction clearly $\sum_{l=1}^{\infty}\sum_{k=1}^{\infty}B^{(l,k)} < \infty$. Hence it follows from dominated convergence that $\nu_{1,n'_m}(\omega) \to 0$ as $n_{m'} \to \infty$. This completes the proof that $\nu_{1,n} = o_p(1)$.

By analogous arguments we also see that $\nu_{i,n} = o_p(1)$ for $i = 2,\ldots,7$. Hence $\nu_n = o_p(1)$ and thus $n^{-1}\widetilde{\mathbf{F}}'_n\widetilde{\boldsymbol{\Sigma}}_n\widetilde{\mathbf{F}}_n - n^{-1}\mathbf{F}'_n\boldsymbol{\Sigma}_n\mathbf{F}_n = o_p(1)$. That $n^{-1}\mathbf{F}'_n\boldsymbol{\Sigma}_n\mathbf{F}_n = O(1)$ follows from the properties maintained for the row and column sums of $(\mathbf{I}_n - \rho_n\mathbf{M}_n)^{-1}$ and the elements of $\boldsymbol{\Sigma}_n$ and $\mathbf{H}_n$. $\blacksquare$

Proof of Theorem 3: We first demonstrate that $\widetilde{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}_n = o_p(1)$, utilizing the expressions for the elements of $\widetilde{\boldsymbol{\Psi}}_n$ and $\boldsymbol{\Psi}_n$ in (C.15) and (C.16). Observe that under the maintained assumption the row and column sums of the absolute elements of the matrices $\mathbf{A}_{r,n}$ and $\mathbf{A}_{s,n}$ are uniformly bounded, and thus clearly are those of the matrices $\mathbf{A}_n = (a_{ij,n})$ with $a_{ij,n} = (a_{ij,r,n} + a_{ji,r,n})(a_{ij,s,n} + a_{ji,s,n})$. It then follows directly from Lemmata C.3 and C.4 that $\widetilde{\psi}^*_{rs,n} - \psi^*_{rs,n} = o_p(1)$, $\psi^*_{rs,n} = O(1)$ and $\widetilde{\psi}^*_{rs,n} = O_p(1)$.

Observing that the row and column sums of $(\mathbf{I}_n - \rho_n\mathbf{M}'_n)(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{I}_n - \rho_n\mathbf{M}_n)$ are bounded uniformly in absolute value it follows from Lemma C.1 that $\widetilde{\boldsymbol{\alpha}}_{r,n} - \boldsymbol{\alpha}_{r,n} = o_p(1)$, $\boldsymbol{\alpha}_{r,n} = O(1)$ and hence $\widetilde{\boldsymbol{\alpha}}_{r,n} = O_p(1)$. By assumption $\widetilde{\mathbf{P}}_n - \mathbf{P}_n = o_p(1)$, $\mathbf{P}_n = O(1)$, and hence $\widetilde{\mathbf{P}}_n = O_p(1)$. Observe that by Lemma C.6 we have $n^{-1}\widetilde{\mathbf{F}}'_n\widetilde{\boldsymbol{\Sigma}}_n\widetilde{\mathbf{F}}_n - n^{-1}\mathbf{F}'_n\boldsymbol{\Sigma}_n\mathbf{F}_n = o_p(1)$, $n^{-1}\mathbf{F}'_n\boldsymbol{\Sigma}_n\mathbf{F}_n = O(1)$, and hence $n^{-1}\widetilde{\mathbf{F}}'_n\widetilde{\boldsymbol{\Sigma}}_n\widetilde{\mathbf{F}}_n = O_p(1)$. Thus $\widetilde{\psi}^{**}_{rs,n} - \psi^{**}_{rs,n} = o_p(1)$, $\psi^{**}_{rs,n} = O(1)$, and $\widetilde{\psi}^{**}_{rs,n} = O_p(1)$. Hence $\widetilde{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}_n = o_p(1)$, $\boldsymbol{\Psi}_n = O(1)$ and $\widetilde{\boldsymbol{\Psi}}_n = O_p(1)$.

