Econometric Reviews, 32(5–6):686–733, 2013
Copyright © Taylor & Francis Group, LLC
ISSN: 0747-4938 print/1532-4168 online
DOI: 10.1080/07474938.2013.741020

Downloaded by [University Of Maryland] at 10:34 28 October 2013

ON TWO-STEP ESTIMATION OF A SPATIAL AUTOREGRESSIVE MODEL WITH AUTOREGRESSIVE DISTURBANCES AND ENDOGENOUS REGRESSORS

David M. Drukker[1], **Peter Egger**[2], **and Ingmar R. Prucha**[3]

[1]*StataCorp, College Station, Texas, USA*
[2]*KOF, ETH Zurich, Zurich, Switzerland*
[3]*Department of Economics, University of Maryland, College Park, Maryland, USA*

□ *In this paper, we consider a spatial-autoregressive model with autoregressive disturbances, where we allow for endogenous regressors in addition to a spatial lag of the dependent variable. We suggest a two-step generalized method of moments (GMM) and instrumental variable (IV) estimation approach extending earlier work by, e.g., Kelejian and Prucha (1998, 1999). In contrast to those papers, we not only prove consistency for our GMM estimator for the spatial-autoregressive parameter in the disturbance process, but we also derive the joint limiting distribution for our GMM estimator and the IV estimator for the regression parameters. Thus the theory allows for a joint test of zero spatial interactions in the dependent variable, the exogenous variables and the disturbances. The paper also provides a Monte Carlo study to illustrate the performance of the estimator in small samples.*

Keywords Cliff–Ord spatial model; Generalized method of moments estimation; Limited information estimation; Two-stage least squares estimation.

JEL Classification C21; C31.

1. INTRODUCTION

Recent years have seen a rapidly growing number of theoretical[1] and applied econometric studies[2] which consider spatial interdependence

[1]Recent theoretical contributions include Baltagi and Li (2001a, 2001b, 2004) Baltagi et al. (2003); Blonigen et al. (2007), Bao and Ullah (2007), Conley (1999), Das et al. (2003), Kelejian and Prucha (1998, 1999, 2001, 2004, 2007, 2010), Kelejian et al. (2004), Kapoor et al. (2007), Lee (2002, 2003, 2004), LeSage (1997, 2000), Lin and Lee (2010), Lin and Lee (2010), Lin and Lee (2011), Pace and Barry (1997), Pinkse and Slade (1998), Pinkse et al. (2002), Rey and Boarnet (2004), and Yang (2010). Classic references concerning spatial models are Anselin (1988), Cliff and Ord (1973, 1981), Cressie (1993).

[2]Recent applied studies include Audretsch and Feldmann (1996), Baltagi et al. (2008), Bell and Bockstael (2000), Besley and Case (1995), Blonigen et al. (2007), Bollinger and Ihlanfeldt (1997), Case (1991), Coughlin and Segev (2000), Case et al. (1993), Dowd and LeSage (1997), Holtz-Eakin (1994), LeSage (1999), Kelejian and Robinson (1997, 2000), Pinkse and Slade (1998), Pinkse et al. (2002), and Shroder (1995).

Address correspondence to Ingmar R. Prucha, Department of Economics, University of Maryland, College Park, MD 20742, USA; E-mail: prucha@econ.umd.edu

On Two-step Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances and Endogenous Regressors

Forthcoming in *Econometric Reviews*

David M. Drukker[1], Peter Egger[2], and Ingmar R. Prucha [3]

April 11, 2012

[1]StataCorp, 4905 Lakeway Drive, College Station, TX 77845, Tel.:+1- 979-696-4600, e-mail: ddrukker@stata.com

[2]KOF, ETH Zurich, WEH E6, Weinbergstrasse 35, 8092 Zurich, Switzerland; Tel.: +41-44-632-4108, e-mail: egger@kof.ethz.ch

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

Abstract

In this paper, we consider a spatial-autoregressive model with autoregressive disturbances, where we allow for endogenous regressors in addition to a spatial lag of the dependent variable. We suggest a two-step generalized method of moments (GMM) and instrumental variable (IV) estimation approach extending earlier work by, e.g., Kelejian and Prucha (1998, 1999). In contrast to those papers, we not only prove consistency for our GMM estimator for the spatial-autoregressive parameter in the disturbance process, but we also derive the joint limiting distribution for our GMM estimator and the IV estimator for the regression parameters. Thus the theory allows for a joint test of zero spatial interactions in the dependent variable, the exogenous variables and the disturbances. The paper also provides a Monte Carlo study to illustrate the performance of the estimator in small samples.[1]

Key Words: Cliff-Ord spatial model; Limited information estimation; Two-stage least squares estimation; Generalized method of moments estimation

JEL Classification: C21, C31

[1]We gratefully acknowledge financial support from the National Institute of Health through the SBIR grant R43 AG027622 and R44 AG027622. We also thank three anonymous referees, the special-issue editor Badi Baltagi, and the editor Esfandiar Maasoumi, for helpful comments and support.

1 Introduction

Recent years have seen a rapidly growing number of theoretical[2] and applied econometric studies[3] which consider spatial interdependence among the cross-sectional units of observation. The original form of the spatial model only considers spatial spillovers in the dependent variable, and employs an endogenous weighted average of dependent variables corresponding to other cross-sectional units on the right-hand side (RHS). This model is commonly referred to as a *spatial-autoregressive model* or SAR (see Cliff and Ord, 1973, 1981, for early examples), the weighted average is dubbed the spatial lag, the corresponding parameter is known as the autoregressive parameter, and the matrix containing the weights as the spatial-weights matrix. Generalized versions of this model also allow for the dependent variable to depend on a set of exogenous variables and spatial lags thereof and, in particular, allow for the disturbances to be generated by a spatial-autoregressive process. The combined *spatial-autoregressive model with (spatial) autoregressive residuals* is often referred to as SARAR (see Anselin and Florax, 1995).

Estimation theory developed for the SARAR model typically assumed that – except for the spatial lag – the regressors are strictly exogenous; see, e.g., Kelejian and Prucha (1998, 1999, 2010) and Lee (2003, 2004, 2007). This may be problematic in the context of many empirical applications. One objective of our study is to relax the assumption of exogenous regressors in a single-equation SARAR framework with homoskedastic innovations. Our work builds on Kelejian and Prucha (2004), who developed a GMM/IV estimation framework for systems of linear equations. However, that paper only derives the limiting distribution of the IV estimators of the regression parameters, but not that of the GMM estimators of the spatial-autoregressive parameters of the disturbance process, although the paper shows that the latter estimators are consistent. Consequently the results in that paper do not allow for the testing of the joint hypothesis of the absence of spatial spillovers in the dependent variables, the exogenous variables and the disturbance process, which may be of interest in empirical work.

In light of the above, the other objective of our study is thus to derive the joint limiting distribution of the IV estimators of the regression parameters and of the GMM estimators of the spatial-autoregressive parameters of the disturbance process, and to provide consistent estimators for the joint asymptotic variance-covariance matrix. Our analysis focuses on two step estimators, because it is computationally convenient. As a by-product, the paper also develops the joint asymptotic distribution of the GMM/IV estimators considered in Kelejian and Prucha (1998, 1999), thus closing an existing lacuna in the estimation theory for those estimators. One reason for this lacuna was that originally the authors lacked a useful central

[2]Recent theoretical contributions include Baltagi and Li (2001a, b, 2004), Baltagi et al. (2003, 2007), Bao and Ullah (2007), Conley (1999), Das et al. (2003), Kelejian and Prucha (1998, 1999, 2001, 2004, 2007, 2010), Kapoor et al. (2007), Lee (2002, 2003, 2004), LeSage (1997, 2000), Lee and Liu (2010), Lin and Lee (2010), Liu and Lee (2011), Pace and Barry (1997), Pinkse and Slade (1998), Pinkse et al. (2002), and Rey and Boarnet (2004). Classic references concerning spatial models are Anselin (1988), Cliff and Ord (1973, 1981), and Cressie (1993).

[3]Recent applied studies include Audretsch and Feldmann (1996), Bell and Bockstael (2000), Besley and Case (1995), Blonigen et al. (2007), Bollinger and Ihlanfeldt (1997), Case (1991), Coughlin and Segev (2000), Case et al. (1993), Dowd and LeSage (1997), Holtz-Eakin (1994), LeSage (1999), Kelejian and Robinson (1997, 2000), Pinkse and Slade (1998), Pinkse et al. (2002), and Shroder (1995).

limit theorem (CLT) for quadratic forms. Such a CLT was only developed in Kelejian and Prucha (2001). Also, in contrast to Kelejian and Prucha (2004), the endogenous regressor variables are not assumed to be generated by a linear system.[4]

The remainder of the paper is organized as follows. Section 2 specifies the SARAR model with "outside" endogenous variables. In Section 3 we define and establish the large sample properties of suggested GMM estimators for the spatial-autoregressive parameter of the disturbance process and IV estimator for the regression parameters. Moreover, in this section we derive the joint large sample distribution of the GMM and IV estimators. In Section 4 we analyze the small sample behavior of our suggested estimator and test statistics via a small Monte Carlo study. Concluding remarks are given in Section 5. All technical details are relegated to the appendices.

2 Model Specification

In this section, we specify a generalized spatial-autoregressive model with autoregressive disturbances, and discuss the underlying assumptions. The specification is fairly general in that it allows for some of the RHS variables to be endogenous in a general, unspecified form. More specifically, we consider the following Cliff-Ord type spatial model relating a cross section of n spatial units:

$$\begin{aligned}\mathbf{y}_n &= \mathbf{X}_n\beta_{0n} + \mathbf{Y}_n\pi_{0n} + \lambda_{0n}\mathbf{W}_n\mathbf{y}_n + \mathbf{u}_n \qquad (1)\\ &= \mathbf{Z}_n\delta_{0n} + \mathbf{u}_n\end{aligned}$$

and

$$\mathbf{u}_n = \rho_{0n}\mathbf{M}_n\mathbf{u}_n + \boldsymbol{\varepsilon}_n, \qquad (2)$$

where $\mathbf{Z}_n = [\mathbf{X}_n, \mathbf{Y}_n, \mathbf{W}_n\mathbf{y}_n]$, and $\delta_{0n} = [\beta_{0n}', \pi_{0n}', \lambda_{0n}]'$. Here $\mathbf{y}_n$ is the $n \times 1$ vector of the dependent variable, $\mathbf{X}_n$ is the $n \times k$ matrix of the non-stochastic exogenous regressors, $\mathbf{Y}_n$ is an $n \times r$ matrix of endogenous regressors, $\mathbf{W}_n$ and $\mathbf{M}_n$ are $n \times n$ observed non-stochastic weights matrices, $\mathbf{u}_n$ is the $n \times 1$ vector of regression disturbances, and $\boldsymbol{\varepsilon}_n$ is an $n \times 1$ vector of innovations. The vectors $\mathbf{W}_n\mathbf{y}_n$ and $\mathbf{M}_n\mathbf{u}_n$ represent spatial lags, the scalars λ_{0n} and ρ_{0n} denote the corresponding true parameters, typically referred to as spatial-autoregressive parameters, and β_{0n} and π_{0n} are $k \times 1$ and $r \times 1$ true parameter vectors. As indicated by the indexation, our specification allows for the elements of all data vectors and matrices, as well as for all parameters to depend on the sample size, i.e., to form triangular arrays. Among other things, the specification thus accommodates formulations, as is frequently the case in applications, where the spatial-weights matrices are normalized. In allowing also for the parameters to

[4] Kelejian and Prucha (2010) derive the joint asymptotic distribution for certain GMM/IV estimators for a model with only exogenous regressors (except for a spatial lag) under the assumption of heteroskedastic innovations. However, even in case all regressors are exogenous the results of that paper do not cover the class of estimators considered in this paper in that we allow for a more general set of moment conditions. In a recent paper Liu and Lee (2011) also allow for endogenous regressors. In contrast to that paper we also allow for spatially correlated disturbances and focus on two-step estimation.

depend on n we can then assume a common parameter space for all sample sizes; see Kelejian and Prucha (2010) for a more detailed discussion.

As discussed in the introduction, much of the existing literature on spatial Cliff-Ord models assumes that, except for spatial lags, all of the RHS variables are exogenous. This is a strong assumption that may not hold in many applications. The above specification explicitly accommodates endogenous regressors in addition to the spatial lag $\mathbf{W}_n\mathbf{y}_n$. Given the inclusion of $\mathbf{Y}_n$ on the RHS of (1), the above model may be viewed as representing a single equation of a system of equations. Our assumptions will be such that (1) could be a single equation of the linear simultaneous equation system considered in Kelejian and Prucha (2004).[5] However for generality we will maintain a high-level assumption regarding $\mathbf{Y}_n$ that allows for (1) to be part of a more general system, e.g., a nonlinear system of equations.

In the above specification, spatial spillovers in the dependent variable and the disturbances are modeled explicitly through the spatial lags $\mathbf{W}_n\mathbf{y}_n$ and $\mathbf{M}_n\mathbf{u}_n$, respectively. However, given that our specification allows for the elements of all data vectors and matrices to depend on the sample size, some or all of the column vectors of $\mathbf{X}_n$ and $\mathbf{Y}_n$ may also represent spatial lags of some underlying exogenous and endogenous variables. Consequently the specification is fairly general.[6]

We next give a detailed list and discussion of the maintained assumptions. In particular, 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)* $\sup_n |\lambda_{0n}| < 1$, $\sup_n |\rho_{0n}| < 1$. *(c) The matrices* $\mathbf{I}_n - \lambda_{0n}\mathbf{W}_n$ *and* $\mathbf{I}_n - \rho_{0n}\mathbf{M}_n$ *are nonsingular for all* $\lambda_{0n} \in (-1, 1)$, *and* $\rho_{0n} \in (-1, 1)$.

Assumption 1(a) is a normalization rule. Assumption 1(b) determines the parameter space of λ_{0n} and ρ_{0n}. This assumption is not restrictive in that we can always re-normalize the weights matrices such that the parameter space for the autoregressive parameters is the interval $(-1, 1)$; see Kelejian and Prucha (2009) for a more detailed discussion. Assumption 1(c) ensures that $\mathbf{y}_n$ and $\mathbf{u}_n$ are uniquely defined by (1) and (2) as

$$\begin{aligned} \mathbf{y}_n &= (\mathbf{I}_n - \lambda_{0n}\mathbf{W}_n)^{-1}\mathbf{Z}_n\delta_{0n} + (\mathbf{I}_n - \lambda_{0n}\mathbf{W}_n)^{-1}\mathbf{u}_n, \\ \mathbf{u}_n &= (\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}\boldsymbol{\varepsilon}_n. \end{aligned} \quad (3)$$

[5] As noted above, Kelejian and Prucha (2004) do not derive the limiting distribution of their estimator for ρ_{0n}, which is one of the objectives of this paper.

[6] Badinger and Egger (2011) and Lee and Liu (2010) consider models with higher order spatial lags in the dependent variable and disturbances, without allowing for general endogenous RHS regressors. We note that in contrast to Lee and Liu (2010) the focus of this paper is on two-step estimation. While computationally simple, it turns out that the technical derivation of the limiting distribution of those estimators is more involved. Liu and Lee (2011) consider a spatial-autoregressive model with endogenous regressors. However this paper does not allow for spatial correlation in the disturbances and focuses on an analysis with many instruments.

We maintain the following set of assumptions with respect to the innovations $\boldsymbol{\varepsilon}_n$.

Assumption 2 *The innovations $\{\varepsilon_{i,n} : 1 \leq i \leq n, n \geq 1\}$ are i.i.d. for each n with $E\varepsilon_{i,n} = 0$, $E(\varepsilon_{i,n}^2) = \sigma^2 > 0$ and $E|\varepsilon_{i,n}|^{4+\eta} < \infty$ for some $\eta > 0$.*

The above assumption imposes homoskedastic innovations. For simplicity of presentation we assume that the moments do not depend on n, but note that this assumption could be readily relaxed. We maintain the following assumption concerning the spatial-weights matrices.

Assumption 3 *The row and column sums of the matrices $\mathbf{W}_n$ and $\mathbf{M}_n$ are bounded uniformly in absolute value by, respectively, one and some finite constant, and the row and column sums of the matrices $(\mathbf{I}_n - \lambda_{0n}\mathbf{W}_n)^{-1}$ and $(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}$ are bounded uniformly in absolute value by some finite constant, and the smallest eigenvalues of $(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n')^{-1}$ are bounded away from zero.*

Given (3), Assumption 2 implies that $E\mathbf{u}_n = 0$, and that the variance-covariance (VC) matrix of $\mathbf{u}_n$ is determined as

$$E\mathbf{u}_n\mathbf{u}_n' = \sigma_{0n}^2(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n')^{-1}.$$

Assumptions 2 and 3 imply that the row and column sums of the 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 $\mathbf{y}_n$).[7]

All estimators considered in the paper will correspond to a set of linear quadratic moment conditions of the form:

$$\begin{aligned} E\boldsymbol{\varepsilon}_n'\mathbf{A}_{sn}\boldsymbol{\varepsilon}_n &= 0,\ s = 1, ..., S, \\ E\mathbf{H}_n'\boldsymbol{\varepsilon}_n &= \mathbf{0}, \end{aligned} \tag{4}$$

where the $n \times n$ weighting matrices $\mathbf{A}_{sn}$ in the quadratic forms, and the $n \times p$ instrument matrices $\mathbf{H}_n$ in the linear forms are non-stochastic, with $p \geq k + r + 1$. Specifications for $\mathbf{A}_{sn}$ and recommendations for $\mathbf{H}_n$ will be given below. For initial estimators we will also consider moment conditions of the form $E\mathbf{H}_n'\mathbf{u}_n = \mathbf{0}$.

Examples of sets of weighting matrices for quadratic moment conditions are

$$\begin{aligned} \mathbf{A}_{1,n} &= v_n\left[\mathbf{M}_n'\mathbf{M}_n - n^{-1}tr(\mathbf{M}_n'\mathbf{M}_n)\mathbf{I}_n\right], \\ v_n &= 1/\left[1 + \left[n^{-1}tr(\mathbf{M}_n'\mathbf{M}_n)\right]^2\right], \\ \mathbf{A}_{2,n} &= \mathbf{M}_n. \end{aligned}$$

[7] Towards providing more insight as to how, say, the assumption that the row sums of the absolute elements of $(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}$ are bounded can be implied, let $\|.\|$ denote the maximum row sum norm. Suppose the row sums are uniformly bounded by one in absolute value, i.e., $\|\mathbf{M}_n\| \leq 1$, then given Assumption 1 we have $\left\|(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}\right\| \leq \sum_{k=1}^{\infty}|\rho_{0n}|^k\|\mathbf{M}_n\|^k \leq 1/[1 - \sup_n|\rho_{0n}|] < \infty$.

and

$$\mathbf{A}_{1,n} = \mathbf{M}_n'\mathbf{M}_n - diag(\mathbf{M}_n'\mathbf{M}_n),$$
$$\mathbf{A}_{2,n} = \mathbf{M}_n.$$

The former are, e.g., used in Kelejian and Prucha (1998, 1999, 2004) and Lee (2003), the latter, e.g., in Kelejian and Prucha (2010).

We maintain the following assumptions concerning the matrices $\mathbf{A}_{sn}$, $\mathbf{X}_n$ and $\mathbf{H}_n$:

Assumption 4 *: The row and column sums of the matrices* $\mathbf{A}_{sn}$*,* $s = 1, ..., S$*, are bounded uniformly in absolute value by some finite constant. Furthermore* $tr(\mathbf{A}_{sn}) = 0$ *for all* $s = 1, ..., S$*. For* $s = 1, ..., S_\star$ *with* $0 \leq S_* \leq S$ *the diagonal elements of the matrices* $\mathbf{A}_{sn}$ *are assumed to be zero.*

Notice that – unlike Kelejian and Prucha (2010) and Arraiz et al. (2010) – we do not require the diagonal elements of $\mathbf{A}_{sn}$ to be zero. However, our approach allows a subset of $\mathbf{A}_{sn}$ to exhibit zero diagonal elements. The adopted ordering of the matrices $\mathbf{A}_{sn}$ in the above assumption is w.l.o.g.

Additionally we maintain the following assumptions regarding the exogenous and endogenous regressors and the instruments.

Assumption 5 *: The exogenous regressors* $\mathbf{X}_n$ *are non-stochastic, and the elements of the matrices* $\mathbf{X}_n$ *are uniformly bounded in absolute value (by some finite constant). The endogenous regressors* $\mathbf{Y}_n$ *have finite* $2 + \delta$ *absolute moments for some* $\delta > 0$*, that are uniformly bounded in absolute value (by some finite constant). Furthermore,* $\mathbf{Y}_n$ *is such that for any* $n \times n$ *real matrix* $\mathbf{A}_n$ *whose row and column sums are bounded uniformly in absolute value*

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

where the variables $\mathbf{Z}_n$ *are defined above after equation (2).*

In appendix D we show that this assumption is satisfied if $\mathbf{Y}_n$ is generated from a system as considered in Kelejian and Prucha (2004).