By Assumption 5 we have $\widetilde{\boldsymbol{\Upsilon}}_n - \boldsymbol{\Upsilon}_n = o_p(1)$, $\boldsymbol{\Upsilon}_n = O(1)$ and thus $\widetilde{\boldsymbol{\Upsilon}}_n = O_p(1)$. Let $\boldsymbol{\Xi}_n = \mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{J}_n$ and $\widetilde{\boldsymbol{\Xi}}_n = \widetilde{\mathbf{J}}'_n\widetilde{\boldsymbol{\Upsilon}}_n\widetilde{\mathbf{J}}_n$, then, as demonstrated in the proof of

47

Theorem 2, $\widetilde{\mathbf{J}}_n = O_p(1)$, $\mathbf{J}_n = O(1)$, $\widetilde{\boldsymbol{\Xi}}_n^+ = O_p(1)$ and $\boldsymbol{\Xi}_n^{-1} = O_p(1)$, and furthermore $\widetilde{\mathbf{J}}_n - \mathbf{J}_n \overset{p}{\to} 0$ and $\widetilde{\boldsymbol{\Xi}}_n^+ - \boldsymbol{\Xi}_n^{-1} \overset{p}{\to} 0$ as $n \to \infty$. (There is a slight difference in the definition of $\widetilde{\boldsymbol{\Xi}}_n$ here and in the proof of Theorem 2, which does not affect the claim.) Given the above results it is now obvious that $\widetilde{\boldsymbol{\Omega}}_{\widetilde{\rho}_n} - \boldsymbol{\Omega}_{\widetilde{\rho}_n} = o_p(1)$. ∎

Proof of Theorem 4: We concentrate again on the case $\mathbf{F}_n = (\mathbf{I}_n - \rho_n \mathbf{M}'_n)^{-1} \mathbf{H}_n$. The first line in (27) is seen to hold in light of Assumption 7 and Theorem 2. We next verify that $\boldsymbol{\xi}_{\circ,n} \overset{d}{\to} N(\mathbf{0}, \mathbf{I}_{p_*+2})$ utilizing the central limit theorem for vectors of linear quadratic forms given above as Theorem A.1. By assumption $\lambda_{\min}(\boldsymbol{\Psi}_{\circ,n}) \geq c^*_{\Psi_\circ} > 0$. In proving Theorem 2 we verified that the innovations $\boldsymbol{\varepsilon}_n$ and the elements of $\mathbf{a}_{r,n}$ and $\mathbf{A}_{r,n}$ appearing in $\mathbf{v}_n$ all satisfy the conditions of the central limit theorem. Since the row and column sums of $(\mathbf{I}_n - \rho_n \mathbf{M}'_n)^{-1}$ and the elements of $\mathbf{H}_n$ are uniformly bounded in absolute value it follows that the elements of $\mathbf{F}_n$ are also uniformly bounded in absolute value. Thus all conditions of Theorem A.1 are satisfied, which verifies the claim.

In proving Theorems 2 and 3 we have shown that $\widetilde{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}_n = o_p(1)$, $\boldsymbol{\Psi}_n = O(1)$ and thus $\widetilde{\boldsymbol{\Psi}}_n = O_p(1)$. By analogous argumentation it is readily seen that the other sub-matrices of $\widetilde{\boldsymbol{\Psi}}_{\circ,n}$ and $\boldsymbol{\Psi}_{\circ,n}$ have the same properties, and thus $\widetilde{\boldsymbol{\Psi}}_{\circ,n} - \boldsymbol{\Psi}_{\circ,n} = o_p(1)$ and $\boldsymbol{\Psi}_{\circ,n} = O(1)$. By assumption $\widetilde{\mathbf{P}}_n - \mathbf{P}_n = o_p(1)$, $\mathbf{P}_n = O(1)$ and thus $\widetilde{\mathbf{P}}_n = O_p(1)$, and furthermore $\widetilde{\boldsymbol{\Upsilon}}_n - \boldsymbol{\Upsilon}_n = o_p(1)$, $\boldsymbol{\Upsilon}_n = O(1)$ and thus $\widetilde{\boldsymbol{\Upsilon}}_n = O_p(1)$. In proving Theorem 2 we have verified that $\widetilde{\mathbf{J}}_n - \mathbf{J}_n \overset{p}{\to} 0$, $\mathbf{J}_n = O(1)$ and $\widetilde{\mathbf{J}}_n = O_p(1)$, and furthermore that $(\widetilde{\mathbf{J}}'_n \widetilde{\boldsymbol{\Upsilon}}_n \widetilde{\mathbf{J}}_n)^+ - (\mathbf{J}'_n \boldsymbol{\Upsilon}_n \mathbf{J}_n)^{-1} = o_p(1)$, $(\widetilde{\mathbf{J}}'_n \widetilde{\boldsymbol{\Upsilon}}_n \widetilde{\mathbf{J}}_n)^+ = O_p(1)$ and $(\mathbf{J}'_n \boldsymbol{\Upsilon}_n \mathbf{J}_n)^{-1} = O(1)$. The claim that $\widetilde{\boldsymbol{\Omega}}_{\circ,n} - \boldsymbol{\Omega}_{\circ,n}(\boldsymbol{\Upsilon}_n) = o_p(1)$ and $\boldsymbol{\Omega}_{\circ,n} = O(1)$ is now obvious. ∎