Assumption 6 *: The instrument matrices* $\mathbf{H}_n$ *are nonstochastic and have full column rank* $p \geq k + r + 1$ *(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 contain, at least, the linearly independent columns of* $(\mathbf{X}_n, \mathbf{M}_n\mathbf{X}_n)$*.*

In treating $\mathbf{X}_n$ and $\mathbf{H}_n$ as non-stochastic our analysis should be viewed as conditional on $\mathbf{X}_n$ and $\mathbf{H}_n$. Also note that moment conditions (4) clearly hold under the above set of assumptions. In the following $\lambda_{\min}(.)$ denotes the smallest eigenvalue of a matrix.

Assumption 7 *: 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) $\mathbf{Q}_{HZ} = \text{plim}_{n\to\infty} n^{-1}\mathbf{H}_n'\mathbf{Z}_n$ *and* $\mathbf{Q}_{HMZ} = \text{plim}_{n\to\infty} n^{-1}\mathbf{H}_n'\mathbf{M}_n\mathbf{Z}_n$ *are finite and have full column rank.*

(c) *Let* $\mathbf{Q}_{HZ*}(\rho_{0n}) = \mathbf{Q}_{HZ} - \rho_{0n}\mathbf{Q}_{HMZ}$*, then* $\lambda_{\min}\left\{\left[\mathbf{Q}_{HZ*}(\rho_{0n})'\mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ*}(\rho_{0n})\right]^{-1}\right\} \geq c$ *for some* $c > 0$*.*

The above assumptions are similar to those maintained in Kelejian and Prucha (1998, 2004, 2010), and Lee (2003), and so a discussion which is quite similar to the one given in those papers also applies here.

As to the selection of instruments, for the case where there are no "outside" RHS endogenous variables present, Kelejian and Prucha (1999) suggested for $\mathbf{H}_n$ to be a subset of the linearly independent columns of

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

where q is a pre-selected finite constant. The motivation for this recommendation was to achieve a computationally simple approximation of the ideal instruments, which are given in terms of the conditional means of the RHS variables.[8] Within our setting, since the system determining $\mathbf{y}_n$ and $\mathbf{Y}_n$ is not completely specified, the ideal instruments are not known. A reasonable suggestion may be to use a set of instruments as above with $\mathbf{X}_n$ augmented by other exogenous variables expected to be part of the reduced form of the system.

Our discussions will also utilize the following spatial Cochrane-Orcutt transformation of (1) and (2):

$$\mathbf{y}_{n*}(\rho_{0n}) = \mathbf{Z}_{n*}(\rho_{0n})\delta_{0n} + \boldsymbol{\varepsilon}_n, \tag{5}$$

where $\mathbf{y}_{n*}(\rho_{0n}) = \mathbf{y}_n - \rho_{0n}\mathbf{M}_n\mathbf{y}_n$ and $\mathbf{Z}_{n*}(\rho_{0n}) = \mathbf{Z}_n - \rho_{0n}\mathbf{M}_n\mathbf{Z}_n$. The transformed model is readily obtained by pre-multiplying (1) by $\mathbf{I}_n - \rho_{0n}\mathbf{M}_n$.

[8] We note that the inclusion of instruments involving $\mathbf{M}_n$ in the instrument matrix $\mathbf{H}_n$ is only need needed for the formulation of IV estimators applied to the spatially Cochrane-Orcutt transformed model (5) given below. However, for simplicity of exposition we work with only one instrument matrix.

3 Two-Step IV/GMM Estimators

In this section, we specify IV/GMM estimators for the model parameters δ_{0n} and ρ_{0n}. Our GMM estimators of ρ_{0n} are based on $n^{1/2}$-consistent residuals obtained in a prior step. As it turns out, the asymptotic distribution of the estimator for ρ_{0n} will generally depend on the estimator for δ_{0n} employed in computing estimates of the disturbances. In the following we will give results on the joint limiting distribution of second-step GMM estimators for δ_{0n} and ρ_{0n}, say $\widehat{\delta}_{0n}$ and $\widehat{\rho}_{0n}$, where $\widehat{\rho}_{0n}$ is based on estimated disturbances which depend on $\widehat{\delta}_{0n}$. We note that in the appendix we give generic results concerning the consistency and joint asymptotic normality of two-step IV/GMM estimators that are sufficiently general so that they can be used readily to obtain the joint limiting distribution for alternative combinations of IV/GMM estimators for δ_{0n} and ρ_{0n}. To conserve space, we only provide results for the final-step estimators.

Each step of the specific two-step IV/GMM estimators defined below consists of sub-steps involving the estimation of ρ_{0n} and δ_{0n}. In step 1, estimates are computed from the original model (1). Those estimates are used in step 2 to compute estimates from the transformed model (3), with ρ_{0n} replaced by an estimator.

3.1 Quadratic Moment Conditions

Define the vector of identically and independently distributed disturbances

$$\boldsymbol{\varepsilon}_n(\theta) = (\mathbf{I}_n - \rho \mathbf{M}_n)\left[\mathbf{y}_n - \mathbf{Z}_n \delta\right] \tag{6}$$

with $\theta = (\delta', \rho)'$ and $\delta = (\beta', \pi', \lambda)'$. For further interpretation let $\mathbf{u}_n(\delta) = \mathbf{y}_n - \mathbf{Z}_n\delta$, and $\overline{\mathbf{u}}_n(\delta) = \mathbf{M}_n\mathbf{u}_n(\delta)$, and observe that $\boldsymbol{\varepsilon}_n(\theta)$ can also be expressed as

$$\boldsymbol{\varepsilon}_n(\theta) = \mathbf{u}_n(\delta) - \rho\overline{\mathbf{u}}_n(\delta) = \mathbf{y}_{n*}(\rho) - \mathbf{Z}_{n*}(\rho)\delta. \tag{7}$$

In the following we assume that our estimators utilize the following set of quadratic moment functions:

$$\mathbf{m}_n(\theta) = n^{-1}\begin{bmatrix} \boldsymbol{\varepsilon}_n(\theta)'\mathbf{A}_{1n}\boldsymbol{\varepsilon}_n(\theta) \\ \vdots \\ \boldsymbol{\varepsilon}_n(\theta)'\mathbf{A}_{Sn}\boldsymbol{\varepsilon}_n(\theta) \end{bmatrix} = n^{-1}\begin{bmatrix} (\mathbf{u}_n(\delta) - \rho\overline{\mathbf{u}}_n(\delta))'\,\mathbf{A}_{1n}\,(\mathbf{u}_n(\delta) - \rho\overline{\mathbf{u}}_n(\delta)) \\ \vdots \\ (\mathbf{u}_n(\delta) - \rho\overline{\mathbf{u}}_n(\delta))'\,\mathbf{A}_{Sn}\,(\mathbf{u}_n(\delta) - \rho\overline{\mathbf{u}}_n(\delta)) \end{bmatrix}. \tag{8}$$

Clearly, $E\mathbf{m}_n(\theta_{0n}) = \mathbf{0}$ under the maintained assumptions because $\boldsymbol{\varepsilon}_n = \boldsymbol{\varepsilon}_n(\theta_{0n})$, with $\theta_{0n} = (\delta_{0n}', \rho_{0n})'$. It proves convenient and it is instructive to re-write the moment conditions as

$$\gamma_n - \boldsymbol{\Gamma}_n \begin{bmatrix} \rho_{0n} \\ \rho_{0n}^2 \end{bmatrix} = \mathbf{0} \tag{9}$$

where

$$\gamma_n = n^{-1}\begin{bmatrix} E\mathbf{u}_n'\mathbf{A}_{1n}\mathbf{u}_n \\ \vdots \\ E\mathbf{u}_n'\mathbf{A}_{Sn}\mathbf{u}_n \end{bmatrix}, \boldsymbol{\Gamma}_n = n^{-1}\begin{bmatrix} 2E\mathbf{u}_n'\mathbf{M}_n'\mathbf{A}_{1n}\mathbf{u}_n & -E\mathbf{u}_n'\mathbf{M}_n'\mathbf{A}_{1n}\mathbf{M}_nE\mathbf{u}_n \\ \vdots & \vdots \\ 2E\mathbf{u}_n'\mathbf{M}_n'\mathbf{A}_{Sn}E\mathbf{u}_n & -E\mathbf{u}_n'\mathbf{M}_n'\mathbf{A}_{Sn}\mathbf{M}_nE\mathbf{u}_n \end{bmatrix}.$$

We maintain the following assumption, which is in essence an identification condition for ρ_{0n}.

Assumption 8 *The smallest eigenvalue of $\mathbf{\Gamma}_n^{'}\mathbf{\Gamma}_n$ is uniformly bounded away from zero.*

3.2 Definition of Two-step IV/GMM Estimators

In the following we describe the steps to compute our suggested IV/GMM estimators.

Step 1a: 2SLS Estimator
In the first step, we apply 2SLS to model (1) using the instrument matrix $\mathbf{H}_n$ in Assumption 5 to estimate δ. The 2SLS estimator, say $\widetilde{\delta}_n$, is then defined as

$$\widetilde{\delta}_n = (\widetilde{\mathbf{Z}}_n'\mathbf{Z}_n)^{-1}\widetilde{\mathbf{Z}}_n'\mathbf{y}_n, \tag{10}$$

where $\widetilde{\mathbf{Z}}_n = \mathbf{P}_{\mathbf{H}_n}\mathbf{Z}_n = (\mathbf{X}_n, \widetilde{\mathbf{Y}}_n, \widetilde{\mathbf{W}_n\mathbf{y}_n})$, $\widetilde{\mathbf{Y}}_n = \mathbf{P}_{\mathbf{H}_n}\mathbf{Y}_n$, $\widetilde{\mathbf{W}_n\mathbf{y}_n} = \mathbf{P}_{\mathbf{H}_n}\mathbf{W}_n\mathbf{y}_n$, and where $\mathbf{P}_{\mathbf{H}_n} = \mathbf{H}_n(\mathbf{H}_n'\mathbf{H}_n)^{-1}\mathbf{H}_n'$.

Step 1b: Initial GMM Estimator of ρ Based on 2SLS Residuals
Let $\widetilde{\mathbf{u}}_n = \mathbf{u}_n(\widetilde{\delta}_n) = \mathbf{y}_n - \mathbf{Z}_n\widetilde{\delta}_n$ denote the 2SLS residuals, and define $\widetilde{\overline{\mathbf{u}}}_n = \mathbf{M}_n\widetilde{\mathbf{u}}_n$ and $\widetilde{\overline{\overline{\mathbf{u}}}}_n = \mathbf{M}_n^2\widetilde{\mathbf{u}}_n$. Consider the following sample moments corresponding to (8) based on estimated 2SLS residuals:

$$\mathbf{m}_n(\widetilde{\delta}_n, \rho) = n^{-1}\begin{bmatrix} (\widetilde{\mathbf{u}}_n - \rho\widetilde{\overline{\mathbf{u}}}_n)'\mathbf{A}_{1n}(\widetilde{\mathbf{u}}_n - \rho\widetilde{\overline{\mathbf{u}}}_n) \\ \vdots \\ (\widetilde{\mathbf{u}}_n - \rho\widetilde{\overline{\mathbf{u}}}_n)'\mathbf{A}_{Sn}(\widetilde{\mathbf{u}}_n - \rho\widetilde{\overline{\mathbf{u}}}_n) \end{bmatrix}. \tag{11}$$

Our initial GMM estimator for ρ is now defined as

$$\widetilde{\rho}_n = \underset{\rho\in[-a^\rho, a^\rho]}{argmin}\left[\mathbf{m}_n(\widetilde{\delta}_n, \rho)'\mathbf{m}_n(\widetilde{\delta}_n, \rho)\right], \tag{12}$$

where $a^\rho \geq 1$.

We note that moment conditions used in step 1b could differ from those in step 2b below, as long as the estimator $\widetilde{\rho}_n$ is consistent. For example, users may simply employ the matrices $\mathbf{A}_{1n}$ and $\mathbf{A}_{2n}$ which leads to the moment conditions used in Kelejian and Prucha (1999), and which have been seen to obtain reasonably good estimates, even in small samples.

Step 2a: GS2SLS Estimator
Analogous to Kelejian and Prucha (1998), we next compute a generalized spatial two-stage least-squares (GS2SLS) estimator of δ. This estimator is defined as the 2SLS estimator of the

Cochrane-Orcutt transformed model (3) with the parameter ρ replaced by $\widetilde{\rho}_n$ computed in step 1b:

$$\widehat{\delta}_n(\widetilde{\rho}_n) = [\widehat{\mathbf{Z}}_{n*}(\widetilde{\rho}_n)'\mathbf{Z}_{n*}(\widetilde{\rho}_n)]^{-1}\widehat{\mathbf{Z}}_{n*}(\widetilde{\rho}_n)'\mathbf{y}_{n*}(\widetilde{\rho}_n) \tag{13}$$

where $\mathbf{y}_{n*}(\widetilde{\rho}_n) = (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n)\mathbf{y}_n$, $\mathbf{Z}_{n*}(\widetilde{\rho}_n) = (\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n)\mathbf{Z}_n$, $\widehat{\mathbf{Z}}_{n*}(\widetilde{\rho}_n) = \mathbf{P}_{\mathbf{H}_n}\mathbf{Z}_{n*}(\widetilde{\rho}_n)$, and $\mathbf{P}_{\mathbf{H}_n} = \mathbf{H}_n(\mathbf{H}_n'\mathbf{H}_n)^{-1}\mathbf{H}_n'$.

Step 2b: Efficient GMM Estimator of ρ Based on GS2SLS Residuals

Let $\widehat{\mathbf{u}}_n = \mathbf{y}_n - \mathbf{Z}_n\widehat{\delta}_n(\widetilde{\rho}_n)$ denote the GS2SLS residuals, and define $\widehat{\overline{\mathbf{u}}}_n = \mathbf{M}_n\widehat{\mathbf{u}}_n$ and $\widehat{\overline{\overline{\mathbf{u}}}}_n = \mathbf{M}_n^2\widehat{\mathbf{u}}_n$. Let $\mathbf{m}(\widehat{\delta}_n, \rho)$ denote the sample moment vector obtained by replacing in (11) the 2SLS residuals by the GS2SLS residuals $\widehat{\mathbf{u}}_n$, $\widehat{\overline{\mathbf{u}}}_n$, and $\widehat{\overline{\overline{\mathbf{u}}}}_n$. The corresponding efficient GMM estimator for ρ based on GS2SLS residuals is then given by

$$\widehat{\rho}_n = \underset{\rho\in[-a^\rho,a^\rho]}{argmin}\left[\mathbf{m}(\widehat{\delta}_n, \rho)'\left(\widehat{\mathbf{\Psi}}_n^{\rho\rho}(\widehat{\delta}_n, \widetilde{\rho}_n)\right)^{-1}\mathbf{m}(\widehat{\delta}_n, \rho)\right], \tag{14}$$

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

$$\begin{aligned}\widehat{\psi}_{rs,n}^{\rho\rho}(\widehat{\delta}_n, \widetilde{\rho}_n) = {}& \widehat{\sigma}_n^4(2n)^{-1}tr\left[\left(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}'\right)\left(\mathbf{A}_{s,n} + \mathbf{A}_{s,n}'\right)\right] \\ &+ \widehat{\sigma}_n^2 n^{-1}\widehat{\mathbf{a}}_{r,n}'\widehat{\mathbf{a}}_{s,n} \\ &+ n^{-1}(\widehat{\mu}_n^{(4)} - 3\widehat{\sigma}_n^4)vec_D(\mathbf{A}_{r,n})'vec_D(\mathbf{A}_{s,n}) \\ &+ n^{-1}\widehat{\mu}_n^{(3)}\left[\widehat{\mathbf{a}}_{r,n}'vec_D(\mathbf{A}_{s,n}) + \widehat{\mathbf{a}}_{s,n}'vec_D(\mathbf{A}_{r,n})\right],\end{aligned} \tag{15}$$

where $\widehat{\mathbf{a}}_{r,n} = \widehat{\mathbf{T}}_n\widehat{\alpha}_{r,n}$ and

$$\begin{aligned}\widehat{\mathbf{T}}_n &= \widehat{\mathbf{T}}_n(\widehat{\delta}_n, \widetilde{\rho}_n) = \mathbf{H}_n\widehat{\mathbf{P}}_n^* \\ \widehat{\mathbf{P}}_n^* &= \widehat{\mathbf{P}}_n^*(\widetilde{\rho}_n) = (n^{-1}\mathbf{H}_n'\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}_n'\mathbf{Z}_{n*}(\widetilde{\rho}_n))\times \\ &\quad \left[(n^{-1}\mathbf{Z}_{n*}'(\widetilde{\rho}_n)\mathbf{H}_n)(n^{-1}\mathbf{H}_n'\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}_n'\mathbf{Z}_{n*}(\widetilde{\rho}_n))\right]^{-1} \\ \widehat{\alpha}_{r,n} &= \widehat{\alpha}_{r,n}(\widehat{\delta}_n, \widetilde{\rho}_n) = -n^{-1}\left[\mathbf{Z}_n'\left(\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n'\right)\left(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}'\right)\left(\mathbf{I}_n - \widetilde{\rho}_n\mathbf{M}_n\right)\widehat{\mathbf{u}}_n\right]\end{aligned}$$

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

For $s = 1, ..., S_\star$ $(0 \leq S_* \leq S)$, the terms involving the third and fourth moments are zero in the above expressions for $\widehat{\psi}_{rs,n}^{\rho\rho}(\widehat{\delta}_n, \widetilde{\rho}_n)$ because we maintained in Assumption 4 that, the diagonal elements of the matrices $\mathbf{A}_{sn}$ are zero for $s = 1, ..., S_\star$ $(0 \leq S_* \leq S)$. If $S_* = S$, then all terms involving the third and fourth moments no longer appear in the above expressions for $\widehat{\psi}_{rs,n}^{\rho\rho}(\widehat{\delta}_n, \widetilde{\rho}_n)$.

3.3 Asymptotic Properties of Two-step IV/GMM Estimators

In the following we give results concerning the joint limiting distribution of the final stage estimators $\widehat{\rho}_n$ and $\widehat{\delta}_n$. As remarked above, in the appendix we also give generic results concerning the consistency and joint asymptotic normality of two-step IV/GMM estimators that are sufficiently general so that they can be used readily to obtain the joint limiting distribution for alternative combinations of IV/GMM estimators for δ_{0n} and ρ_{0n}.

In preparation of the next theorem we first give, without proof, an expression for the variance-covariance matrix used to normalize the limiting distribution of $\widehat{\rho}_n$ and $\widehat{\delta}_n$. In particular, consider

$$\mathbf{\Omega}_n = \begin{bmatrix} \mathbf{\Omega}_n^{\delta\delta} & \mathbf{\Omega}_n^{\delta\rho} \\ \mathbf{\Omega}_n^{\delta\rho\prime} & \mathbf{\Omega}_n^{\rho\rho} \end{bmatrix} \qquad \text{and} \qquad \mathbf{\Psi}_n = \begin{bmatrix} \mathbf{\Psi}_n^{\delta\delta} & \mathbf{\Psi}_n^{\delta\rho} \\ \mathbf{\Psi}_n^{\delta\rho\prime} & \mathbf{\Psi}_n^{\rho\rho} \end{bmatrix} \tag{16}$$

with

$$\begin{aligned}
\mathbf{\Omega}_n^{\delta\delta} &= \mathbf{P}_n^{*\prime}\mathbf{\Psi}_n^{\delta\delta}\mathbf{P}_n^{*}, \\
\mathbf{\Omega}_n^{\delta\rho} &= \mathbf{P}_n^{*\prime}\mathbf{\Psi}_n^{\delta\rho}(\mathbf{\Psi}_n^{\rho\rho})^{-1}\mathbf{J}_n\left[\mathbf{J}_n'(\mathbf{\Psi}_n^{\rho\rho})^{-1}\mathbf{J}_n\right]^{-1}, \\
\mathbf{\Omega}_n^{\rho\rho} &= \left[\mathbf{J}_n'(\mathbf{\Psi}_n^{\rho\rho})^{-1}\mathbf{J}_n\right]^{-1}, \\
\mathbf{\Psi}_n^{\delta\delta} &= \sigma^2 n^{-1}\mathbf{H}_n'\mathbf{H}_n, \\
\mathbf{\Psi}_n^{\delta\rho} &= \sigma^2 n^{-1}\mathbf{H}_n'\left[\mathbf{a}_{1,n},\ldots,\mathbf{a}_{S,n}\right] + \mu^{(3)}n^{-1}\mathbf{H}_n'\left[vec_D(\mathbf{A}_{1,n}),\ldots,vec_D(\mathbf{A}_{S,n})\right], \\
\mathbf{\Psi}_n^{\rho\rho} &= (\psi_{rs,n}^{\rho\rho})_{r,s=1,\ldots,S},
\end{aligned}$$

where

$$\mathbf{J}_n = \mathbf{\Gamma}_n \begin{bmatrix} 1 \\ 2\rho_{0n} \end{bmatrix}.$$

The elements $\psi_{rs,n}^{\rho\rho}$ are defined as

$$\begin{aligned}
\psi_{rs,n}^{\rho\rho} &= \sigma^4 n^{-1} tr\left[(\mathbf{A}_{r,n}' + \mathbf{A}_{r,n})\mathbf{A}_{s,n}\right] \\
&+ \sigma^2 n^{-1}\mathbf{a}_{r,n}'\mathbf{a}_{s,n} \\
&+ (\mu^{(4)} - 3\sigma^4)n^{-1}vec_D(\mathbf{A}_{r,n})'vec_D(\mathbf{A}_{s,n}) \\
&+ \mu^{(3)}n^{-1}\left[\mathbf{a}_{r,n}'vec_D(\mathbf{A}_{s,n}) + vec_D(\mathbf{A}_{r,n})'\mathbf{a}_{s,n}\right]
\end{aligned}$$

with $\mu^{(3)} = E\varepsilon_{i,n}^3$ and $\mu^{(4)} = E\varepsilon_{i,n}^4$, and where $\mathbf{a}_{r,n} = \mathbf{T}_n\alpha_{r,n}$ with

$$\begin{aligned}
\mathbf{T}_n &= \mathbf{H}_n\mathbf{P}_n^{*}, \\
\mathbf{P}_n^{*} &= \mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_{0n})[\mathbf{Q}_{\mathbf{HZ}*}'(\rho_{0n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_{0n})]^{-1}, \\
\alpha_{r,n} &= -n^{-1}E\mathbf{Z}_n'\left(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n'\right)\left(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}'\right)\left(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n\right)\mathbf{u}_n.
\end{aligned}$$

In the appendix we prove the following theorem concerning the joint limiting distribution of $\widehat{\rho}_n$ and $\widehat{\delta}_n$.

Theorem 1 *Suppose Assumptions 1-8 hold,* $\sup_n |\beta_{0n}| < \infty$*,* $\sup_n |\pi_{0n}| < \infty$*, and that* $\lambda_{\min}(\mathbf{\Psi}_n) \geq const > 0$*. Then,* $\widehat{\rho}_n$ *is efficient among the class of GMM estimators based on GS2SLS residuals, and*

$$\begin{bmatrix} n^{1/2}(\widehat{\delta}_n - \delta_{0n}) \\ n^{1/2}(\widehat{\rho}_n - \rho_{0n}) \end{bmatrix} = \begin{bmatrix} \mathbf{P}_n^{*\prime} & 0 \\ 0 & \left[\mathbf{J}_n'(\mathbf{\Psi}_n^{\rho\rho})^{-1}\mathbf{J}_n\right]^{-1}\mathbf{J}_n'(\mathbf{\Psi}_n^{\rho\rho})^{-1} \end{bmatrix} \mathbf{\Psi}_n^{1/2}\xi_n + o_p(1),$$

where

$$\xi_n \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{p+S}).$$

Furthermore $\lambda_{\min}(\mathbf{\Omega}_n) \geq const > 0$*.*

Kelejian and Prucha (1998) considered two-step GMM/IV estimation for a model without outside endogenous variables. They only considered the first stage GMM estimator $\widetilde{\rho}_n$ and established its consistency. They also derived the limiting distribution of the S2SLS and GS2SLS estimator for their setting, but did not derive the joint limiting distribution of the GMM/IV estimators. The above joint-asymptotic-normality result fills in an important gap in their theory toward allowing a joint test for the absence of spatial dependencies, i.e., for a joint test $H_0 : \lambda_{0n} = 0, \rho_{0n} = 0$. We also note that in contrast to Arraiz et al. (2010) and Kelejian and Prucha (2010), and as in Kelejian and Prucha (1998), we cover here quadratic moment conditions corresponding to weights matrices $\mathbf{A}_{sn}$ with non-zero diagonal elements, which leads to terms involving the third and fourth moments of the innovations in $\mathbf{\Psi}_n$.

Observe that

$$\mathbf{\Omega}_n = \begin{bmatrix} \mathbf{P}_n^{*\prime} & 0 \\ 0 & \left[\mathbf{J}_n'(\mathbf{\Psi}_n^{\rho\rho})^{-1}\mathbf{J}_n\right]^{-1}\mathbf{J}_n'(\mathbf{\Psi}_n^{\rho\rho})^{-1} \end{bmatrix} \mathbf{\Psi}_n \begin{bmatrix} \mathbf{P}_n^{*} & 0 \\ 0 & (\mathbf{\Psi}_n^{\rho\rho})^{-1}\mathbf{J}_n\left[\mathbf{J}_n'(\mathbf{\Psi}_n^{\rho\rho})^{-1}\mathbf{J}_n\right]^{-1} \end{bmatrix}.$$

Given that $\lambda_{\min}(\mathbf{\Omega}_n) \geq const > 0$ it follows from Corollary F.4 in Pötscher and Prucha (1997) that

$$\begin{bmatrix} \widehat{\delta}_n - \delta_{0n} \\ \widehat{\rho}_n - \rho_{0n} \end{bmatrix} \sim AN(0, \mathbf{\Omega}_n/n). \tag{17}$$

We next define a consistent estimator for $\mathbf{\Omega}_n$. In particular, consider

$$\widehat{\mathbf{\Omega}}_n = \begin{bmatrix} \widehat{\mathbf{\Omega}}_n^{\delta\delta} & \widehat{\mathbf{\Omega}}_n^{\delta\rho} \\ \widehat{\mathbf{\Omega}}_n^{\delta\rho\prime} & \widehat{\mathbf{\Omega}}_n^{\rho\rho} \end{bmatrix}, \qquad \widehat{\mathbf{\Psi}}_n = \begin{bmatrix} \widehat{\mathbf{\Psi}}_n^{\delta\delta} & \widehat{\mathbf{\Psi}}_n^{\delta\rho} \\ \widehat{\mathbf{\Psi}}_n^{\delta\rho\prime} & \widehat{\mathbf{\Psi}}_n^{\rho\rho} \end{bmatrix} \tag{18}$$

with

$$\widehat{\mathbf{\Omega}}_n^{\delta\delta} = \widehat{\mathbf{P}}_n^{*\prime}\widehat{\mathbf{\Psi}}_n^{\delta\delta}\widehat{\mathbf{P}}_n^{*},$$

$$\widehat{\mathbf{\Omega}}_n^{\delta\rho} = \widehat{\mathbf{P}}_n^{*\prime}\widehat{\mathbf{\Psi}}_n^{\delta\rho}\left(\widehat{\mathbf{\Psi}}_n^{\rho\rho}\right)^{-1}\widehat{\mathbf{J}}_n\left[\widehat{\mathbf{J}}_n'\left(\widehat{\mathbf{\Psi}}_n^{\rho\rho}\right)^{-1}\widehat{\mathbf{J}}_n\right]^{-1},$$

$$\widehat{\mathbf{\Omega}}_n^{\rho\rho} = \left[\widehat{\mathbf{J}}_n'\left(\widehat{\mathbf{\Psi}}_n^{\rho\rho}\right)^{-1}\widehat{\mathbf{J}}_n\right]^{-1},$$

$$\widehat{\mathbf{\Psi}}_n^{\delta\delta} = \widehat{\sigma}_n^2 n^{-1}\mathbf{H}_n'\mathbf{H}_n,$$

$$\widehat{\mathbf{\Psi}}_n^{\delta\rho} = \widehat{\sigma}_n^2 n^{-1}\mathbf{H}_n'\left[\widehat{\mathbf{a}}_{1,n},\ldots,\widehat{\mathbf{a}}_{S,n}\right] + \widehat{\mu}_n^{(3)} n^{-1}\mathbf{H}_n'\left[vec_D(\mathbf{A}_{1,n}),\ldots,vec_D(\mathbf{A}_{S,n})\right],$$

$$\widehat{\mathbf{\Psi}}_n^{\rho\rho} = (\widehat{\psi}_{rs,n}^{\rho\rho})$$

where

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

and where $\widehat{\mathbf{P}}_n^{*}$, $\widehat{\mathbf{a}}_{s,n}$, $\widehat{\psi}_{rs,n}^{\rho\rho}$, and the estimators for the moments of the innovations $\varepsilon_{i,n}$ are as defined after (14), but with $\widetilde{\rho}_n$ replaced by $\widehat{\rho}_n$.[9] For interpretation, observe that $\widehat{\mathbf{\Omega}}_n^{\delta\delta} = \widehat{\sigma}_n^2[\widehat{\mathbf{Z}}_{n*}(\widehat{\rho}_n)'\mathbf{Z}_{n*}(\widehat{\rho}_n)]^{-1}$, i.e., the above expression for the estimator of variance-covariance matrix of the joint distribution of $\widehat{\delta}_n$ and $\widehat{\rho}_n$ delivers the usual estimator for the variance-covariance matrix of the GS2SLS estimator as a special case.

The next theorem establishes the consistency of $\widehat{\mathbf{\Omega}}_n(\widehat{\rho}_n)$ for $\mathbf{\Omega}_n$.

Theorem 2 *Suppose Assumptions 1-8 hold,* $\sup_n |\beta_{0n}| < \infty$, $\sup_n |\pi_{0n}| < \infty$, *and that* $\lambda_{\min}(\mathbf{\Psi}_n) \geq const > 0$. *Then* $\widehat{\mathbf{\Psi}}_n(\widehat{\rho}_n) - \mathbf{\Psi}_n = o_p(1)$ *and* $\widehat{\mathbf{\Omega}}_n(\widehat{\rho}_n) - \mathbf{\Omega}_n = o_p(1)$. *Furthermore* $\mathbf{\Psi}_n = O(1)$, $\mathbf{\Psi}_n^{-1} = O(1)$, $\mathbf{\Omega}_n = O(1)$ *and* $\mathbf{\Omega}_n^{-1} = O(1)$.

In light of the theorem inference can be based on (17) with $\mathbf{\Omega}_n$ replaced by $\widehat{\mathbf{\Omega}}_n$. The Monte Carlo study in section 4 shows that this approximation of the small-sample distribution works well.

4 Monte Carlo Study

In this section, we report on a small Monte Carlo study to explore the finite-sample properties of the estimators proposed in this paper. The Monte Carlo design is influenced by those used in Anselin and Florax (1995), Kelejian and Prucha (1999, 2007), and Arraiz et al. (2010).

[9] We note that the consistency result given below holds as long as a consistent estimator for ρ_{0n} is used in the formulation of the estimators $\widehat{\mathbf{\Omega}}_n$ and $\widehat{\mathbf{\Psi}}_n$. The suggestion of using $\widehat{\rho}_n$ rather than $\widetilde{\rho}_n$ is motivated by the fact $\widehat{\rho}_n$ is a more efficient estimator.

4.1 Monte Carlo Experiments

The model we employ in our Monte Carlo analysis is a special case of the one specified in (1) and (2) with $\mathbf{M}_n = \mathbf{W}_n$, one outside endogenous variable in $\mathbf{Y}_n = \widetilde{\mathbf{y}}_n$ and one exogenous variable in $\mathbf{X}_n = \mathbf{x}_n$:

$$\begin{aligned} \mathbf{y}_n &= \mathbf{x}_n\beta + \widetilde{\mathbf{y}}_n\pi + \lambda\mathbf{W}_n\mathbf{y}_n + \mathbf{u}_n \\ \mathbf{u}_n &= \rho\mathbf{W}_n\mathbf{u}_n + \boldsymbol{\epsilon}_n \end{aligned} \tag{19}$$

We have dropped the zero subscripts used to identify the true parameter values to simplify the notation. The outside endogenous variable is assumed to be generated as:

$$\widetilde{\mathbf{y}}_n = \widetilde{\mathbf{x}}_n\widetilde{\beta} + \mathbf{y}_n\widetilde{\pi} + \widetilde{\boldsymbol{\epsilon}}_n \tag{20}$$

where $\widetilde{\mathbf{x}}_n$ is an exogenous variable, $\widetilde{\boldsymbol{\epsilon}}_n$ is an $n \times 1$ vector of innovations, and $\widetilde{\beta}$ and $\widetilde{\pi}$ are coefficients. To simulate the data we employ the reduced form of the models, which is given by

$$\begin{pmatrix} \mathbf{y} \\ \widetilde{\mathbf{y}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} - \lambda\mathbf{W} & -\mathbf{\Pi} \\ -\widetilde{\mathbf{\Pi}} & \mathbf{I} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{m}_1 \\ \mathbf{m}_2 \end{pmatrix}, \tag{21}$$

where $\mathbf{m}_1 = \mathbf{x}\beta + \mathbf{u}$, $\mathbf{m}_2 = \widetilde{\mathbf{x}}\widetilde{\beta} + \widetilde{\boldsymbol{\epsilon}}$, $\mathbf{u}_n = (\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}\boldsymbol{\epsilon}_n$, $\mathbf{\Pi} = \pi\mathbf{I}_n$, and $\widetilde{\mathbf{\Pi}} = \widetilde{\pi}\mathbf{I}_n$. We focus on estimating the parameters $\boldsymbol{\delta} = (\beta, \pi, \lambda)'$ and ρ. For each configuration of parameters, we generated 2500 Monte Carlo runs based on the (correlated) draws of $\widetilde{\boldsymbol{\epsilon}}_n$ and $\boldsymbol{\epsilon}_n$.

The two $n \times 1$ regressors $\mathbf{x}_n$ and $\widetilde{\mathbf{x}}_n$ are normalized versions of income per-capita and the proportion of housing units, respectively, which are rental in 1980 for 760 counties in US mid-western states. The data were taken from Kelejian and Robinson (1995). We use the same normalization of the original data as in Arraiz et al. (2010), by subtracting from each observation of each variable the corresponding sample average, and then dividing that result by the sample standard deviation. We stack each vector of normalized observations twice underneath each other and draw the first n values of these normalized variables in our Monte Carlo experiments of sample size n. Hence, for sample sizes larger than 760 the last $n - 760$ observations were repeated. The set of normalized observations on these variables is fixed in repeated samples in our Monte Carlo runs.

We considered five values for λ and for ρ, namely: $-.8, -.3, 0, .3. .8$. In all of our experiments we chose $\beta = \widetilde{\beta} = 2$, $\pi = 1$, and $\widetilde{\pi} = -1$. We assume that the innovations corresponding to unit i, $\boldsymbol{\epsilon}_n$ and $\widetilde{\boldsymbol{\epsilon}}_n$, are i.i.d. draws from a bivariate normal distribution with mean zero and covariance matrix

$$s^2 \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

We consider two values for s, namely, .5 and 1.

We chose $\pi = 1$ and $\widetilde{\pi} = -1$ to avoid identification problems. At the conventional choices of $\pi = \widetilde{\pi} = 1$ the matrix inverse in equation (21) does not exist when $\lambda = 0$. We plan to discuss the restricted parameter space in future research.

We consider 4 different spatial-weights matrices, that we refer to as weights matrix 1, ..., 4. We use north-east modified-rook matrices as in Arraiz et al. (2010).

The north-east modified-rook matrix corresponds to a space in which units located in the northeastern portion are closer to each other and have more neighbors than the units in other quadrants. This leads to a distribution of units in space akin to the one of northeastern versus western states of the US. To define these matrices, assume a square grid with both the x and y coordinates only taking on the values 1, 1.5, 2, 2.5, ..., $\bar{m}$. Let the units in the northeastern quadrant of this matrix be located at the indicated discrete coordinates: $m \leq x \leq \bar{m}$ and $m \leq y \leq \bar{m}$. Let the remaining units be located only at integer values of the coordinates: $x = 1, 2, ..., m-1$ and $y = 1, 2, ..., m-1$. Accordingly, the number of units located in the northeastern quadrant is inversely related to m in this set-up.

For this space, define the Euclidean distance between any two units, i_1 and i_2 with coordinates of (x_1, y_1) and (x_2, y_2), respectively, as $d(i_1, i_2) = \left[(x_1 - x_2)^2 + (y_1 - y_2)^2\right]^{1/2}$.Now, define the (i, j)-th element of our row-normalized weights matrix $\mathbf{W}_n$ as

$$w_{ij} = w_{ij}^* / \sum_{j=1}^{n} w_{ij}^*,$$

$$w_{ij}^* = \begin{cases} 1 & \text{if} \quad 0 < d(i_1, i_2) \leq 1 \\ 0 & \text{else} \end{cases} .$$

We used four configurations of sample size n and reference coordinates on the lattice captured by m and $\bar{m}$. We refer to the corresponding spatial-weights matrices as matrices 1 through 4. In terms of tuples $(n; m; \bar{m})$ these matrices are described follows: Matrix 1 $(486; 5; 15)$, which has approximately 25% units in the north-east; Matrix 2 $(974; 7; 21)$, which has approximately 25% units in the north-east; Matrix 3 $(485; 14; 20)$, which has approximately 75% units in the north-east; and Matrix 4 $(945; 20; 28)$, which has approximately 76% units in the north-east. A north-east modified-rook matrix is illustrated in Figure 1 in Arraiz et al. (2010) for the case in which $m = 2$ and $\overline{m} = 5$.

The performance of our GMM/IV estimator depends on the relative noise in the system. To provide information on the relative noise we now report for each experiments the Monte Carlo average of an R^2 measure. This R^2 measure is calculate as the squared sample correlation coefficient between $\mathbf{y}$ and

$$E\mathbf{y} = \begin{pmatrix} \mathbf{I} - \lambda\mathbf{W} & -\mathbf{\Pi} \\ -\widetilde{\mathbf{\Pi}} & \mathbf{I} \end{pmatrix}^{-1} \begin{pmatrix} E\mathbf{m}_1 \\ E\mathbf{m}_2 \end{pmatrix},$$

with $E\mathbf{m}_1 = \mathbf{x}\beta$ and $E\mathbf{m}_2 = \widetilde{\mathbf{x}}\widetilde{\beta}$.

The number of Monte Carlo repetitions was $2,500$.

4.2 Monte Carlo Results

The detailed Monte Carlo results are given in Tables 1–20 in appendix E. The tables present the results for the final-stage GS2SLS estimator $\widehat{\delta}_n = (\widehat{\beta}_n', \widehat{\pi}_n', \widehat{\lambda}_n)'$ and GMM estimator $\widehat{\rho}_n$ defined by (13) and (14), and Wald tests based on the variance covariance estimator $\widehat{\mathbf{\Omega}}_n$

defined by (18). There are four parameters of interest (β, π, λ, and ρ) and 5 true values of ρ, causing there to be 20 tables of results. In all cases, we used the instrument matrix $\mathbf{H}_n = [\mathbf{x}, \widetilde{\mathbf{x}}, \mathbf{Wx}, \mathbf{W}\widetilde{\mathbf{x}}, \mathbf{W}^2\mathbf{x}, \mathbf{W}^2\widetilde{\mathbf{x}}]$.

Each table presents the results for the estimator for a specific parameter and contains 12 columns. The first column specifies the true values of λ, the second column specifies the $\mathbf{W}$-matrix. The third and fourth columns contain the median of the point estimates for the parameter for the cases of $s = .5$ and $s = 1$, respectively. The fifth and sixth columns contain the standard deviations (Std.Dev.) of the point estimates when $s = .5$ and $s = 1$. The seventh and eighth columns contain the means of the estimated standard deviations (Est.Std.Dev.) of the parameter estimates when $s = .5$ and $s = 1$. The ninth and tenth columns contain the rejection rates (Rej.Rate) of Wald test at the 5% level against the true null hypothesis that the parameter equals its true value when $s = .5$ and $s = 1$. The eleventh and twelfth columns contain the average R^2 values when $s = .5$ and $s = 1$.

The Monte Carlo results are encouraging, and suggest that the derived large-sample distributions for our estimators provides a reasonable approximation to the actual small-sample distributions. An inspection of the Monte Carlo results suggests (1) that the small-sample biases of the estimators are small for each of the parameters, (2) that the rejection rates are close to nominal size, (3) that the means of the estimated standard deviations of the parameter estimators over the Monte Carlo repetitions are close to the actual standard deviations, and (4) our estimator and our large-sample approximation to its distribution work well for the considered experiments.

5 Conclusion

In this paper we develop an IV/GMM estimation framework for the estimation of a spatial-autoregressive model with autoregressive disturbances, where the right-hand-side variables may include "outside"endogenous variables in addition to a spatial lag in the dependent variables, as well as exogenous variables. Our model includes the standard SARAR(1,1) model as a special case. The focus of our paper is on two-step estimators, because of their computational simplicity. Our analysis implicitly also contains the basic modules for an analysis of one-step estimators, which is actually mathematically less challenging.

The paper establishes the consistency of our IV/GMM estimators. Furthermore, the paper derives their joint asymptotic distribution, which can be readily used for, e.g., a Wald test of the joint hypothesis that there are no spatial interactions in the dependent variable, the disturbances and exogenous variables. Apart from deriving the asymptotic properties of the IV/GMM estimators, we provide a Monte Carlo analysis which shows that the small-sample distributions are well approximated by the derived large-sample distributions.