D Appendix: Proofs for Section 4

Proof of Lemma 3: Utilizing (3) it follows from Assumptions 3, 8 and $\sup_n \|\boldsymbol{\beta}_n\| < \infty$ that all columns of $\mathbf{Z}_n = [\mathbf{X}_n, \mathbf{W}_n \mathbf{y}_n]$ are of the form $\vartheta_n = \boldsymbol{\pi}_n + \boldsymbol{\Pi}_n \boldsymbol{\varepsilon}_n$, where the elements of the $n \times 1$ vector $\boldsymbol{\pi}_n$ are bounded in absolute value and the row and column sums of the $n \times n$ matrix $\boldsymbol{\Pi}_n$ are uniformly bounded in absolute value by some finite constant; compare, e.g., Remark A.1. in Kelejian and Prucha (2004). The claims that the fourth moments of $\mathbf{D}_n = -\mathbf{Z}_n$ are uniformly bounded by a finite constant, and that Assumption 6 holds, now follows directly from Lemma C.2.

Clearly
$$n^{1/2}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n) = \widetilde{\mathbf{P}}'_n n^{-1/2} \mathbf{F}'_n \boldsymbol{\varepsilon}_n$$

where $\widetilde{\mathbf{P}}_n$ is defined in the lemma and $\mathbf{F}_n = (\mathbf{I}_n - \rho_n \mathbf{M}'_n)^{-1} \mathbf{H}_n$. Given Assumption 10 clearly $\widetilde{\mathbf{P}}_n = \mathbf{P}_n + o_p(1)$ and $\mathbf{P}_n = O(1)$, with $\mathbf{P}_n$ defined in the lemma. Since by Assumption 3 the row and column sums of $(\mathbf{I}_n - \rho_n \mathbf{M}_n)^{-1}$ are uniformly

bounded in absolute value, and since by Assumption 9 the elements of $\mathbf{H}_n$ are uniformly bounded in absolute value, it follows that the elements of $\mathbf{F}_n$ are uniformly bounded in absolute value. By Assumption 2, $E(\boldsymbol{\varepsilon}_n) = 0$ and its diagonal VC matrix, $\boldsymbol{\Sigma}_n$, has uniformly bounded elements. Therefore, $En^{-1/2}\mathbf{F}'_n\boldsymbol{\varepsilon}_n = 0$ and the elements of $VC(n^{-1/2}\mathbf{F}'_n\boldsymbol{\varepsilon}_n) = n^{-1}\mathbf{F}'_n\boldsymbol{\Sigma}_n\mathbf{F}_n$ are also uniformly bounded in absolute value. Thus, by Chebyshev's inequality $n^{-1/2}\mathbf{F}'_n\boldsymbol{\varepsilon}_n = O_p(1)$, and consequently $n^{1/2}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_n) = \mathbf{P}'_n n^{-1/2}\mathbf{F}'_n\boldsymbol{\varepsilon}_n + o_p(1)$ and $\mathbf{P}'_n n^{-1/2}\mathbf{F}'_n\boldsymbol{\varepsilon}_n = O_p(1)$. This completes the proof recalling that $\mathbf{T}_n = \mathbf{F}_n\mathbf{P}_n$. ∎