A Appendix: Asymptotic Linearity of S2SLS and GS2SLS Estimators

Lemma A.1 *: For fixed q let $\mathbf{a}_n = (a_{i,n})$ be some $qn \times 1$ vector where the absolute elements are uniformly bounded in n by some finite constant, and let $\mathbf{A}_n = (a_{ij,n})$ be some $qn \times qn$ matrix, where the row sums of the absolute elements are bounded uniformly in n by some finite constant. Let $\eta_n = (\eta_{i,n})$ be some $qn \times 1$ random vector with $\sup_n \max_{i=1}^{qn} E\left|\eta_{i,n}\right|^p < \infty$ for some $p > 1$, and let $\xi_n = (\xi_{i,n}) = \mathbf{a}_n + \mathbf{A}_n \eta_n$. Then $\sup_n \max_{i=1}^{qn} E\left|\xi_{i,n}\right|^p < \infty$.*

Proof. Let $C_a = \sup_n \max_{i=1}^{qn} |a_{i,n}|$, $C_A = \sup_n \max_{i=1}^{qn} \sum_{j=1}^{qn} |a_{ij,n}|$ and $C_\eta = \sup_n \max_{i=1}^{qn} E\left|\eta_{i,n}\right|^p$. Clearly

$$\begin{aligned}
|\xi_{in}| &= |a_{i,n}| + \left|\sum_{j=1}^{qn} a_{ij,n}\eta_{jn}\right| \leq |a_{i,n}| + \sum_{j=1}^{qn} |a_{ij,n}|\,|\eta_{jn}| \\
&= |a_{i,n}| + \left(\sum_{j=1}^{qn} |a_{ij,n}|\right) \sum_{j=1}^{qn} b_{j,n}\,|\eta_{jn}| \leq C_a + C_A \sum_{j=1}^{qn} b_{j,n}\,|\eta_{jn}|
\end{aligned}$$

with $b_{jn} = |a_{ij,n}| / \left(\sum_{j=1}^{qn} |a_{ij,n}|\right)$. Since $0 \leq b_{jn} \leq 1$ and $\sum_{j=1}^{qn} b_{j,n} = 1$ it follows from Lyapunov's inequality that

$$\sum_{j=1}^{qn} b_{j,n}\,|\eta_{jn}| \leq \left[\sum_{j=1}^{qn} b_{j,n}\,|\eta_{jn}|^p\right]^{1/p}$$

and hence

$$\begin{aligned}
E\left|\xi_{in}\right|^p &\leq 2^p \left\{C_a^p + C_A^p \sum_{j=1}^{qn} b_{j,n} E\left|\eta_{jn}\right|^p\right\} \leq 2^p \left\{C_a^p + C_A^p C_\eta \sum_{j=1}^{qn} b_{j,n}\right\} \\
&= 2^p \left\{C_a^p + C_A^p C_\eta\right\} < \infty
\end{aligned}$$

which proves the claim since C_a, C_A and C_η do not depend on i and n. ■

Lemma A.2 *: Given Assumptions 1-3 and 5, and given that $\sup_n |\beta_{0n}| < \infty$ and $\sup_n |\pi_{0n}| < \infty$ we have*

$$E\left|z_{ij,n}\right|^{2+\delta} \leq C < \infty \tag{A.1}$$

where C does not depend on i,j and n.

Proof. Recall that $\mathbf{Z}_n = [\mathbf{X}_n, \mathbf{Y}_n, \mathbf{W}\mathbf{y}_n]$. By assumption $\mathbf{X}_n$ is non-stochastic with $\sup_n \sup_{i,k} |x_{ik,n}| < \infty$, and so (A.1) holds trivially if $z_{ij,n}$ corresponds to an element of $\mathbf{X}_n$. By assumption $\sup_n \sup_{i,l} E|y_{il,n}|^{2+\delta} < \infty$ and thus (A.1) also holds trivially if $z_{ij,n}$ corresponds to an element of $\mathbf{Y}_n$. Next consider

$$\begin{aligned}
\overline{\mathbf{y}}_n &= \mathbf{W}_n \mathbf{y}_n = \varsigma_{1,n} + \varsigma_{2,n} + \varsigma_{3,n}, \\
\varsigma_{1,n} &= (\mathbf{I}_n - \lambda_{0n}\mathbf{W}_n)^{-1}\mathbf{X}_n\beta_{0n}, \\
\varsigma_{2,n} &= (\mathbf{I}_n - \lambda_{0n}\mathbf{W}_n)^{-1}\mathbf{Y}_n\pi_{0n}, \\
\varsigma_{3,n} &= (\mathbf{I}_n - \lambda_{0n}\mathbf{W}_n)^{-1}(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}\varepsilon_n.
\end{aligned}$$

Observe further that

$$E\left|\overline{y}_{i,n}\right|^{2+\delta} \leq 3^{1+\delta}\left\{|\varsigma_{1i,n}|^{2+\delta} + E|\varsigma_{2i,n}|^{2+\delta} + E|\varsigma_{3i,n}|^{2+\delta}\right\}.$$

Recall that by assumption the row and column sums of the absolute elements of $(\mathbf{I}_n - \lambda_{0n}\mathbf{W}_n)^{-1}$ and $(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}$, and hence those of $(\mathbf{I}_n - \lambda_{0n}\mathbf{W}_n)^{-1}(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}$ are bounded uniformly in n by some finite constant. Given that the elements of $\mathbf{X}_n\beta_{0n}$ are uniformly bounded clearly $\sup_n \sup_i |\varsigma_{1i,n}|^{2+\delta} < \infty$. Under the maintained assumptions the $2+\delta$ absolute moments of the elements of $\mathbf{Y}_n\pi_{0n}$ and ε_n are uniformly bounded. Consequently we also have by Lemma A.1 that $\sup_n \sup_i |\varsigma_{2i,n}|^{2+\delta} < \infty$ and $\sup_n \sup_i |\varsigma_{3i,n}|^{2+\delta} < \infty$, which completes the proof. ■

Lemma A.3 *: Suppose Assumptions 1-3, 6 and 7 hold. Consider the S2SLS estimator*

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

where $\widehat{\mathbf{Z}}_n = \mathbf{P}_{\mathbf{H}_n}\mathbf{Z}_n$ *and* $\mathbf{P}_{\mathbf{H}_n} = \mathbf{H}_n(\mathbf{H}_n'\mathbf{H}_n)^{-1}\mathbf{H}_n'$. *Then*
(a) $n^{1/2}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_{0n}) = n^{-1/2}\mathbf{T}_n'\varepsilon_n + o_p(1)$ *with* $\mathbf{T}_n = \mathbf{F}_n\mathbf{P}$ *and where*

$$\begin{aligned}
\mathbf{P} &= \mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}}[\mathbf{Q}_{\mathbf{HZ}}'\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}}]^{-1}, \\
\mathbf{F}_n &= \left(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n'\right)^{-1}\mathbf{H}_n.
\end{aligned}$$

(b) $n^{-1/2}\mathbf{T}_n'\varepsilon_n = O_p(1)$.
(c) $\mathbf{P}$ *is a finite matrix and* $\widetilde{\mathbf{P}}_n - \mathbf{P} = o_p(1)$ *for*

$$\begin{aligned}
\widetilde{\mathbf{P}}_n = &(n^{-1}\mathbf{H}_n'\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}_n'\mathbf{Z}_n)\times \\
&[(n^{-1}\mathbf{Z}_n'\mathbf{H}_n)(n^{-1}\mathbf{H}_n'\mathbf{H}_n)^{-1}(n^{-1}\mathbf{H}_n'\mathbf{Z}_n)]^{-1}.
\end{aligned}$$

(d) $\lambda_{\min}(n^{-1}\mathbf{T}_n'\mathbf{T}_n) \geq c$ *for some* $c > 0$ *for all large* n.

Proof. Clearly

$$n^{1/2}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_{0n}) = \widetilde{\mathbf{P}}_n' n^{-1/2}\mathbf{F}_n'\varepsilon_n,$$

where $\widetilde{\mathbf{P}}_n$ and $\mathbf{F}_n$ are defined in the lemma. Assumption 7 entails $\widetilde{\mathbf{P}}_n = \mathbf{P} + o_p(1)$ with $\mathbf{P}$ finite, which establishes (b). Since by Assumption 3 the row and column sums of $(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}$ are uniformly bounded in absolute value, and since by Assumption 6 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, $E(\varepsilon_n) = 0$ and $E(\varepsilon_n\varepsilon_n') = \sigma^2\mathbf{I}_n$. Therefore, $En^{-1/2}\mathbf{F}_n'\varepsilon_n = 0$ and the elements of $VC(n^{-1/2}\mathbf{F}_n'\varepsilon_n) = \sigma^2 n^{-1}\mathbf{F}_n'\mathbf{F}_n$ are also uniformly bounded in absolute value. Thus, by Chebyshev's inequality $n^{-1/2}\mathbf{F}_n'\varepsilon_n = O_p(1)$, and consequently $n^{1/2}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_{0n}) = \mathbf{P}'n^{-1/2}\mathbf{F}_n'\varepsilon_n + o_p(1)$ and $\mathbf{P}'n^{-1/2}\mathbf{F}_n'\varepsilon_n = O_p(1)$. This establishes (a) and (b), recalling that $\mathbf{T}_n = \mathbf{F}_n\mathbf{P}$. Next observe that

$$\begin{aligned}\lambda_{\min}(n^{-1}\mathbf{T}_n'\mathbf{T}_n) \geq\ & \lambda_{\min}\left[\left(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n'\right)^{-1}\left(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n\right)^{-1}\right]\lambda_{\min}\left[n^{-1}\mathbf{H}_n'\mathbf{H}_n\right] \\ & \lambda_{\min}\left\{[\mathbf{Q}_{\mathbf{HZ}}'\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}}]^{-1}\mathbf{Q}_{\mathbf{HZ}}'\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}}[\mathbf{Q}_{\mathbf{HZ}}'\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}}]^{-1}\right\} \\ & \geq c\end{aligned}$$

for some $c > 0$ in light of Assumptions 3 and 7, since $\lambda_{\min}\left[n^{-1}\mathbf{H}_n'\mathbf{H}_n\right] \geq \lambda_{\min}\mathbf{Q}_{\mathbf{HH}}/2 > 0$ for n sufficiently large. This establishes (d). ■

Lemma A.4 : *Suppose Assumptions 1-3, 6 and 7 hold. Consider the GS2SLS estimator*

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

where $\widehat{\mathbf{Z}}_{n*}(\widehat{\rho}_n) = \mathbf{P}_{\mathbf{H}_n}\ \mathbf{Z}_{n*}(\widehat{\rho}_n)$*, where* $\widehat{\rho}_n$ *is any* $n^{1/2}$*-consistent estimator for* ρ_{0n}*. Then*
(a) $n^{1/2}[\widehat{\boldsymbol{\delta}}_n(\widehat{\rho}_n) - \boldsymbol{\delta}_{0n}] = n^{-1/2}\mathbf{T}_n^{*\prime}\varepsilon_n + o_p(1)$ *with* $\mathbf{T}_n^* = \mathbf{F}_n^*\mathbf{P}_n^*$ *and where*

$$\begin{aligned}\mathbf{P}_n^* &= \mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_{0n})[\mathbf{Q}_{\mathbf{HZ}*}'(\rho_{0n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_{0n})]^{-1} \\ \mathbf{F}_n^* &= \mathbf{H}_n.\end{aligned}$$

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

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

(d) $\lambda_{\min}(n^{-1}\mathbf{T}_n^{*\prime}\mathbf{T}_n^*) \geq c$ *for some* $c > 0$ *for all large* n*.*

Proof. Note from (1) and (2) that

$$\mathbf{y}_{n*}(\hat{\rho}_n) = \mathbf{Z}_{n*}(\hat{\rho}_n)\boldsymbol{\delta}_{0n} + \varepsilon_n - (\hat{\rho}_n - \rho_{0n})\mathbf{M}_n\mathbf{u}_n$$

and hence

$$\begin{aligned}& n^{1/2}[\widehat{\boldsymbol{\delta}}_n(\hat{\rho}_n) - \boldsymbol{\delta}_{0n}] \\ &= \left[n^{-1}\widehat{\mathbf{Z}}_{n*}'(\hat{\rho}_n)\mathbf{Z}_{n*}(\hat{\rho}_n)\right]^{-1} n^{-1/2}\widehat{\mathbf{Z}}_{n*}'(\hat{\rho}_n)\left[\varepsilon_n - (\hat{\rho}_n - \rho_{0n})\mathbf{M}_n\mathbf{u}_n\right] \\ &= \widehat{\mathbf{P}}_n^{*\prime}\left[n^{-1/2}\mathbf{F}_n^{*\prime}\varepsilon_n - (\hat{\rho}_n - \rho_{0n})n^{-1/2}\mathbf{F}_n^{**\prime}\varepsilon_n\right],\end{aligned}$$

where $\widehat{\mathbf{P}}_n^*$ is defined in the lemma, and with $\mathbf{F}_n^* = \mathbf{H}_n$ and $\mathbf{F}_n^{**} = (\mathbf{I}_n - \rho_{0n}\mathbf{M}_n')^{-1}M_n'\mathbf{H}_n$. In light of Assumption 7, and since $\hat{\rho}_n$ is $n^{1/2}$-consistent, it follows that

$$n^{-1}\widehat{\mathbf{Z}}_{n*}'(\hat{\rho}_n)\mathbf{Z}_{n*}(\hat{\rho}_n) - \mathbf{Q}_{\mathbf{HZ}*}'(\rho_{0n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_{0n}) = o_p(1).$$

Since by Assumption 7 we have $\mathbf{Q}_{\mathbf{HZ}*}'(\rho_{0n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_{0n}) = O(1)$ and $[\mathbf{Q}_{\mathbf{HZ}*}'(\rho_{0n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_{0n})]^{-1} = O(1)$ it follows that

$$[n^{-1}\widehat{\mathbf{Z}}_{n*}'(\hat{\rho}_n)\mathbf{Z}_{n*}(\hat{\rho}_n)]^{-1} - [\mathbf{Q}_{\mathbf{HZ}*}'(\rho_{0n})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}*}(\rho_{0n})]^{-1} = o_p(1);$$

compare, e.g., Pötscher and Prucha (1997), Lemma F1. In light of this it follows further that $\widehat{\mathbf{P}}_n^* - \mathbf{P}_n^* = o_p(1)$ and $\mathbf{P}_n^* = O(1)$, with $\mathbf{P}_n^*$ defined in the lemma. By argumentation analogous to that in the proof of Lemma A.3 it is readily seen that $n^{-1/2}\mathbf{F}_n^{*\prime}\varepsilon_n = O_p(1)$ and $n^{-1/2}\mathbf{F}_n^{**\prime}\varepsilon_n = O_p(1)$. Consequently $n^{1/2}[\widehat{\boldsymbol{\delta}}_n(\hat{\rho}_n) - \boldsymbol{\delta}_{0n}] = \mathbf{P}_n^{*\prime}n^{-1/2}\mathbf{F}_n^{*\prime}\varepsilon_n + o_p(1)$ and $\mathbf{P}_n^{*\prime}n^{-1/2}\mathbf{F}_n^{*\prime}\varepsilon_n = O_p(1)$, observing again that $\hat{\rho}_n - \rho_{0n} = o_p(1)$. This establishes (a)-(c) recalling that $\mathbf{T}_n^* = \mathbf{F}_n^*\mathbf{P}_n^*$. Next observe that

$$\begin{aligned}\lambda_{\min}(n^{-1}\mathbf{T}_n^{*\prime}\mathbf{T}_n^*) \geq{}& \lambda_{\min}\left[\mathbf{Q}_{\mathbf{HH}}^{\prime -1/2}n^{-1}\mathbf{H}_n'\mathbf{H}_n\mathbf{Q}_{\mathbf{HH}}^{-1/2}\right] \\ & \lambda_{\min}\left\{[\mathbf{Q}_{\mathbf{HZ}_*}'(\rho_{n0})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}_*}(\rho_{n0})]^{-1}\right. \\ & \left.\times\mathbf{Q}_{\mathbf{HZ}_*}'(\rho_{n0})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}_*}(\rho_{n0})[\mathbf{Q}_{\mathbf{HZ}_*}'(\rho_{n0})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}_*}(\rho_{n0})]^{-1}\right\} \\ & \geq \lambda_{\min}\left[\mathbf{Q}_{\mathbf{HH}}^{-1}\right]\lambda_{\min}\left\{[\mathbf{Q}_{\mathbf{HZ}_*}'(\rho_{n0})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}_*}(\rho_{n0})]^{-1}\right\}\lambda_{\min}\left[n^{-1}\mathbf{H}_n'\mathbf{H}_n\right] \\ & \geq \lambda_{\min}\left[\mathbf{Q}_{\mathbf{HH}}^{-1}\right]\lambda_{\min}\left\{[\mathbf{Q}_{\mathbf{HZ}_*}'(\rho_{n0})\mathbf{Q}_{\mathbf{HH}}^{-1}\mathbf{Q}_{\mathbf{HZ}_*}(\rho_{n0})]^{-1}\right\}\lambda_{\min}\left[\mathbf{Q}_{\mathbf{HH}}^{-1}\right]/2\end{aligned}$$

for some $c > 0$ in light of Assumption 7, since $\lambda_{\min}\left[n^{-1}\mathbf{H}_n'\mathbf{H}_n\right] \geq \lambda_{\min}\mathbf{Q}_{\mathbf{HH}}/2 > 0$ for n sufficiently large. This establishes (d). ■

B Appendix: Generic Consistency and Asymptotic Normality of Two-step Estimators

In the following we establish a generic consistency and asymptotic normality result for IV/GMM estimators. In (10) - (12) we used the notation $\widetilde{\boldsymbol{\delta}}_n$, $\widetilde{\mathbf{u}}$ and $\widetilde{\rho}_n$ for the 2SLS estimator for $\boldsymbol{\delta}_{0n}$, 2SLS residuals and the initial GMM estimator for ρ_{0n}. In abuse of notation, we use in the following the same symbols to denote a generic estimator for $\boldsymbol{\delta}_{0n}$, correspondingly defined residuals, and a generic GMM estimator for ρ_{0n}.

B.1 Generic Consistency

In the following we provide a general consistency result for some generic GMM estimator for ρ_{0n} defined as

$$\widetilde{\rho}_n = \arg \min_{\rho\in[-a^\rho,a^\rho]} \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n,\rho)'\widetilde{\boldsymbol{\Upsilon}}_n\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n,\rho) \tag{B.1}$$

where $(-1,1) \subset [-a^\rho, a^\rho]$, $a^\rho > 1$. We will maintain the following assumptions regarding $\widetilde{\boldsymbol{\delta}}_n$ and $\widetilde{\boldsymbol{\Upsilon}}_n$.

Assumption B.1 *: The estimator $\widetilde{\boldsymbol{\delta}}_n$ is asymptotically linear in the sense that*

$$n^{1/2}[\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_{0n}] = n^{-1/2}\mathbf{T}_n'\varepsilon_n + o_p(1) \tag{B.2}$$

with $\mathbf{T}_n = \mathbf{F}_n\mathbf{P}_n$, where $\mathbf{F}_n$ and $\mathbf{P}_n$ are nonstochastic $n\times p$ and $p\times p$ matrices whose elements are uniformly bounded in absolute value and where the smallest eigenvalue of $n^{-1}(\mathbf{T}_n'\mathbf{T}_n)$ is bounded away from zero.

Assumption B.2 *: The moment weights matrix $\widetilde{\boldsymbol{\Upsilon}}_n$ satisfies that $\widetilde{\boldsymbol{\Upsilon}}_n - \boldsymbol{\Upsilon}_n = o_p(1)$, where $\boldsymbol{\Upsilon}_n$ are $S\times S$ non-stochastic symmetric positive definite matrices. Furthermore the smallest eigenvalue of $\boldsymbol{\Upsilon}_n$ is bounded away from zero, and the largest eigenvalue of $\boldsymbol{\Upsilon}_n$ is bounded from above (by a finite constant).*

We now have the following consistency result for $\widetilde{\rho}_n$.