Proof of Lemma 4: Note from (1) and (2) that

$$\mathbf{y}_{n*}(\hat{\rho}_n) = \mathbf{Z}_{n*}(\hat{\rho}_n)\boldsymbol{\delta}_n + \boldsymbol{\varepsilon}_n - (\hat{\rho}_n - \rho_n)\mathbf{M}_n\mathbf{u}_n$$

and hence

$$
\begin{aligned}
&n^{1/2}[\widehat{\boldsymbol{\delta}}_n(\hat{\rho}_n) - \boldsymbol{\delta}_n] \\
={} &\left[n^{-1}\widehat{\mathbf{Z}}'_{n*}(\hat{\rho}_n)\mathbf{Z}_{n*}(\hat{\rho}_n)\right]^{-1} n^{-1/2}\widehat{\mathbf{Z}}'_{n*}(\hat{\rho}_n)\left[\boldsymbol{\varepsilon}_n - (\hat{\rho}_n - \rho_n)\mathbf{M}_n\mathbf{u}_n\right] \\
={} &\widetilde{\mathbf{P}}_n^{*\prime}\left[n^{-1/2}\mathbf{F}_n^{*\prime}\boldsymbol{\varepsilon}_n - (\hat{\rho}_n - \rho_n)n^{-1/2}\mathbf{F}_n^{**\prime}\boldsymbol{\varepsilon}_n\right],
\end{aligned}
$$

where $\widetilde{\mathbf{P}}_n^*$ is defined in the lemma, and with $\mathbf{F}_n^* = \mathbf{H}_n$ and $\mathbf{F}_n^{**} = (\mathbf{I}_n - \rho_n\mathbf{M}'_n)^{-1}\mathbf{M}'_n\mathbf{H}_n$. In light of Assumption 10, and since $\hat{\rho}_n$ is $n^{1/2}$-consistent, it follows that

$$n^{-1}\widehat{\mathbf{Z}}'_{n*}(\hat{\rho}_n)\mathbf{Z}_{n*}(\hat{\rho}_n) - \mathbf{Q}'_{\mathbf{HZ}*}(\rho_n)\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_n) = o_p(1).$$

Since by Assumption 10 we have $\mathbf{Q}'_{\mathbf{HZ}*}(\rho_n)\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_n) = O(1)$ and $[\mathbf{Q}'_{\mathbf{HZ}*}(\rho_n)\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_n)]^{-1} = O(1)$ it follows that

$$[n^{-1}\widehat{\mathbf{Z}}'_{n*}(\hat{\rho}_n)\mathbf{Z}_{n*}(\hat{\rho}_n)]^{-1} - [\mathbf{Q}'_{\mathbf{HZ}*}(\rho_n)\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_n)]^{-1} = o_p(1);$$

compare, e.g., Pötscher and Prucha (1997), Lemma F1. In light of this it follows further that $\widetilde{\mathbf{P}}_n^* - \mathbf{P}_n^* = o_p(1)$ and $\mathbf{P}_n^* = O(1)$, with $\mathbf{P}_n^*$ defined in the lemma. By argumentation analogous to that in the proof of Lemma 3 it is readily seen that $n^{-1/2}\mathbf{F}_n^{*\prime}\boldsymbol{\varepsilon}_n = O_p(1)$ and $n^{-1/2}\mathbf{F}_n^{**\prime}\boldsymbol{\varepsilon}_n = O_p(1)$. Consequently $n^{1/2}[\widehat{\boldsymbol{\delta}}_n(\hat{\rho}_n) - \boldsymbol{\delta}_n] = \mathbf{P}_n^{*\prime}n^{-1/2}\mathbf{F}_n^{*\prime}\boldsymbol{\varepsilon}_n + o_p(1)$ and $\mathbf{P}_n^{*\prime}n^{-1/2}\mathbf{F}_n^{*\prime}\boldsymbol{\varepsilon}_n = O_p(1)$, observing again that $\hat{\rho}_n - \rho_n = o_p(1)$. This completes the proof recalling that $\mathbf{T}_n^* = \mathbf{F}_n^*\mathbf{P}_n^*$. ∎

References

[1] Amemiya, T., The Estimation of the Variances in a Variance-Component Model. *International Economic Review,* 12 (1971), 1-13.