Theorem B.1 ***(Theorem on Consistency of*** $\widetilde{\rho}_n$*)* *Suppose Assumptions 1-5, 8, B.1 and B.2 hold, and suppose* $\sup_n|\beta_{0n}| < \infty$ *and* $\sup_n|\pi_{0n}| < \infty$*. Then,*

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

Proof.Theorem on Consistency of $\widetilde{\rho}_n$.[10] 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,

$$\begin{aligned} R_n(\omega,\rho) &= \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n,\rho)'\widetilde{\boldsymbol{\Upsilon}}_n\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n,\rho) \\ &= \left[\widetilde{\boldsymbol{\Gamma}}_n(\rho,\rho^2)'-\widetilde{\gamma}_n\right]'\widetilde{\boldsymbol{\Upsilon}}_n\left[\widetilde{\boldsymbol{\Gamma}}_n(\rho,\rho^2)'-\widetilde{\gamma}_n\right] \\ \overline{R}_n(\rho) &= \left[\boldsymbol{\Gamma}_n(\rho,\rho^2)'-\gamma_n\right]'\boldsymbol{\Upsilon}_n\left[\boldsymbol{\Gamma}_n(\rho,\rho^2)'-\gamma_n\right], \end{aligned}$$

where

$$\widetilde{\gamma}_n = n^{-1}\begin{bmatrix} \widetilde{\mathbf{u}}_n'\mathbf{A}_{1n}\widetilde{\mathbf{u}}_n \\ \vdots \\ \widetilde{\mathbf{u}}_n'\mathbf{A}_{Sn}\widetilde{\mathbf{u}}_n \end{bmatrix}, \qquad \widetilde{\boldsymbol{\Gamma}}_n = n^{-1}\begin{bmatrix} 2\widetilde{\mathbf{u}}_n'\mathbf{M}_n'\mathbf{A}_{1n}\widetilde{\mathbf{u}}_n & -\widetilde{\mathbf{u}}_n'\mathbf{M}_n'\mathbf{A}_{1n}\mathbf{M}_n\widetilde{\mathbf{u}}_n \\ \vdots & \vdots \\ 2\widetilde{\mathbf{u}_n}'\mathbf{M}_n'\mathbf{A}_{Sn}\widetilde{\mathbf{u}}_n & -\widetilde{\mathbf{u}}_n'\mathbf{M}_n'\mathbf{A}_{Sn}\mathbf{M}_n\widetilde{\mathbf{u}}_n \end{bmatrix},$$

and the nonstochastic counter parts γ_n and $\boldsymbol{\Gamma}_n$ are defined in (9). 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_{0n} is an identifiably unique sequence of minimizers of $\overline{R}_n$. Observe that $\overline{R}_n(\rho)\geq 0$ and that $\overline{R}_n(\rho_{0n})=0$, since $\gamma_n=\boldsymbol{\Gamma}_n[\rho_{0n},\rho_{0n}^2]'$ by (9)). Utilizing Assumption 8 and B.2 we get

$$\begin{aligned} \overline{R}_n(\rho)-\overline{R}_n(\rho_{0n}) &= \overline{R}_n(\rho) \\ &= \left[\rho-\rho_{0n},\rho^2-\rho_{0n}^2\right]\boldsymbol{\Gamma}_n'\boldsymbol{\Upsilon}_n\boldsymbol{\Gamma}_n\left[\rho-\rho_{0n},\rho^2-\rho_{0n}^2\right]' \\ &\geq \lambda_{\min}(\boldsymbol{\Upsilon}_n)\lambda_{\min}(\boldsymbol{\Gamma}_n'\boldsymbol{\Gamma}_n)\left[\rho-\rho_{0n},\rho^2-\rho_{0n}^2\right]\left[\rho-\rho_{0n},\rho^2-\rho_{0n}^2\right]' \\ &\geq \lambda_*\left[\rho-\rho_{0n}\right]^2 \end{aligned}$$

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

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

which proves that ρ_{0n} is identifiably unique. Next let $\boldsymbol{\Phi}_n=[\boldsymbol{\Gamma}_n,-\gamma_n]$ and $\widetilde{\boldsymbol{\Phi}}_n=[\widetilde{\boldsymbol{\Gamma}}_n,-\widetilde{\gamma}_n]$, then

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

[10] The basic structure of the proof is similar to the consistency proof for the GMM estimator for the spatial-autoregressive parameter considered in Kelejian and Prucha (1999). The earlier paper considered a more restrictive set of moment conditions and did not allow for "outside" endogenous variables.

Observe that

$$\widetilde{\mathbf{u}}_n - \mathbf{u}_n = \mathbf{D}_n \mathbf{\Delta}_n$$

where $\mathbf{D}_n = -\mathbf{Z}_n$ and $\mathbf{\Delta}_n = \widetilde{\boldsymbol{\delta}}_n - \delta_{0n}$. In light of Lemma A.2 we have $E\,|d_{ij,n}|^{2+\delta} \leq C < \infty$. Under Assumption B.1 we have $n^{-1/2}\mathbf{T}_n'\varepsilon_n = O_p(1)$ and hence $n^{1/2}\left\|\mathbf{\Delta}_n\right\| = O_p(1)$. Next observe that the elements of $\mathbf{\Phi}_n$ and $\widetilde{\mathbf{\Phi}}_n$ are all of the form $n^{-1}Eu_n'\mathbf{A}_n u_n$ 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 part (a) of Lemma C.1 in Kelejian and Prucha (2010) that $\widetilde{\mathbf{\Phi}}_n - \mathbf{\Phi}_n \xrightarrow{p} 0$, and that the elements of $\widetilde{\mathbf{\Phi}}_n$ and $\mathbf{\Phi}_n$ are, respectively, $O_p(1)$ and $O(1)$. The elements of $\widetilde{\mathbf{\Upsilon}}_n$ and $\mathbf{\Upsilon}_n$ have the analogous properties in light of Assumption B.2. 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.,

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

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

B.2 Generic Asymptotic Normality

We now establish the joint asymptotic normality of the estimator $\widetilde{\rho}_n$ of Theorem B.1 and some estimator $\widetilde{\widetilde{\boldsymbol{\delta}}}_n$. We maintain the following assumption regarding $\widetilde{\widetilde{\boldsymbol{\delta}}}_n$.

Assumption B.3 *: The estimator $\widetilde{\widetilde{\boldsymbol{\delta}}}_n$ is asymptotically linear in the sense that*

$$n^{1/2}[\widetilde{\widetilde{\boldsymbol{\delta}}}_n - \boldsymbol{\delta}_{0n}] = n^{-1/2}\mathbf{T}_n^{\bullet\prime}\varepsilon_n + o_p(1) \tag{B.3}$$

with $\mathbf{T}_n^\bullet = \mathbf{F}_n^\bullet\mathbf{P}_n^\bullet$, where $\mathbf{F}_n^\bullet$ and $\mathbf{P}_n^\bullet$ are nonstochastic $n \times p^\bullet$ and $p^\bullet \times p^\bullet$ matrices whose elements are uniformly bounded in absolute value and where the smallest eigenvalue of $n^{-1}(\mathbf{T}_n^{\bullet\prime}\mathbf{T}_n^\bullet)$ is bounded away from zero.

The limiting distribution of $\widetilde{\widetilde{\theta}}_n = \left(\widetilde{\widetilde{\boldsymbol{\delta}}}_n', \widetilde{\rho}_n\right)'$ will depend on the limiting distribution of $\mathbf{v}_n = (\mathbf{v}_n^{\boldsymbol{\delta}\prime}, \mathbf{v}_n^{\rho\prime})'$ where

$$\mathbf{v}_n^{\boldsymbol{\delta}} = n^{-1/2}\mathbf{F}_n^{\bullet\prime}\varepsilon_n, \tag{B.4}$$

$$\mathbf{v}_n^{\rho} = n^{-1/2}\begin{bmatrix} \varepsilon_n'\mathbf{A}_{1,n}\varepsilon_n + \mathbf{a}_{1,n}'\varepsilon_n \\ \vdots \\ \varepsilon_n'\mathbf{A}_{S,n}\varepsilon_n + \mathbf{a}_{S,n}'\varepsilon_n \end{bmatrix},$$

with $\mathbf{a}_{r,n} = \mathbf{T}_n \alpha_{r,n}$ with $\alpha_{r,n} = -n^{-1} E\mathbf{Z}_n' (\mathbf{I}_n - \rho_{0n}\mathbf{M}_n') (\mathbf{A}_{r,n} + \mathbf{A}_{r,n}') (\mathbf{I}_n - \rho_{0n}\mathbf{M}_n) \mathbf{u}_n$. By Lemma A.1 in Kelejian and Prucha (2010) we have under Assumptions 1-5, 8, B.1 and B.3 that $\mathbf{v}_n$ has mean zero and its variance-covariance matrix is given by

$$\mathbf{\Psi}_n = \begin{bmatrix} \mathbf{\Psi}_n^{\delta\delta} & \mathbf{\Psi}_n^{\delta\rho} \\ \mathbf{\Psi}_n^{\delta\rho\prime} & \mathbf{\Psi}_n^{\rho\rho} \end{bmatrix} \tag{B.5}$$

with

$$\begin{aligned}
\mathbf{\Psi}_n^{\delta\delta} &= \sigma^2 n^{-1} \mathbf{F}_n^{\bullet\prime} \mathbf{F}_n^{\bullet}, \\
\mathbf{\Psi}_n^{\delta\rho} &= \sigma^2 n^{-1} \mathbf{F}_n^{\bullet\prime} \left[\mathbf{a}_{1,n}, \ldots, \mathbf{a}_{S,n}\right] + \mu^{(3)} n^{-1} \mathbf{F}_n^{\bullet\prime} \left[vec_D(\mathbf{A}_{1,n}), \ldots, vec_D(\mathbf{A}_{S,n})\right], \\
\mathbf{\Psi}_n^{\rho\rho} &= (\psi_{rs,n}^{\rho\rho})_{r,s=1,\ldots,S}
\end{aligned}$$

with

$$\begin{aligned}
\psi_{rs,n}^{\rho\rho} &= \sigma^4 n^{-1} tr\left[(\mathbf{A}_{r,n}' + \mathbf{A}_{r,n})\mathbf{A}_{s,n}\right] + \sigma^2 n^{-1} \mathbf{a}_{r,n}' \mathbf{a}_{s,n} \\
&\quad + (\mu^{(4)} - 3\sigma^4) n^{-1} vec_D(\mathbf{A}_{r,n})' vec_D(\mathbf{A}_{s,n}) \\
&\quad + \mu^{(3)} n^{-1} \left[\mathbf{a}_{r,n}' vec_D(\mathbf{A}_{s,n}) + vec_D(\mathbf{A}_{r,n})' \mathbf{a}_{s,n}\right]
\end{aligned}$$

where $\mu^{(3)} = E\varepsilon_{i,n}^3$ and $\mu^{(4)} = E\varepsilon_{i,n}^4$. Also let

$$\mathbf{J}_n = \mathbf{\Gamma}_n \begin{bmatrix} 1 \\ 2\rho_{0n} \end{bmatrix} \tag{B.6}$$

and

$$\mathbf{\Omega}_n = \begin{bmatrix} \mathbf{\Omega}_n^{\delta\delta} & \mathbf{\Omega}_n^{\delta\rho} \\ \mathbf{\Omega}_n^{\delta\rho\prime} & \mathbf{\Omega}_n^{\rho\rho} \end{bmatrix} \tag{B.7}$$

with

$$\begin{aligned}
\mathbf{\Omega}_n^{\delta\delta} &= \mathbf{P}_n^{\bullet\prime} \mathbf{\Psi}_n^{\delta\delta} \mathbf{P}_n^{\bullet}, \\
\mathbf{\Omega}_n^{\delta\rho} &= \mathbf{P}_n^{\bullet\prime} \mathbf{\Psi}_n^{\delta\rho} \mathbf{\Upsilon}_n \mathbf{J}_n (\mathbf{J}_n' \mathbf{\Upsilon}_n \mathbf{J}_n)^{-1}, \\
\mathbf{\Omega}_n^{\rho\rho} &= (\mathbf{J}_n' \mathbf{\Upsilon}_n \mathbf{J}_n)^{-1} \mathbf{J}_n' \mathbf{\Upsilon}_n \mathbf{\Psi}_n^{\rho\rho} \mathbf{\Upsilon}_n \mathbf{J}_n (\mathbf{J}_n' \mathbf{\Upsilon}_n \mathbf{J}_n)^{-1}.
\end{aligned}$$

We next give results concerning the asymptotic normality $\widetilde{\rho}_n$, and consequently concerning the joint asymptotic normality of $\widetilde{\widetilde{\theta}}_n = \left(\widetilde{\widetilde{\boldsymbol{\delta}}}_n', \widetilde{\rho}_n\right)'$.

Theorem B.2 ***(Theorem on Asymptotic Normality of*** $\widetilde{\rho}_n$*) Let* $\widetilde{\rho}_n$ *be the weighted nonlinear least squares estimators defined by (B.1). Suppose Assumptions 1-5, 8, B.1 and B.2 hold,* $\sup_n |\beta_{0n}| < \infty$ *and* $\sup_n |\pi_{0n}| < \infty$*, and suppose the smallest eigenvalue of* $\mathbf{\Psi}_n^{\rho\rho}$ *is bounded away from zero, then*

$$n^{1/2}(\widetilde{\rho}_n - \rho_n) = -(\mathbf{J}_n' \mathbf{\Upsilon}_n \mathbf{J}_n)^{-1} \mathbf{J}_n' \mathbf{\Upsilon}_n (\mathbf{\Psi}_n^{\rho\rho})^{1/2} \xi_n^{\rho} + o_p(1), \tag{B.8}$$

where

$$\xi_n^{\rho} = -\,(\boldsymbol{\Psi}_n^{\rho\rho})^{-1/2}\,\mathbf{v}_n^{\rho} \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_S).$$

Furthermore $n^{1/2}(\widetilde{\rho}_n - \rho_{0n}) = O_p(1)$ *and*

$$\boldsymbol{\Omega}_n^{\rho\rho}(\boldsymbol{\Upsilon}_n) = (\mathbf{J}_n'\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1}\mathbf{J}_n'\boldsymbol{\Upsilon}_n\boldsymbol{\Psi}_n^{\rho\rho}\boldsymbol{\Upsilon}_n\mathbf{J}_n(\mathbf{J}_n'\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1} \geq const > 0.$$

Proof.Theorem on Asymptotic Normality of $\widetilde{\rho}_n$.[11] We have shown in Theorem B.1 that the GMM estimator $\widetilde{\rho}_n$ is consistent. Apart on a set whose probability tends to zero the estimator satisfies the following first order condition

$$\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)'\widetilde{\boldsymbol{\Upsilon}}_n\frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)}{\partial \rho} = 0.$$

Substituting the mean value theorem expression

$$\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \widetilde{\rho}_n) = \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \rho_{0n}) + \frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \overline{\rho}_n)}{\partial \rho}(\widetilde{\rho}_n - \rho_{0n})$$

into the first order condition yields

$$\frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)}{\partial \rho'}\widetilde{\boldsymbol{\Upsilon}}_n\frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \overline{\rho}_n)}{\partial \rho}n^{1/2}(\widetilde{\rho}_n - \rho_{0n}) = -\frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)}{\partial \rho'}\widetilde{\boldsymbol{\Upsilon}}_n n^{1/2}\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \rho_{0n}) \tag{B.9}$$

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

$$\frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \rho)}{\partial \rho} = -\widetilde{\boldsymbol{\Gamma}}_n\begin{bmatrix} 1 \\ 2\rho \end{bmatrix} \tag{B.10}$$

and consider the nonnegative scalars

$$\begin{aligned}
\widetilde{\boldsymbol{\Xi}}_n &= \frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)}{\partial \rho'}\widetilde{\boldsymbol{\Upsilon}}_n\frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \overline{\rho}_n)}{\partial \rho} = \begin{bmatrix} 1 \\ 2\widetilde{\rho}_n \end{bmatrix}'\widetilde{\boldsymbol{\Gamma}}_n'\widetilde{\boldsymbol{\Upsilon}}_n\widetilde{\boldsymbol{\Gamma}}_n\begin{bmatrix} 1 \\ 2\overline{\rho}_n \end{bmatrix}, \qquad \text{(B.11)} \\
\boldsymbol{\Xi}_n &= \begin{bmatrix} 1 \\ 2\rho_{0n} \end{bmatrix}'\boldsymbol{\Gamma}_n'\boldsymbol{\Upsilon}_n\boldsymbol{\Gamma}_n\begin{bmatrix} 1 \\ 2\rho_{0n} \end{bmatrix}.
\end{aligned}$$

In proving Theorem B.1 we have demonstrated that $\widetilde{\boldsymbol{\Gamma}}_n - \boldsymbol{\Gamma}_n \xrightarrow{p} 0$ and that the elements of $\widetilde{\boldsymbol{\Gamma}}_n$ and $\boldsymbol{\Gamma}_n$ are $O_p(1)$ and $O(1)$, respectively. By Assumption B.2 we have $\widetilde{\boldsymbol{\Upsilon}}_n - \boldsymbol{\Upsilon}_n = o_p(1)$ and also that the elements of $\widetilde{\boldsymbol{\Upsilon}}_n$ and $\boldsymbol{\Upsilon}_n$ are $O_p(1)$ and $O(1)$, respectively. Since $\widetilde{\rho}_n$ and $\overline{\rho}_n$ are consistent and bounded in absolute value, clearly

$$\widetilde{\boldsymbol{\Xi}}_n - \boldsymbol{\Xi}_n \xrightarrow{p} 0 \tag{B.12}$$

[11] The structure of the proof is similar to that of the asymptotic normality proof for the GMM estimator of the spatial-autoregressive paramter given in Kelejian and Prucha (2009). The earlier paper considered a more restrictive set of moment conditions and did not allow for "outside" endogenous variables.

as $n \to \infty$, and furthermore $\widetilde{\Xi}_n = O_p(1)$ and $\Xi_n = O(1)$. In particular $\Xi_n \leq \lambda_{\Xi}^{**}$ where λ_{Ξ}^{**} is some finite constant. In light of Assumptions B.2 and 8 we have $\Xi_n \geq \lambda_{\min}(\mathbf{\Upsilon}_n)\lambda_{\min}(\mathbf{\Gamma}_n'\mathbf{\Gamma}_n)(1+4\rho_n^2) \geq \lambda_{\Xi}^{*}$ for some $\lambda_{\Xi}^{*} > 0$. Hence $0 < \Xi_n^{-1} \leq 1/\lambda_{\Xi}^{*} < \infty$, and thus we also have $\Xi_n^{-1} = O(1)$. Let $\widetilde{\Xi}_n^{+}$ denote the generalized inverse of $\widetilde{\Xi}_n$. It then follows as a special case of Lemma F1 in Pötscher and Prucha (1997) that $\widetilde{\Xi}_n$ is nonsingular eventually with probability tending to one, that $\widetilde{\Xi}_n^{+} = O_p(1)$, and that

$$\widetilde{\Xi}_n^{+} - \Xi_n^{-1} \xrightarrow{p} 0 \tag{B.13}$$

as $n \to \infty$.