[2] Anselin, L., *Spatial Econometrics: Methods and Models.* Boston: Kluwer Academic Publishers, 1988.

[3] Anselin, L. and Florax, R., *New Directions in Spatial Econometrics.* London: Springer, 1995.

[4] Audretsch, D.B. and Feldmann, M.P., R&D Spillovers and the Geography of Innovation and Production. *American Economic Review* 86 (1996), 630-640.

[5] Baltagi, B.H., Egger, P. and Pfaffermayr, M. Estimating Models of Complex FDI: Are There Third-Country Effects? Department of Economics, Texas A&M University, 2005, forthcoming in *Journal of Econometrics.*

[6] Baltagi, B.H. and Li, D., Prediction in Panel Data Models with Spatial Correlation. In L. Anselin and R.J.G.M. Raymond, eds., *New Advances in Spatial Econometrics.* New York: Springer Verlag, 1999, forthcoming.

[7] Baltagi, B.H. and Li, D., Double Length Artificial Regressions for Testing Spatial Dependence. *Econometric Reviews* 20 (2001a), 31-40.

[8] Baltagi, B.H. and Li, D., LM Test for Functional Form and Spatial Error Correlation. *International Regional Science Review* 24 (2001b), 194-225.

[9] Baltagi, B.H., Song, S.H., Jung, B.C. and Koh, W., Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data. Department of Economics, Texas A&M University, 2004, forthcoming in *Journal of Econometrics.*

[10] Baltagi, B.H., Song, S.H. and Koh, W., Testing Panel Data Regression Models with Spatial Error Correlation, *Journal of Econometrics* 117 (2003), 123-150.

[11] Bao, Y., and Ullah, A., Finite Sample Properties of Maximum Likelihood Estimators in Spatial Models. Department of Economics, University of California, Riverside, 2003.

[12] Bell, K.P. and Bockstael, N.E., Applying the Generalized-Moments Estimation Approach to Spatial Problems Involving Microlevel Data. *Review of Economics and Statistics* 82 (2000), 72-82.

[13] Bennett, R. and Hordijk, L., Regional Econometric and Dynamic Models. In Nijkamp P., ed., *Handbook of Regional and Urban Economics*, vol. 1. Amsterdam: North Holland, 1986, 407-441.

[14] Betrand, M., Luttmer, E.F.P., and Mullainathan, S., Net-work Effects and Welfare Cultures. *Quarterly Journal of Economics* 115 (2000), 1019-1055.

[15] Besley, T. and Case, A., Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition. *American Economic Review* 85 (1995), 25-45.

[16] Bierens, H.J., *Topics in Advanced Econometrics.* Cambridge: Cambridge University Press, 1994.

[17] Case, A., Spatial Patterns in Household Demand. *Econometrica* 59 (1991), 953-966.

[18] Case, A., Hines Jr., J. and Rosen, H., Budget Spillovers and Fiscal Policy Independence: Evidence from the States. *Journal of Public Economics* 52 (1993), 285-307.

[19] Cliff, A. and Ord, J., *Spatial Autocorrelation,* London: Pion, 1973.

[20] Cliff, A. and Ord, J., *Spatial Processes, Models and Applications.* London: Pion, 1981.

[21] Cohen, J.P, and Morrison Paul, C.J., Public Infrastructure Investment, Interstate Spatial Spillovers, and Manufacturing Cost, *Review of Economic and Statistics* 86 (2004), 551-560.

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

[23] Cressie, N.A.C., *Statistics of Spatial Data.* New York: Wiley, 1993.

[24] Das, D., Kelejian, H.H. and Prucha, I.R., Small Sample Properties of Estimators of Spatial Autoregressive Models with Autoregressive Disturbances. *Papers in Regional Science* 82 (2003), 1-26.

[25] Dhrymes, P.J., *Introductory Econometrics.* New York: Springer Verlag, 1978.

[26] Driscoll, J. and Kraay, A., Consistent Covariance Matrix Estimation with Spatially Dependent Panel Data. *The Review of Economics and Statistics* LXXX (1998), 549-560.

[27] Fuller, W.A., *Introduction to Statistical Time Series.* New York: Wiley, 1976.

[28] Gänssler, P. and Stute, W., *Wahrscheinlichkeitstheorie.* New York: Springer Verlag, 1977.