Premultiplying (B.9) by $\widetilde{\Xi}_n^{+}$ and rearranging terms yields

$$n^{1/2}(\widetilde{\rho}_n - \rho_n) = \left[1 - \widetilde{\Xi}_n^{+}\widetilde{\Xi}_n\right] n^{1/2}(\widetilde{\rho}_n - \rho_n) - \widetilde{\Xi}_n^{+}\frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)}{\partial \rho}\widetilde{\mathbf{\Upsilon}}_n n^{1/2}\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \rho_{0n}).$$

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

$$n^{1/2}(\widetilde{\rho}_n - \rho_n) = -\widetilde{\Xi}_n^{+}\frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)}{\partial \rho}\widetilde{\mathbf{\Upsilon}}_n n^{1/2}\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \rho_{0n}) + o_p(1). \tag{B.14}$$

Observe that

$$\widetilde{\Xi}_n^{+}\frac{\partial \mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)}{\partial \rho}\widetilde{\mathbf{\Upsilon}}_n - \Xi_n^{-1}\begin{bmatrix} 1 \\ 2\rho_{0n} \end{bmatrix}'\mathbf{\Gamma}_n'\mathbf{\Upsilon}_n = o_p(1), \tag{B.15}$$

and observe that

$$\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \rho_{0n}) = n^{-1}\begin{bmatrix} \widetilde{\mathbf{u}}_n'\mathbf{C}_{1,n}(\rho_{0n})\widetilde{\mathbf{u}}_n \\ \vdots \\ \widetilde{\mathbf{u}}_n'\mathbf{C}_{S,n}(\rho_{0n})\widetilde{\mathbf{u}}_n \end{bmatrix}$$

with $\mathbf{C}_{1,n}(\rho) = (1/2)\left(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n'\right)\left(A_{s,n} + A_{s,n}'\right)\left(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n\right)$, and that

$$\widetilde{\mathbf{u}}_n - \mathbf{u}_n = \mathbf{D}_n\mathbf{\Delta}_n$$

where $\mathbf{D}_n = -\mathbf{Z}_n$ and $\mathbf{\Delta}_n = \widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_{0n}$. In light of Lemma A.2 we have $E\left|d_{ij,n}\right|^{2+\delta} \leq C < \infty$. Under Assumption B.1 we have $n^{-1/2}\mathbf{T}_n'\varepsilon_n = O_p(1)$ and hence $n^{1/2}\left\|\mathbf{\Delta}_n\right\| = O_p(1)$. It now follows from part (c) of Lemma C.1 in Kelejian and Prucha (2010) that the elements of $\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \rho_{0n})$ can be expressed as

$$n^{-1/2}\widetilde{\mathbf{u}}_n'\mathbf{C}_{r,n}(\rho_{0n})\widetilde{\mathbf{u}}_n = n^{-1/2}\mathbf{u}_n'\mathbf{C}_{r,n}(\rho_{0n})\mathbf{u}_n + \alpha_{r,n}'n^{1/2}\left(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}_{0n}\right) + o_p(1),$$

where

$$\alpha_{s,n} = -2n^{-1}E\mathbf{Z}_n'\mathbf{C}_{s,n}(\rho_{0n})\mathbf{u}_n.$$

Furthermore, the lemma implies that the elements of $\alpha_{s,n}$ are uniformly bounded in absolute value. Utilizing $u_n = (\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}\varepsilon_n$ and Assumption B.1 we have

$$n^{1/2}\mathbf{m}_n(\widetilde{\boldsymbol{\delta}}_n, \rho_{0n}) = n^{-1/2}\begin{bmatrix} \frac{1}{2}\varepsilon_n'(\mathbf{A}_{1,n} + \mathbf{A}_{1,n}')\varepsilon_n + \mathbf{a}_{1,n}'\varepsilon_n \\ \vdots \\ \frac{1}{2}\varepsilon_n'(\mathbf{A}_{S,n} + \mathbf{A}_{S,n}')\varepsilon_n + \mathbf{a}_{S,n}'\varepsilon_n \end{bmatrix} + o_p(1) \tag{B.16}$$

where $\mathbf{a}_{s,n} = \mathbf{T}_n \alpha_{s,n}$, $s = 1, \ldots, S$. Observe that the elements of $\mathbf{a}_{s,n}$ are uniformly bounded in absolute value. As discussed before the theorem the VC matrix of the vector of quadratic forms on the RHS of (B.16) is given by $\Psi_n^{\rho\rho} = (\psi_{rs,n})$ as defined in (B.5). By assumption $\lambda_{\min}(\Psi_n^{\rho\rho}) \geq const > 0$. Since the matrices $A_{r,n}$, the vectors $\mathbf{a}_{r,n}$, and the innovations ε_n satisfy all of the remaining assumptions of the central limit theorem for vectors of linear quadratic forms given as Theorem A.1 in Kelejian and Prucha (2010) it follows that

$$\xi_n^{\rho} = (\mathbf{\Psi}_n^{\rho\rho})^{-1/2} n^{-1/2} \begin{bmatrix} \frac{1}{2}\varepsilon_n'(\mathbf{A}_{1,n} + \mathbf{A}_{1,n}')\varepsilon_n + \mathbf{a}_{1,n}'\varepsilon_n \\ \vdots \\ \frac{1}{2}\varepsilon_n'(\mathbf{A}_{S,n} + \mathbf{A}_{S,n}')\varepsilon_n + \mathbf{a}_{S,n}'\varepsilon_n \end{bmatrix} \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_S). \tag{B.17}$$

Since the row and column sums of the matrices $A_{s,n}$ are uniformly bounded in absolute value, and since the elements of $\mathbf{a}_{s,n}$ and the variances are uniformly bounded by finite constants it is readily seen from (B.5) that the elements of $\mathbf{\Psi}_n^{\rho\rho}$, and hence those of $(\mathbf{\Psi}_n^{\rho\rho})^{1/2}$ are uniformly bounded. It now follows from (B.14), (B.15) and (B.17) that

$$n^{1/2}(\widetilde{\rho}_n - \rho_{0n}) = -\mathbf{\Xi}_n^{-1} \begin{bmatrix} 1 \\ 2\rho_{0n} \end{bmatrix}' \mathbf{\Gamma}_n' \mathbf{\Upsilon}_n (\mathbf{\Psi}_n^{\rho\rho})^{1/2} \xi_n^{\rho} + o_p(1). \tag{B.18}$$

Observing that $\mathbf{\Xi}_n = \mathbf{J}_n' \mathbf{\Upsilon}_n \mathbf{J}_n$, where $\mathbf{J}_n = \mathbf{\Gamma}_n[1, 2\rho_{0n}]'$, this establishes (B.8). Since all of the nonstochastic terms on the RHS of (B.18) are $O(1)$ it follows that $n^{1/2}(\widetilde{\rho}_n - \rho_{0n}) = O_p(1)$. Next recall that $0 < \lambda_{\Xi}^* \leq \mathbf{\Xi}_n \leq \lambda_{\Xi}^{**} < \infty$. Hence

$$\begin{aligned} \mathbf{\Omega}_n^{\rho\rho} &= \mathbf{\Xi}_n^{-1} \mathbf{J}_n' \mathbf{\Upsilon}_n \mathbf{\Psi}_n^{\rho\rho} \mathbf{\Upsilon}_n \mathbf{J}_n \mathbf{\Xi}_n^{-1} \\ &\geq \lambda_{\min}(\mathbf{\Psi}_n^{\rho\rho}) [\lambda_{\min}(\mathbf{\Upsilon}_n)]^2 \lambda_{\min}(\mathbf{\Gamma}_n'\mathbf{\Gamma}_n)(1 + 4\rho_{0n}^2)/(\lambda_{\Xi}^{**})^2 \geq const > 0. \end{aligned}$$

This establishes the last claim of the theorem. ■

Theorem B.3 ***(Theorem on Asymptotic Normality of*** $\widetilde{\widetilde{\theta}}_n = \left(\widetilde{\widetilde{\boldsymbol{\delta}}}_n', \widetilde{\rho}_n\right)'$***)*** *Let $\widetilde{\rho}_n$ be the weighted nonlinear least squares estimators defined by (B.1). Suppose Assumptions 1-5, 8, and B.1-B.3 hold,* $\sup_n |\beta_{0n}| < \infty$ *and* $\sup_n |\pi_{0n}| < \infty$, *and suppose additionally the smallest eigenvalues of $\mathbf{\Psi}_n$ and $\mathbf{P}_n^{\bullet\prime}\mathbf{P}_n^{\bullet}$ are bounded away from zero, then*

$$\begin{bmatrix} n^{1/2}(\widetilde{\widetilde{\boldsymbol{\delta}}}_n - \boldsymbol{\delta}_{0n}) \\ n^{1/2}(\widetilde{\rho}_n - \rho_{0n}) \end{bmatrix} = \begin{bmatrix} \mathbf{P}_n^{\bullet\prime} & 0 \\ 0 & (\mathbf{J}_n'\mathbf{\Upsilon}_n\mathbf{J}_n)^{-1}\mathbf{J}_n'\mathbf{\Upsilon}_n \end{bmatrix} \mathbf{\Psi}_n^{1/2} \xi_n + o_p(1),$$

where

$$\xi_n = \mathbf{\Psi}_n^{-1/2} \mathbf{v}_n \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{p^\bullet + S}).$$

Furthermore $n^{1/2}(\widetilde{\widetilde{\boldsymbol{\delta}}}_n - \boldsymbol{\delta}_{0n}) = O_p(1)$, $n^{1/2}(\widetilde{\rho}_n - \rho_{0n}) = O_p(1)$ *and*

$$\lambda_{\min}(\mathbf{\Omega}_n) = \lambda_{\min}\left\{ \begin{bmatrix} \mathbf{P}_n^{\bullet\prime} & 0 \\ 0 & (\mathbf{J}_n'\mathbf{\Upsilon}_n\mathbf{J}_n)^{-1}\mathbf{J}_n'\mathbf{\Upsilon}_n \end{bmatrix} \mathbf{\Psi}_n \begin{bmatrix} \mathbf{P}_n^{\bullet} & 0 \\ 0 & \mathbf{\Upsilon}_n\mathbf{J}_n(\mathbf{J}_n'\mathbf{\Upsilon}_n\mathbf{J}_n)^{-1} \end{bmatrix} \right\} \geq const > 0.$$

Proof.Theorem on Asymptotic Normality of $\widetilde{\widetilde{\theta}}_n = \left(\widetilde{\widetilde{\boldsymbol{\delta}}}'_n, \widetilde{\rho}_n\right)'$. By Assumption B.3 and Theorem B.2 we have

$$\begin{bmatrix} n^{1/2}(\widetilde{\widetilde{\boldsymbol{\delta}}}_n - \boldsymbol{\delta}_{0n}) \\ n^{1/2}(\widetilde{\rho}_n - \rho_{0n}) \end{bmatrix} = \begin{bmatrix} \mathbf{P}_n^{\bullet\prime}\mathbf{v}_n^{\delta} \\ (\mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1}\mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{v}_n^{\rho} \end{bmatrix} + o_p(1)$$

and hence clearly

$$\begin{bmatrix} n^{1/2}(\widetilde{\widetilde{\boldsymbol{\delta}}}_n - \boldsymbol{\delta}_{0n}) \\ n^{1/2}(\widetilde{\rho}_n - \rho_{0n}) \end{bmatrix} = \begin{bmatrix} \mathbf{P}_n^{\bullet\prime} & 0 \\ 0 & (\mathbf{J}'_n\Upsilon_n\mathbf{J}_n)^{-1}\mathbf{J}'_n\Upsilon_n \end{bmatrix} \boldsymbol{\Psi}_n^{1/2}\xi_n + o_p(1)$$

where $\xi_n = \boldsymbol{\Psi}_n^{-1/2}\mathbf{v}_n$. It is readily checked that the linear quadratic forms composing $\mathbf{v}_n$ satisfy all the assumptions of Theorem A.1 in Kelejian and Prucha (2010), and hence the claim that $\xi_n \xrightarrow{d} N(\mathbf{0}, \mathbf{I}_{p^\bullet+S})$ follows directly from that theorem.

Next observe that

$$\begin{aligned} \lambda_{\min}(\boldsymbol{\Omega}_n) &\geq \lambda_{\min}(\boldsymbol{\Psi}_n)\lambda_{\min}\left\{\begin{bmatrix} \mathbf{P}_n^{\bullet\prime}\mathbf{P}_n^{\bullet} & 0 \\ 0 & (\mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1}\mathbf{J}'_n\boldsymbol{\Upsilon}_n\boldsymbol{\Upsilon}_n\mathbf{J}_n(\mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1} \end{bmatrix}\right\} \\ &\geq \lambda_{\min}(\boldsymbol{\Psi}_n)\min\left\{\lambda_{\min}(\mathbf{P}_n^{\bullet\prime}\mathbf{P}_n^{\bullet}), \lambda_{\min}(\Theta_n)\right\} \end{aligned}$$

with $\Theta_n = (\mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1}\mathbf{J}'_n\boldsymbol{\Upsilon}_n\boldsymbol{\Upsilon}_n\mathbf{J}_n(\mathbf{J}'_n\boldsymbol{\Upsilon}_n\mathbf{J}_n)^{-1}$. Analogous as in the consistency proof of $\widetilde{\rho}_n$ we have

$$\lambda_{\min}(\Theta_n) \geq [\lambda_{\min}(\boldsymbol{\Upsilon}_n)]^2 \lambda_{\min}(\boldsymbol{\Gamma}'_n\boldsymbol{\Gamma}_n)(1 + 4\rho_{0n}^2)/(\lambda_{\Xi}^{**})^2 \geq const > 0.$$

Since $\lambda_{\min}(\boldsymbol{\Psi}_n)$ and $\lambda_{\min}(\mathbf{P}_n^{\bullet\prime}\mathbf{P}_n^{\bullet})$ are bounded away from zero by assumption it follows that $\lambda_{\min}(\boldsymbol{\Omega}_n) \geq const > 0$ as claimed. ■

C Appendix: Proofs of Consistency and Asymptotic Normality of Two-step IV/GMM Estimators

Lemma C.1 *: Suppose Assumptions 1-8 hold,* $\sup_n |\beta_{0n}| < \infty$ *and* $\sup_n |\pi_{0n}| < \infty$*, and suppose additionally that the smallest eigenvalues of* $\mathbf{\Psi}_n^{\rho\rho}$ *are bounded away from zero. Let* $\widetilde{\rho}_n$ *and* $\widehat{\rho}_n$ *be the GMM estimators for* ρ_{0n} *defined in (12) and (14), respectively. Then* $\widetilde{\rho}_n - \rho_{0n} = o_p(1)$ *and* $\widehat{\rho}_n - \rho_{0n} = o_p(1)$*.*

Proof. The proof utilizes Theorem B.1. Given the maintained assumptions it suffices to verify Assumptions B.1 and B.2 for the respective estimators.

(a) Proof of $\widetilde{\rho}_n - \rho_{0n} = o_p(1)$: The estimators $\widetilde{\rho}_n$ is based on S2SLS residuals. Assumption B.1 is clearly satisfied in light of Lemma A.3. Next observe that for $\widetilde{\rho}_n$ we have $\widetilde{\mathbf{\Upsilon}}_n = \mathbf{\Upsilon}_n = \mathbf{I}_S$. Thus also Assumption B.2 clearly holds, which completes the proof of the claim.

(b) Proof of $\widehat{\rho}_n - \rho_{0n} = o_p(1)$: The estimator $\widehat{\rho}_n$ is based on GS2SLS residuals. Assumption B.1 is clearly satisfied in light of Lemma A.4. Next observe that for $\widehat{\rho}_n$ we have $\widetilde{\mathbf{\Upsilon}}_n = \left[\widehat{\mathbf{\Psi}}_n^{\rho\rho}(\widehat{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)\right]^{-1}$ and $\mathbf{\Upsilon}_n = [\mathbf{\Psi}_n^{\rho\rho}]^{-1}$, where $\mathbf{\Psi}_n^{\rho\rho}$ is defined in (16). By assumption the smallest eigenvalues of $\mathbf{\Psi}_n^{\rho\rho}$ are bounded away from zero, and thus the largest eigenvalues of $\mathbf{\Upsilon}_n = [\mathbf{\Psi}_n^{\rho\rho}]^{-1}$ are bounded from above. We next show that the smallest eigenvalue of $\mathbf{\Upsilon}_n = [\mathbf{\Psi}_n^{\rho\rho}]^{-1}$ is bounded away from zero by showing that the largest eigenvalue of $\mathbf{\Psi}_n^{\rho\rho}$ is bounded from above. For that it suffices to show that the elements of $\mathbf{\Psi}_n^{\rho\rho}$ are uniformly bounded in n. First note that the elements of $\alpha_{r,n}$ are uniformly bounded in absolute value in light of part (b) of Lemma C.1 in Kelejian and Prucha (2010). Hence

$$n^{-1}\mathbf{a}'_{r,n}\mathbf{a}_{r,n} = n^{-1}\alpha'_{r,n}\mathbf{P}_n^{*\prime}\mathbf{H}'_n\mathbf{H}_n\mathbf{P}_n^{*}\alpha_{r,n} = O(1)$$

in light of Assumptions 6 and 7. Using the Cauchy-Schwartz inequality and taking into account Assumption 4 it is then readily seen that the elements of $\mathbf{\Psi}_n^{\rho\rho}$ are $O(1)$.

We next show that $\left[\widehat{\mathbf{\Psi}}_n^{\rho\rho}(\widehat{\boldsymbol{\delta}}_n, \widetilde{\rho}_n)\right]^{-1} - [\mathbf{\Psi}_n^{\rho\rho}]^{-1} = o_p(1)$. Utilizing Lemma F.1 in Kelejian and Prucha (1997) it suffices to show that $\widehat{\mathbf{\Psi}}_n^{\rho\rho} - \mathbf{\Psi}_n^{\rho\rho} = o_p(1)$. In light of Assumption 5 and Lemma C.1 in Kelejian and Prucha (2010) we have $\widehat{\alpha}_{r,n} - \alpha_{r,n} = o_p(1)$ and $\alpha_{r,n} = O(1)$. By Lemma A.4 we have furthermore $\widehat{\mathbf{P}}_n^{*} - \mathbf{P}_n^{*} = o_p(1)$ and $\mathbf{P}_n^{*} = O(1)$. Since $n^{-1}\mathbf{H}'_n\mathbf{H}_n = O(1)$ by Assumption 6 we have

$$\begin{aligned} n^{-1}\widehat{\mathbf{a}}'_{r,n}\widehat{\mathbf{a}}_{s,n} &= \widehat{\alpha}'_{r,n}\widehat{\mathbf{P}}_n^{*\prime}\left(n^{-1}\mathbf{H}'_n\mathbf{H}_n\right)\widehat{\mathbf{P}}_n^{*}\widehat{\alpha}_{s,n} = n^{-1}\alpha'_{r,n}\mathbf{P}_n^{*\prime}\mathbf{H}'_n\mathbf{H}_n\mathbf{P}_n^{*}\alpha_{s,n} + o_p(1) \\ &= n^{-1}\mathbf{a}'_{r,n}\mathbf{a}_{s,n} + o_p(1) \end{aligned}$$

and $n^{-1}\mathbf{a}'_{r,n}\mathbf{a}_{s,n} = O(1)$. Similarly

$$n^{-1}\widehat{\mathbf{a}}'_{r,n}vec_D(\mathbf{A}_{s,n}) = n^{-1}\mathbf{a}'_{r,n}vec_D(\mathbf{A}_{s,n}) + o_p(1)$$

and $n^{-1}\mathbf{a}'_{r,n}vec_D(\mathbf{A}_{s,n}) = O(1)$. Observing that $\widehat{\sigma}_n^2 = \sigma^2 + o_p(1)$, $\widehat{\mu}_n^{(3)} = \mu^3 + o_p(1)$, $\widehat{\mu}_n^{(4)} = \mu^4 + o_p(1)$ it now follows that $\widehat{\psi}_{rs,n}^{\rho\rho}(\widehat{\boldsymbol{\delta}}_n, \widetilde{\rho}_n) = \psi_{rs,n}^{\rho\rho} + o_p(1)$, which completes the proof. For later use we note that the above arguments also hold if $\widetilde{\rho}_n$ is replaced by any other consistent estimator. ■

Proof of Theorem 1: The proof utilizes Theorem B.3, which considers estimators $\widetilde{\widetilde{\boldsymbol{\delta}}}_n$ and $\widetilde{\rho}_n(\widetilde{\boldsymbol{\delta}}_n)$, where the latter is defined in (B.1). For the problem at hand both $\widetilde{\widetilde{\boldsymbol{\delta}}}_n$ and $\widetilde{\boldsymbol{\delta}}_n$ correspond to the GS2SLS estimator $\widehat{\boldsymbol{\delta}}_n$. Thus in this case Assumptions B.1 for $\widetilde{\boldsymbol{\delta}}_n$ and Assumption B.3 for $\widetilde{\widetilde{\boldsymbol{\delta}}}_n$ coincide. It has been verified in the proof of Lemma C.1 that the GS2SLS estimator $\widehat{\boldsymbol{\delta}}_n$ satisfies these assumptions. Furthermore, for the problem at hand $\widetilde{\boldsymbol{\Upsilon}}_n = \left(\widehat{\boldsymbol{\Psi}}_n^{\rho\rho}\right)^{-1}$ and $\boldsymbol{\Upsilon}_n = (\boldsymbol{\Psi}_n^{\rho\rho})^{-1}$ and it has also been verified in the proof of Lemma C.1 that those moment weights matrices satisfy Assumption B.2. Also, given that $\boldsymbol{\Upsilon}_n = (\boldsymbol{\Psi}_n^{\rho\rho})^{-1}$ it follows immediately from Theorem B.2 that $\widehat{\rho}_n$ is efficient. For $\mathbf{P}_n^*$ as defined in Lemma A.4 pertaining the the GS2SLS estimator we have $\mathbf{P}_n^{*\prime}\mathbf{P}_n^* = \left[\mathbf{Q}_{HZ*}(\rho_{0n})'\mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ*}(\rho_{0n})\right]^{-1}$ and thus $\lambda_{\min}(\mathbf{P}_n^{*\prime}\mathbf{P}_n^*) \geq c$ for some $c > 0$ by Assumption 7. Observing that all other assumptions utilized by Theorem B.3 concludes the proof. ■