[29] Hanushek, E.A., Kain, J.F., Markman, J.M., and Rivkin, S.G., Does Peer Ability Affect Student Achievement? *Journal of Applied Econometrics* 18 (2003), 527-544.

[30] Holtz-Eakin, D., Public Sector Capital and the Productivity Puzzle. *Review of Economics and Statistics* 76 (1994), 12-21.

[31] Horn, R.A. and Johnson, C.R. *Matrix Analysis.* Cambridge: Cambridge University Press, 1985.

[32] Kapoor, M., Kelejian, H.H., and Prucha, I.R., Panel Data Models with Spatially Correlated Error Components. Department of Economics, University of Maryland, 2004, forthcoming in *Journal of Econometrics.*

[33] Kelejian, H.H. and Prucha, I.R., 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 (1998), 99-121.

[34] Kelejian, H.H. and Prucha, I.R., A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review* 40 (1999), 509-533.

[35] Kelejian, H.H. and Prucha, I.R., On the Asymptotic Distribution of the Moran I Test Statistic with Applications. *Journal of Econometrics* 104 (2001), 219-257.

[36] Kelejian, H.H. and Prucha, I.R., 2SLS and OLS in a Spatial Autoregressive Model with Equal Spatial Weights. *Regional Science and Urban Economics* 32 (2002), 691-707.

[37] Kelejian, H.H. and Prucha, I.R., Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations. *Journal of Econometrics*, 118 (2004), 27-50.

[38] Kelejian, H.H. and Prucha, I.R., HAC Estimation in a Spatial Framework. Department of Economics, University of Maryland, 2005, forthcoming in *Journal of Econometrics.*

[39] Kelejian, H.H., Prucha, I.R. and Yuzefovich, E., Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results. In J. LeSage and K. Pace (eds.) *Advances in Econometrics: Spatial and Spatiotemporal Econometrics.* Elsevier, New York, 2004, 163-198.

[40] Korniotis, G.M., A Dynamic Panel Estimator with Both Fixed and Spatial Effects. Department of Finance, University of Notre Dame, 2005.

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

[42] Lee, L.F., Best Spatial Two-Stage Least Squares Estimators for a Spatial Autoregressive Model with Autoregressive Disturbances. *Econometric Reviews* 22 (2003), 307-335.

[43] Lee, L.F., Asymptotic Distributions of Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica* 72 (2004), 1899-1925.

[44] Lee, L.F., The Method of Elimination and Substitution in the GMM Estimation of Mixed Regressive, Spatial Autoregressive Models. Department of Economics, Ohio State University, 2005, forthcoming in *Journal of Econometrics*.

[45] Lee, L.F., GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models, *Journal of Econometrics*, 137 (2007), 489-514.

[46] Lin, X. and Lee, L.F., GMM Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity. Department of Economics, Ohio State University, November 2005.

[47] Pinkse, J. and Slade, M.E., Contracting in Space: An Application of Spatial Statistics to Discrete-Choice Models. *Journal of Econometrics* 85 (1998), 125-154.

[48] Pinske, J., Slade, M.E. and Brett, C., Spatial Price Competition: a Semiparametric Approach. *Econometrica* 70 (2002), 1111-1153.

[49] Pötscher, B.M. and Prucha, I.R., *Dynamic Nonlinear Econometric Models, Asymptotic Theory.* New York: Springer Verlag, 1997.

[50] Sacredote, B., Peer Effects with random Assignment: Results of Dartmouth Roommates. *Quarterly Journal of Economics* 116 (2001), 681-704.

[51] Shroder, M., Games the States Don't Play: Welfare Benefits and the Theory of Fiscal Federalism. *Review of Economics and Statistics* 77 (1995), 183-191.

[52] Topa, G., Social Interactions, Local Spillovers and Unemployment. *Review of Economic Studies* 68 (2001), 261-295.

[53] Whittle, P., On Stationary Processes in the Plane. *Biometrica* 41 (1954), 434-449.

[54] Zhenlin Yang, Z., Quasi-Maximum Likelihood Estimation for Spatial Panel Data Regressions. Paper presented at the Spatial Econometrics Workshop, Kiel Institute for World Economics, April 2005.