Proof of Theorem 2: In the proof of Lemma C.1 we have demonstrated that $\widehat{\boldsymbol{\Psi}}_n^{\rho\rho} - \boldsymbol{\Psi}_n^{\rho\rho} = o_p(1)$, $\left[\widehat{\boldsymbol{\Psi}}_n^{\rho\rho}\right]^{-1} - [\boldsymbol{\Psi}_n^{\rho\rho}]^{-1} = o_p(1)$ and that $\boldsymbol{\Psi}_n^{\rho\rho} = O(1)$ and $[\boldsymbol{\Psi}_n^{\rho\rho}]^{-1} = O(1)$. By similar arguments we see that $\widehat{\boldsymbol{\Psi}}_n^{\delta\rho} - \boldsymbol{\Psi}_n^{\delta\rho} = o_p(1)$, $\widehat{\boldsymbol{\Psi}}_n^{\delta\delta} - \boldsymbol{\Psi}_n^{\delta\delta} = o_p(1)$, and $\boldsymbol{\Psi}_n^{\delta\rho} = O(1)$ and $\boldsymbol{\Psi}_n^{\delta\delta} = O(1)$. This establishes that $\widehat{\boldsymbol{\Psi}}_n - \boldsymbol{\Psi}_n = o_p(1)$ and $\boldsymbol{\Psi}_n = O(1)$. In the proof of Lemma C.1 it has been verified that the moment weights matrices $\widetilde{\boldsymbol{\Upsilon}}_n = \left(\widehat{\boldsymbol{\Psi}}_n^{\rho\rho}\right)^{-1}$ and $\boldsymbol{\Upsilon}_n = (\boldsymbol{\Psi}_n^{\rho\rho})^{-1}$ satisfy Assumption B.2. Hence it follows from the discussion around (B.12) and (B.13) in the proof of Theorem B.2 that $\left[\widehat{\mathbf{J}}_n'\left(\widehat{\boldsymbol{\Psi}}_n^{\rho\rho}\right)^{-1}\widehat{\mathbf{J}}_n\right]^{-1} - \left[\mathbf{J}_n'\left(\boldsymbol{\Psi}_n^{\rho\rho}\right)^{-1}\mathbf{J}_n\right]^{-1} = o_p(1)$ and $\left[\mathbf{J}_n'\left(\boldsymbol{\Psi}_n^{\rho\rho}\right)^{-1}\mathbf{J}_n\right]^{-1} = O(1)$. From the proof of Theorem B.1 we also have $\widehat{\mathbf{J}}_n - \mathbf{J}_n = o_p(1)$ and $\mathbf{J}_n = O(1)$, and by Lemma A.4 we have $\widehat{\mathbf{P}}_n^* - \mathbf{P}_n^* = o_p(1)$ and $\mathbf{P}_n^* = O(1)$. Given this it is readily seen from (18) that $\widehat{\boldsymbol{\Omega}}_n - \boldsymbol{\Omega}_n = o_p(1)$ and $\boldsymbol{\Omega}_n = O(1)$. Since $\lambda_{\min}(\boldsymbol{\Psi}_n) \geq \mathit{const} > 0$ by assumption, and $\lambda_{\min}(\boldsymbol{\Omega}_n) \geq \mathit{const} > 0$ by Theorem 1 it follows further that $\boldsymbol{\Psi}_n^{-1} = O(1)$ and $\boldsymbol{\Omega}_n^{-1} = O(1)$. ■

D Appendix: Proof of Claims in Text

Assumption 5 maintains that the endogenous regressors $\mathbf{Y}_n$ have finite uniformly bounded $2+\delta$ moments for some $\delta > 0$, and that

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

for any matrix $\mathbf{A}_n$ whose row and column sums are uniformly bounded in absolute value. In the following we verify the claim, made in the text after this assumption, that those properties are automatically implied if $\mathbf{y}_n$ and $\mathbf{Y}_n$ are generated by a simultaneous system of m equations with Cliff-Ord type spatial interactions as considered in Kelejian and Prucha (2004). For definiteness, let $\mathbf{y}_n^r$ denote the $n \times 1$ vector of observations of the r-th endogenous variable in the system, and w.l.o.g. let (1) represent the first equation in the system so that $\mathbf{y}_n = \mathbf{y}_n^1$. Under the assumptions maintained by Kelejian and Prucha (2004) it is readily seen that in reduced form

$$\mathbf{y}_n^r = \mathbf{c}_n^r + \mathbf{C}_n^r\mathbf{v}_n \tag{D.2}$$

where $\mathbf{c}_n^r$ is a $n \times 1$ nonstochastic vector with uniformly bounded element, $\mathbf{C}_n^r$ is a $n \times nm$ nonstochastic matrix whose row and column sums are uniformly bounded in absolute value, and $\mathbf{v}_n$ as an $nm \times 1$ vector of i.i.d. $(0, 1)$ innovations and finite fourth moments. (The vector $\mathbf{c}_n^r$ represents the reduced-form mean and is a function of the nonstochastic exogenous regressors and parameters.) Let $\mathbf{v}_n^1$ denote the first $n \times 1$ subvector of $\mathbf{v}_n$, then $\varepsilon_n = \sigma\mathbf{v}_n^1$.

Since the elements of $\mathbf{v}_n$ are i.i.d. with finite fourth moments it follows immediately from Lemma A.1 that for any δ, $0 < \delta \leq 2$ the $2 + \delta$ absolute moments of the elements of $\mathbf{y}_n^r$ and thus of the elements of $\mathbf{Y}_n$ are uniformly bounded as claimed.

Next consider

$$n^{-1}\mathbf{Z}_n'\mathbf{A}_n\mathbf{u}_n = n^{-1}\mathbf{Z}_n'\mathbf{A}_n\mathbf{u}_n = \sigma n^{-1}\mathbf{Z}_n'\mathbf{B}_n\mathbf{v}_n^1$$

with $\mathbf{B}_n = \mathbf{A}_n(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}$. Since by assumption row and column sums of the absolute elements of $(\mathbf{I}_n - \rho_{0n}\mathbf{M}_n)^{-1}$ are uniformly bounded, it follows that $\mathbf{B}_n$ has the same property. The elements of $n^{-1}\mathbf{Z}_n'\mathbf{A}_n\mathbf{u}_n$ are now seen to be of one of the following kind:

$$\begin{aligned}
\varsigma_{1n} &= n^{-1}\mathbf{x}_n^{k\prime}\mathbf{B}_n\mathbf{v}_n^1, \\
\varsigma_{2n} &= n^{-1}\mathbf{y}_n^{r\prime}\mathbf{B}_n\mathbf{v}_n^1 = n^{-1}\mathbf{c}_n^{r\prime}\mathbf{B}_n\mathbf{v}_n^1 + n^{-1}\mathbf{v}_n'\mathbf{C}_n^{r\prime}\mathbf{B}_n\mathbf{v}_n^1, \\
\varsigma_{3n} &= n^{-1}\mathbf{y}_n'\mathbf{W}_n'\mathbf{B}_n\mathbf{v}_n^1 = n^{-1}\mathbf{c}_n^{1\prime}\mathbf{W}_n'\mathbf{B}_n\mathbf{v}_n^1 + n^{-1}\mathbf{v}_n'\mathbf{C}_n^{1\prime}\mathbf{W}_n'\mathbf{B}_n\mathbf{v}_n^1.
\end{aligned} \tag{D.3}$$

The mean of ς_{jn} clearly exists. Thus to prove (D.1) it suffices to show that the variances of ς_{in} are $o(1)$. Since the covariances can be bounded by variances it suffices to show that the variances of each of the terms on the RHS of (D.3) are $o_p(1)$. In light of, e.g., Remark A.1 in Kelejian and Prucha (2004) it follows that each of the first terms on the RHS of (D.3) are of the form $n^{-1}\mathbf{d}_n'\mathbf{v}_n^1$ where the elements of $\mathbf{d}_n$ are uniformly bounded by some finite constant, say K, and hence $var(n^{-1}\mathbf{d}_n'\mathbf{v}_n^1) \leq n^{-1}K^2 = o(1)$. In light of, e.g., Remark A.1 in Kelejian and Prucha (2004) it follows furthermore that each of the second terms on the RHS of (D.3) are of the form

$$n^{-1}\sum_{j=1}^{m}\mathbf{v}_n^{j\prime}\mathbf{D}_n^j\mathbf{v}_n^1 = n^{-1}\mathbf{v}_n'\mathbf{D}_n\mathbf{v}_n, \qquad \mathbf{D}_n = \begin{bmatrix} \mathbf{D}_1 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \mathbf{D}_m & 0 & \dots & 0 \end{bmatrix}$$

where the row and column sums of the $n \times n$ matrices $\mathbf{D}_n^j$ and thus those of the matrix $\mathbf{D}_n$ are uniformly bounded by some finite constant. In light of, e.g., Lemma A.1 in Kelejian and Prucha (2004) we have $var(n^{-1}\mathbf{v}_n'\mathbf{D}_n\mathbf{v}_n) = n^{-2}tr\left\{[\mathbf{D}_n + \mathbf{D}_n'][\mathbf{D}_n + \mathbf{D}_n']\right\}/4$. Referring again to Remark A.1 Kelejian and Prucha (2004) we see that also the row and column sums of the absolute elements of $[\mathbf{D}_n + \mathbf{D}_n'][\mathbf{D}_n + \mathbf{D}_n']$ are uniformly bounded by some finite constant, again say K. Thus $var(n^{-1}\mathbf{v}_n'\mathbf{D}_n\mathbf{v}_n) \leq n^{-1}K/4 = o(1)$, which completes the proof of the claim.

E Tables

E.1 Results for $\widehat{\beta}$

This subsection contains the tables of results for $\widehat{\beta}$ when the true value of $\beta = 2$.

Table 1: $\widehat{\beta}$ results when ρ=-0.8

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	2.001	2.003	.046	.091	.046	.091	.05	.046	.812	.52
-.8	2	2.000	1.998	.034	.067	.033	.066	.056	.05	.803	.503
-.8	3	2.003	2.011	.049	.101	.049	.098	.045	.059	.757	.44
-.8	4	2.002	2.006	.036	.071	.035	.07	.058	.053	.752	.433
-.3	1	1.999	2.003	.046	.095	.047	.094	.046	.052	.835	.559
-.3	2	2.000	2.002	.035	.067	.034	.068	.058	.044	.827	.544
-.3	3	2.001	2.008	.052	.106	.052	.104	.051	.056	.8	.5
-.3	4	2.002	2.004	.037	.072	.037	.073	.054	.048	.793	.49
0	1	2.001	1.999	.049	.097	.048	.096	.054	.056	.847	.581
0	2	1.999	2.003	.035	.069	.034	.069	.05	.048	.838	.565
0	3	2.003	2.007	.054	.106	.053	.106	.053	.049	.822	.536
0	4	1.999	2.001	.038	.073	.037	.075	.056	.047	.813	.521
.3	1	2.000	2.003	.049	.095	.048	.097	.05	.045	.859	.605
.3	2	1.999	1.998	.034	.071	.035	.069	.044	.054	.85	.586
.3	3	2.000	2.002	.054	.107	.053	.107	.058	.051	.842	.572
.3	4	2.000	2.003	.038	.076	.038	.076	.051	.05	.832	.552
.8	1	1.998	1.998	.05	.099	.049	.097	.053	.052	.884	.657
.8	2	2.000	1.999	.035	.071	.035	.07	.043	.048	.873	.631
.8	3	2.002	1.995	.054	.107	.054	.108	.056	.048	.88	.645
.8	4	2.002	2.000	.038	.079	.038	.076	.054	.065	.863	.613

Table 2: $\widehat{\beta}$ results when ρ=-0.3

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	2.001	1.996	.042	.081	.042	.083	.053	.049	.854	.595
-.8	2	2.000	1.999	.031	.061	.03	.061	.05	.048	.845	.577
-.8	3	1.999	1.999	.042	.082	.041	.083	.056	.049	.848	.583
-.8	4	1.998	1.999	.03	.059	.03	.06	.048	.046	.843	.574
-.3	1	1.999	1.998	.043	.086	.043	.086	.051	.054	.862	.609
-.3	2	2.000	1.998	.031	.065	.031	.063	.054	.057	.853	.592
-.3	3	2.001	1.998	.045	.087	.044	.087	.054	.048	.858	.603
-.3	4	1.998	1.997	.032	.062	.031	.063	.052	.042	.852	.59
0	1	2.002	1.999	.044	.089	.044	.088	.054	.056	.868	.621
0	2	1.998	1.998	.032	.065	.032	.064	.047	.053	.858	.603
0	3	1.998	1.997	.046	.09	.045	.09	.053	.05	.866	.618
0	4	1.999	1.996	.032	.064	.032	.065	.048	.045	.858	.601
.3	1	2.001	1.997	.045	.091	.045	.089	.051	.052	.874	.635
.3	2	1.999	1.997	.033	.065	.032	.065	.056	.053	.865	.616
.3	3	1.999	1.999	.047	.094	.047	.093	.052	.048	.875	.636
.3	4	2.000	1.997	.033	.066	.033	.067	.05	.042	.865	.616
.8	1	1.998	1.999	.046	.092	.046	.092	.062	.052	.892	.673
.8	2	1.999	1.996	.033	.065	.033	.066	.052	.049	.88	.647
.8	3	2.002	2.002	.049	.099	.048	.097	.051	.06	.896	.682
.8	4	1.998	1.996	.035	.068	.034	.069	.057	.051	.881	.648

Table 3: $\widehat{\beta}$ results when ρ= 0.0

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	1.999	1.994	.04	.082	.041	.082	.044	.052	.861	.607
-.8	2	2.001	1.998	.03	.059	.03	.06	.051	.047	.851	.589
-.8	3	2.000	1.998	.041	.081	.041	.082	.048	.052	.86	.605
-.8	4	1.999	1.998	.03	.061	.03	.06	.052	.05	.855	.596
-.3	1	1.999	1.996	.042	.082	.042	.083	.054	.049	.865	.616
-.3	2	1.999	1.999	.03	.062	.03	.061	.05	.054	.857	.599
-.3	3	1.998	1.995	.04	.082	.041	.083	.046	.053	.864	.614
-.3	4	1.999	1.998	.031	.061	.03	.061	.056	.052	.858	.602
0	1	2.000	1.995	.042	.084	.042	.085	.05	.055	.87	.624
0	2	2.000	2.000	.031	.062	.031	.062	.048	.053	.861	.607
0	3	1.999	1.998	.043	.085	.042	.085	.054	.051	.869	.624
0	4	1.999	1.998	.031	.061	.031	.061	.045	.049	.862	.609
.3	1	1.997	1.995	.043	.086	.043	.086	.055	.049	.875	.636
.3	2	1.998	1.997	.031	.063	.031	.063	.044	.049	.865	.617
.3	3	1.999	2.000	.043	.087	.043	.087	.048	.051	.876	.639
.3	4	2.000	1.996	.031	.062	.031	.062	.054	.056	.866	.618
.8	1	2.000	1.995	.045	.089	.044	.088	.053	.053	.89	.668
.8	2	1.998	2.000	.032	.064	.032	.064	.048	.047	.878	.643
.8	3	2.000	1.996	.046	.092	.045	.09	.055	.06	.893	.674
.8	4	2.000	2.001	.032	.065	.032	.064	.046	.056	.878	.643

Table 4: $\widehat{\beta}$ results when ρ= 0.3

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	2.001	1.997	.043	.088	.042	.085	.05	.06	.863	.611
-.8	2	2.000	1.997	.031	.062	.031	.062	.051	.052	.853	.593
-.8	3	1.997	1.997	.044	.085	.044	.088	.053	.048	.863	.611
-.8	4	1.999	2.001	.033	.064	.032	.064	.058	.047	.859	.603
-.3	1	2.000	1.993	.042	.086	.042	.084	.048	.055	.865	.615
-.3	2	1.999	1.998	.031	.06	.031	.061	.048	.046	.856	.598
-.3	3	2.000	1.993	.044	.087	.043	.085	.058	.061	.863	.612
-.3	4	1.999	1.998	.031	.063	.031	.063	.046	.052	.857	.601
0	1	1.998	1.997	.043	.085	.042	.084	.059	.054	.867	.62
0	2	1.998	1.999	.031	.062	.031	.061	.061	.054	.859	.603
0	3	1.997	1.993	.042	.083	.042	.085	.049	.046	.866	.617
0	4	1.999	1.999	.031	.061	.031	.062	.049	.051	.858	.602
.3	1	1.998	1.992	.042	.084	.042	.084	.056	.058	.871	.628
.3	2	2.000	1.998	.031	.062	.031	.061	.045	.05	.862	.609
.3	3	1.998	1.994	.042	.084	.042	.084	.049	.05	.87	.625
.3	4	1.998	1.999	.03	.061	.031	.062	.043	.049	.86	.606
.8	1	1.999	1.994	.042	.084	.043	.085	.047	.048	.883	.653
.8	2	1.999	1.996	.031	.061	.031	.062	.043	.044	.871	.628
.8	3	1.998	1.992	.043	.085	.043	.085	.052	.05	.883	.654
.8	4	1.999	1.998	.03	.062	.031	.062	.044	.055	.867	.62

Table 5: $\widehat{\beta}$ results when ρ= 0.8

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	1.999	2.007	.05	.102	.05	.099	.053	.058	.846	.58
-.8	2	2.001	1.999	.035	.07	.035	.07	.049	.046	.836	.561
-.8	3	2.001	2.006	.054	.109	.054	.108	.058	.055	.841	.568
-.8	4	2.000	1.997	.039	.077	.038	.076	.055	.055	.836	.561
-.3	1	2.004	2.009	.05	.098	.049	.098	.055	.053	.839	.565
-.3	2	2.001	2.004	.036	.072	.035	.069	.056	.058	.829	.549
-.3	3	2.001	2.006	.054	.107	.053	.106	.048	.058	.826	.544
-.3	4	2.000	2.006	.038	.076	.038	.075	.052	.053	.819	.531
0	1	2.001	2.009	.05	.098	.048	.096	.059	.057	.834	.557
0	2	2.002	2.002	.035	.069	.034	.068	.056	.05	.824	.54
0	3	2.003	2.011	.052	.103	.052	.104	.05	.046	.819	.53
0	4	1.999	2.005	.037	.073	.037	.074	.048	.044	.809	.513
.3	1	2.002	2.005	.047	.095	.047	.095	.046	.046	.829	.548
.3	2	2.002	2.007	.034	.069	.034	.067	.056	.056	.818	.53
.3	3	2.006	2.013	.053	.101	.051	.101	.058	.05	.812	.518
.3	4	2.002	2.003	.036	.073	.036	.072	.05	.056	.798	.496
.8	1	2.001	2.008	.046	.092	.046	.091	.051	.052	.821	.534
.8	2	2.000	2.002	.033	.065	.033	.065	.052	.048	.806	.509
.8	3	2.006	2.005	.048	.096	.048	.095	.057	.049	.802	.503
.8	4	1.999	2.004	.035	.069	.034	.069	.058	.05	.778	.468

E.2 Results for $\widehat{\pi}$

This subsection contains the tables of results for $\widehat{\pi}$ when the true value of $\pi = 1$.

Table 6: $\widehat{\pi}$ results when ρ=-0.8

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	1.000	0.994	.031	.064	.032	.063	.047	.054	.812	.52
-.8	2	1.000	0.997	.023	.045	.023	.045	.06	.056	.803	.503
-.8	3	1.000	0.999	.033	.067	.033	.066	.056	.054	.757	.44
-.8	4	1.001	1.001	.024	.047	.024	.047	.05	.056	.752	.433
-.3	1	0.999	0.996	.032	.066	.032	.064	.058	.058	.835	.559
-.3	2	0.999	1.000	.023	.046	.023	.046	.046	.052	.827	.544
-.3	3	1.000	1.000	.034	.067	.033	.067	.06	.048	.8	.5
-.3	4	1.001	1.000	.024	.047	.024	.048	.055	.051	.793	.49
0	1	0.999	0.997	.033	.066	.032	.064	.059	.055	.847	.581
0	2	1.000	1.000	.023	.047	.023	.046	.051	.056	.838	.565
0	3	1.000	0.999	.034	.067	.033	.067	.055	.047	.822	.536
0	4	0.999	1.000	.025	.048	.024	.049	.054	.048	.813	.521
.3	1	0.999	0.996	.032	.065	.032	.064	.051	.052	.859	.605
.3	2	0.999	0.998	.023	.046	.023	.046	.045	.053	.85	.586
.3	3	1.001	0.997	.034	.066	.033	.067	.058	.045	.842	.572
.3	4	1.000	0.999	.024	.049	.024	.049	.054	.05	.832	.552
.8	1	0.998	0.995	.033	.063	.032	.064	.057	.047	.884	.657
.8	2	1.000	0.998	.023	.047	.023	.047	.055	.054	.873	.631
.8	3	1.000	0.995	.034	.066	.033	.067	.055	.048	.88	.645
.8	4	1.001	0.999	.024	.051	.024	.049	.046	.064	.863	.613

Table 7: $\widehat{\pi}$ results when ρ=-0.3

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	1.000	0.997	.033	.064	.032	.065	.056	.046	.854	.595
-.8	2	1.000	0.997	.023	.046	.023	.045	.052	.057	.845	.577
-.8	3	0.999	0.997	.033	.067	.033	.065	.059	.058	.848	.583
-.8	4	0.998	0.999	.023	.046	.023	.046	.049	.053	.843	.574
-.3	1	0.999	0.995	.032	.064	.032	.065	.044	.055	.862	.609
-.3	2	1.000	0.999	.023	.046	.023	.046	.056	.053	.853	.592
-.3	3	0.999	1.000	.033	.066	.033	.066	.048	.049	.858	.603
-.3	4	0.999	0.998	.023	.046	.023	.047	.044	.049	.852	.59
0	1	1.000	0.997	.033	.065	.033	.065	.06	.051	.868	.621
0	2	0.998	0.998	.023	.046	.023	.046	.044	.052	.858	.603
0	3	1.000	0.996	.034	.066	.033	.066	.052	.052	.866	.618
0	4	0.999	0.997	.023	.047	.023	.047	.04	.052	.858	.601
.3	1	0.999	0.994	.033	.065	.033	.065	.05	.052	.874	.635
.3	2	1.000	0.998	.024	.046	.023	.046	.067	.049	.865	.616
.3	3	0.998	0.997	.033	.066	.033	.066	.055	.057	.875	.636
.3	4	0.999	0.998	.024	.047	.024	.047	.053	.051	.865	.616
.8	1	0.999	0.996	.033	.065	.033	.066	.055	.048	.892	.673
.8	2	1.000	0.997	.023	.046	.023	.046	.047	.05	.88	.647
.8	3	0.999	0.997	.034	.066	.033	.067	.05	.048	.896	.682
.8	4	0.999	0.997	.024	.047	.024	.048	.054	.047	.881	.648

Table 8: $\widehat{\pi}$ results when ρ= 0.0

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	0.999	0.994	.033	.067	.033	.066	.051	.054	.861	.607
-.8	2	0.998	0.999	.023	.045	.023	.046	.045	.052	.851	.589
-.8	3	0.999	0.998	.033	.067	.033	.066	.05	.054	.86	.605
-.8	4	0.998	0.998	.024	.047	.024	.047	.055	.052	.855	.596
-.3	1	0.999	0.997	.033	.066	.033	.066	.055	.05	.865	.616
-.3	2	0.999	0.997	.023	.046	.023	.046	.051	.054	.857	.599
-.3	3	0.999	0.997	.033	.065	.033	.066	.048	.052	.864	.614
-.3	4	0.999	0.996	.024	.047	.023	.047	.063	.049	.858	.602
0	1	0.998	0.997	.033	.066	.033	.066	.051	.053	.87	.624
0	2	1.000	1.000	.023	.047	.023	.046	.052	.056	.861	.607
0	3	0.998	0.999	.033	.066	.033	.066	.054	.049	.869	.624
0	4	0.999	0.999	.023	.047	.023	.047	.049	.05	.862	.609
.3	1	0.999	0.995	.034	.066	.033	.066	.055	.053	.875	.636
.3	2	0.999	0.998	.023	.047	.023	.046	.05	.053	.865	.617
.3	3	0.998	0.997	.033	.066	.033	.066	.053	.052	.876	.639
.3	4	1.000	0.996	.023	.046	.023	.047	.048	.052	.866	.618
.8	1	0.998	0.996	.034	.068	.033	.066	.057	.055	.89	.668
.8	2	0.999	0.999	.023	.045	.023	.046	.049	.05	.878	.643
.8	3	1.000	0.997	.033	.065	.033	.066	.048	.048	.893	.674
.8	4	0.999	0.998	.023	.047	.023	.047	.048	.057	.878	.643

Table 9: $\widehat{\pi}$ results when ρ= 0.3

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	1.000	0.997	.033	.069	.033	.067	.048	.057	.863	.611
-.8	2	1.001	1.000	.023	.046	.023	.046	.052	.052	.853	.593
-.8	3	1.000	0.993	.035	.067	.034	.068	.061	.049	.863	.611
-.8	4	0.999	0.998	.024	.049	.024	.049	.049	.05	.859	.603
-.3	1	0.998	0.996	.033	.069	.033	.067	.051	.064	.865	.615
-.3	2	0.999	0.998	.024	.046	.023	.046	.054	.049	.856	.598
-.3	3	1.000	0.994	.033	.068	.033	.067	.056	.056	.863	.612
-.3	4	1.000	1.000	.024	.048	.024	.048	.05	.053	.857	.601
0	1	0.998	0.996	.033	.067	.033	.066	.047	.056	.867	.62
0	2	1.000	0.997	.023	.047	.023	.046	.052	.052	.859	.603
0	3	0.999	0.995	.033	.067	.033	.066	.053	.059	.866	.617
0	4	0.999	0.998	.023	.047	.024	.047	.052	.047	.858	.602
.3	1	0.999	0.994	.033	.067	.033	.066	.054	.053	.871	.628
.3	2	0.999	0.999	.023	.046	.023	.046	.061	.053	.862	.609
.3	3	0.999	0.994	.033	.066	.033	.066	.049	.052	.87	.625
.3	4	0.999	0.998	.023	.046	.023	.047	.048	.045	.86	.606
.8	1	0.998	0.994	.034	.066	.033	.066	.054	.054	.883	.653
.8	2	0.999	0.999	.023	.047	.023	.046	.046	.054	.871	.628
.8	3	0.999	0.994	.034	.065	.033	.065	.06	.055	.883	.654
.8	4	0.999	0.997	.023	.046	.023	.046	.05	.048	.867	.62

Table 10: $\widehat{\pi}$ results when ρ= 0.8

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	0.999	1.000	.034	.069	.034	.068	.048	.05	.846	.58
-.8	2	1.000	0.997	.024	.047	.024	.048	.051	.046	.836	.561
-.8	3	1.000	0.998	.035	.07	.035	.07	.054	.047	.841	.568
-.8	4	0.999	0.999	.026	.053	.026	.052	.052	.057	.836	.561
-.3	1	1.001	0.998	.034	.069	.034	.068	.054	.042	.839	.565
-.3	2	1.000	0.999	.024	.049	.024	.048	.055	.06	.829	.549
-.3	3	1.000	0.999	.035	.07	.035	.07	.054	.053	.826	.544
-.3	4	1.000	1.001	.025	.051	.026	.051	.046	.051	.819	.531
0	1	0.999	0.996	.034	.07	.034	.068	.05	.058	.834	.557
0	2	1.001	1.001	.024	.047	.024	.047	.049	.053	.824	.54
0	3	1.000	1.000	.035	.067	.035	.069	.052	.04	.819	.53
0	4	1.000	1.000	.025	.05	.025	.051	.052	.048	.809	.513
.3	1	1.000	0.997	.033	.067	.034	.068	.044	.047	.829	.548
.3	2	1.000	1.002	.023	.047	.023	.047	.052	.053	.818	.53
.3	3	1.002	1.000	.035	.068	.034	.069	.057	.051	.812	.518
.3	4	1.000	1.001	.025	.05	.025	.05	.045	.05	.798	.496
.8	1	1.001	1.001	.034	.067	.034	.067	.058	.05	.821	.534
.8	2	1.000	1.001	.024	.047	.023	.046	.059	.057	.806	.509
.8	3	1.003	0.997	.033	.067	.033	.067	.052	.046	.802	.503
.8	4	1.000	1.002	.024	.049	.024	.048	.048	.055	.778	.468

E.3 Results for $\widehat{\rho}$

This subsection contains the tables of results for $\widehat{\rho}$. The true value of ρ is given in the title of each table.

Table 11: $\widehat{\rho}$ results when ρ=-0.8

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	-0.798	-0.789	.062	.082	.062	.071	.04	.06	.812	.52
-.8	2	-0.801	-0.796	.045	.054	.044	.05	.051	.052	.803	.503
-.8	3	-0.796	-0.793	.04	.052	.04	.048	.051	.053	.757	.44
-.8	4	-0.799	-0.796	.029	.037	.029	.035	.056	.057	.752	.433
-.3	1	-0.798	-0.796	.061	.075	.06	.066	.047	.044	.835	.559
-.3	2	-0.798	-0.798	.043	.048	.043	.047	.05	.05	.827	.544
-.3	3	-0.798	-0.796	.038	.044	.038	.043	.051	.053	.8	.5
-.3	4	-0.799	-0.798	.027	.032	.028	.031	.044	.056	.793	.49
0	1	-0.800	-0.796	.063	.072	.06	.064	.052	.049	.847	.581
0	2	-0.798	-0.798	.042	.047	.042	.046	.051	.044	.838	.565
0	3	-0.799	-0.797	.038	.043	.038	.041	.047	.052	.822	.536
0	4	-0.800	-0.798	.027	.03	.027	.03	.052	.046	.813	.521
.3	1	-0.801	-0.796	.061	.071	.06	.063	.042	.04	.859	.605
.3	2	-0.801	-0.798	.042	.046	.042	.045	.041	.055	.85	.586
.3	3	-0.800	-0.798	.037	.042	.037	.04	.049	.062	.842	.572
.3	4	-0.800	-0.800	.027	.029	.027	.029	.048	.047	.832	.552
.8	1	-0.801	-0.799	.059	.07	.059	.061	.046	.054	.884	.657
.8	2	-0.800	-0.799	.041	.044	.042	.044	.044	.044	.873	.631
.8	3	-0.799	-0.798	.038	.039	.037	.038	.056	.047	.88	.645
.8	4	-0.800	-0.800	.027	.028	.027	.028	.051	.056	.863	.613

Table 12: $\widehat{\rho}$ results when ρ=-0.3

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	-0.300	-0.299	.093	.109	.094	.111	.049	.052	.854	.595
-.8	2	-0.300	-0.303	.068	.08	.068	.083	.051	.044	.845	.577
-.8	3	-0.298	-0.297	.071	.086	.071	.086	.049	.054	.848	.583
-.8	4	-0.300	-0.295	.051	.064	.051	.064	.049	.05	.843	.574
-.3	1	-0.304	-0.298	.092	.106	.094	.107	.053	.055	.862	.609
-.3	2	-0.304	-0.301	.068	.079	.067	.079	.057	.052	.853	.592
-.3	3	-0.298	-0.295	.07	.08	.07	.083	.052	.045	.858	.603
-.3	4	-0.302	-0.299	.052	.06	.051	.061	.055	.052	.852	.59
0	1	-0.306	-0.302	.09	.103	.093	.105	.048	.047	.868	.621
0	2	-0.301	-0.302	.067	.076	.067	.077	.049	.051	.858	.603
0	3	-0.303	-0.301	.069	.08	.069	.08	.053	.051	.866	.618
0	4	-0.301	-0.297	.051	.059	.05	.059	.056	.056	.858	.601
.3	1	-0.305	-0.308	.092	.103	.092	.102	.064	.059	.874	.635
.3	2	-0.302	-0.299	.067	.075	.066	.075	.062	.058	.865	.616
.3	3	-0.305	-0.299	.068	.078	.069	.078	.05	.056	.875	.636
.3	4	-0.300	-0.299	.049	.058	.05	.058	.048	.051	.865	.616
.8	1	-0.309	-0.302	.09	.099	.091	.098	.052	.063	.892	.673
.8	2	-0.304	-0.304	.065	.071	.065	.072	.055	.048	.88	.647
.8	3	-0.302	-0.300	.067	.074	.068	.074	.05	.049	.896	.682
.8	4	-0.301	-0.304	.049	.054	.049	.054	.053	.05	.881	.648

Table 13: $\widehat{\rho}$ results when ρ= 0.0

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	-0.007	-0.003	.095	.11	.094	.11	.054	.06	.861	.607
-.8	2	-0.001	-0.005	.067	.085	.068	.082	.053	.072	.851	.589
-.8	3	-0.001	-0.002	.074	.086	.074	.088	.056	.055	.86	.605
-.8	4	-0.000	0.003	.054	.064	.054	.065	.06	.047	.855	.596
-.3	1	-0.007	-0.005	.093	.107	.094	.109	.048	.053	.865	.616
-.3	2	-0.002	-0.002	.065	.081	.067	.081	.048	.051	.857	.599
-.3	3	-0.002	-0.004	.074	.086	.074	.087	.051	.047	.864	.614
-.3	4	-0.000	-0.000	.055	.065	.054	.065	.062	.054	.858	.602
0	1	-0.005	-0.004	.091	.107	.093	.108	.052	.05	.87	.624
0	2	-0.004	-0.002	.068	.08	.067	.08	.052	.056	.861	.607
0	3	-0.004	-0.003	.072	.085	.074	.086	.049	.049	.869	.624
0	4	0.001	-0.003	.054	.065	.054	.064	.051	.056	.862	.609
.3	1	-0.005	-0.005	.092	.105	.093	.106	.054	.05	.875	.636
.3	2	-0.004	-0.005	.066	.08	.067	.078	.052	.059	.865	.617
.3	3	-0.002	-0.003	.073	.085	.073	.085	.055	.053	.876	.639
.3	4	-0.000	0.000	.053	.062	.053	.063	.055	.047	.866	.618
.8	1	-0.009	-0.006	.094	.099	.092	.102	.062	.046	.89	.668
.8	2	-0.004	-0.007	.067	.077	.066	.075	.054	.057	.878	.643
.8	3	-0.007	-0.005	.074	.08	.072	.082	.058	.049	.893	.674
.8	4	-0.001	0.000	.053	.062	.053	.061	.052	.054	.878	.643

Table 14: $\widehat{\rho}$ results when ρ= 0.3

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	0.296	0.294	.083	.093	.083	.095	.054	.047	.863	.611
-.8	2	0.299	0.297	.058	.07	.06	.07	.046	.054	.853	.593
-.8	3	0.298	0.299	.066	.076	.068	.077	.048	.051	.863	.611
-.8	4	0.300	0.296	.049	.056	.049	.056	.056	.052	.859	.603
-.3	1	0.295	0.290	.083	.098	.083	.097	.051	.054	.865	.615
-.3	2	0.293	0.297	.06	.072	.06	.072	.052	.058	.856	.598
-.3	3	0.295	0.292	.069	.082	.068	.079	.057	.061	.863	.612
-.3	4	0.298	0.295	.049	.057	.049	.059	.047	.049	.857	.601
0	1	0.294	0.293	.082	.1	.083	.097	.047	.06	.867	.62
0	2	0.298	0.295	.06	.073	.06	.072	.055	.053	.859	.603
0	3	0.299	0.295	.069	.082	.068	.08	.059	.056	.866	.617
0	4	0.300	0.299	.05	.06	.049	.059	.051	.054	.858	.602
.3	1	0.291	0.291	.083	.1	.083	.097	.05	.062	.871	.628
.3	2	0.295	0.294	.06	.074	.06	.072	.05	.058	.862	.609
.3	3	0.297	0.296	.069	.08	.068	.08	.055	.053	.87	.625
.3	4	0.298	0.295	.049	.06	.05	.06	.049	.047	.86	.606
.8	1	0.288	0.290	.083	.095	.083	.096	.049	.055	.883	.653
.8	2	0.297	0.296	.059	.072	.06	.071	.054	.052	.871	.628
.8	3	0.297	0.294	.067	.078	.068	.079	.048	.046	.883	.654
.8	4	0.298	0.295	.05	.061	.049	.06	.056	.055	.867	.62

Table 15: $\widehat{\rho}$ results when ρ= 0.8

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	0.799	0.796	.042	.047	.042	.045	.046	.052	.846	.58
-.8	2	0.798	0.799	.03	.033	.03	.033	.047	.05	.836	.561
-.8	3	0.801	0.799	.035	.038	.035	.037	.053	.048	.841	.568
-.8	4	0.800	0.798	.026	.027	.025	.027	.059	.049	.836	.561
-.3	1	0.797	0.793	.044	.053	.043	.048	.051	.056	.839	.565
-.3	2	0.799	0.797	.031	.036	.031	.035	.052	.056	.829	.549
-.3	3	0.798	0.795	.037	.042	.036	.04	.056	.062	.826	.544
-.3	4	0.799	0.799	.026	.029	.026	.028	.046	.055	.819	.531
0	1	0.796	0.792	.046	.056	.043	.05	.058	.06	.834	.557
0	2	0.798	0.796	.032	.037	.031	.036	.054	.046	.824	.54
0	3	0.797	0.795	.036	.045	.036	.041	.049	.056	.819	.53
0	4	0.798	0.798	.027	.03	.026	.029	.057	.048	.809	.513
.3	1	0.797	0.790	.046	.062	.044	.053	.049	.065	.829	.548
.3	2	0.798	0.793	.033	.041	.032	.038	.05	.061	.818	.53
.3	3	0.797	0.792	.038	.046	.037	.043	.053	.055	.812	.518
.3	4	0.799	0.798	.026	.032	.027	.031	.048	.058	.798	.496
.8	1	0.792	0.785	.048	.067	.046	.058	.06	.072	.821	.534
.8	2	0.796	0.794	.034	.045	.033	.042	.052	.064	.806	.509
.8	3	0.797	0.788	.039	.053	.038	.047	.058	.071	.802	.503
.8	4	0.798	0.795	.028	.036	.028	.034	.049	.062	.778	.468

E.4 Results for $\widehat{\lambda}$

This subsection contains the tables of results for $\widehat{\lambda}$. The true value of λ is given in each table.

Table 16: $\widehat{\lambda}$ results when ρ=-0.8

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	-0.804	-0.814	.052	.109	.052	.103	.05	.066	.812	.52
-.8	2	-0.801	-0.803	.04	.083	.04	.079	.056	.061	.803	.503
-.8	3	-0.806	-0.820	.052	.107	.052	.103	.054	.067	.757	.44
-.8	4	-0.803	-0.809	.04	.082	.039	.078	.053	.064	.752	.433
-.3	1	-0.302	-0.310	.042	.089	.043	.085	.047	.062	.835	.559
-.3	2	-0.301	-0.302	.034	.067	.033	.066	.061	.058	.827	.544
-.3	3	-0.302	-0.309	.044	.088	.043	.086	.049	.053	.8	.5
-.3	4	-0.301	-0.305	.033	.066	.033	.065	.049	.053	.793	.49
0	1	-0.001	-0.003	.037	.076	.037	.073	.056	.058	.847	.581
0	2	-0.002	-0.002	.028	.057	.028	.057	.046	.051	.838	.565
0	3	-0.002	-0.004	.038	.077	.037	.074	.056	.06	.822	.536
0	4	0.000	-0.006	.028	.056	.028	.056	.06	.049	.813	.521
.3	1	0.300	0.295	.031	.063	.031	.062	.053	.058	.859	.605
.3	2	0.301	0.299	.024	.05	.024	.048	.048	.056	.85	.586
.3	3	0.299	0.297	.031	.063	.031	.061	.052	.056	.842	.572
.3	4	0.300	0.297	.023	.047	.023	.047	.053	.054	.832	.552
.8	1	0.799	0.798	.023	.047	.023	.046	.048	.057	.884	.657
.8	2	0.800	0.799	.017	.035	.017	.035	.05	.046	.873	.631
.8	3	0.800	0.798	.022	.045	.022	.044	.054	.048	.88	.645
.8	4	0.800	0.801	.016	.033	.016	.033	.053	.047	.863	.613

Table 17: $\widehat{\lambda}$ results when ρ=-0.3

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	-0.802	-0.805	.049	.098	.048	.096	.051	.055	.854	.595
-.8	2	-0.801	-0.800	.039	.076	.038	.076	.058	.054	.845	.577
-.8	3	-0.801	-0.806	.041	.08	.04	.081	.055	.057	.848	.583
-.8	4	-0.799	-0.803	.03	.06	.031	.061	.046	.052	.843	.574
-.3	1	-0.301	-0.304	.044	.086	.043	.087	.053	.052	.862	.609
-.3	2	-0.301	-0.300	.034	.07	.034	.068	.054	.062	.853	.592
-.3	3	-0.301	-0.304	.039	.076	.038	.076	.058	.053	.858	.603
-.3	4	-0.299	-0.302	.03	.058	.029	.059	.057	.053	.852	.59
0	1	-0.001	-0.002	.039	.08	.039	.078	.052	.056	.868	.621
0	2	0.001	0.002	.031	.062	.031	.062	.05	.055	.858	.603
0	3	-0.001	-0.003	.034	.071	.035	.07	.047	.054	.866	.618
0	4	0.000	-0.002	.027	.055	.027	.055	.05	.055	.858	.601
.3	1	0.299	0.299	.035	.072	.035	.069	.051	.057	.874	.635
.3	2	0.301	0.299	.027	.056	.027	.054	.049	.054	.865	.616
.3	3	0.300	0.298	.03	.063	.031	.063	.041	.057	.875	.636
.3	4	0.299	0.300	.025	.049	.025	.049	.048	.052	.865	.616
.8	1	0.800	0.799	.028	.057	.028	.056	.058	.048	.892	.673
.8	2	0.801	0.800	.021	.043	.021	.043	.051	.055	.88	.647
.8	3	0.800	0.799	.025	.052	.025	.051	.052	.053	.896	.682
.8	4	0.800	0.799	.02	.039	.02	.039	.049	.051	.881	.648

Table 18: $\widehat{\lambda}$ results when ρ= 0.0

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	-0.802	-0.804	.048	.095	.047	.095	.053	.06	.861	.607
-.8	2	-0.799	-0.799	.037	.075	.038	.075	.049	.058	.851	.589
-.8	3	-0.802	-0.803	.036	.072	.036	.072	.045	.05	.86	.605
-.8	4	-0.799	-0.804	.027	.054	.027	.054	.051	.053	.855	.596
-.3	1	-0.300	-0.301	.045	.09	.045	.09	.049	.056	.865	.616
-.3	2	-0.299	-0.303	.036	.073	.036	.071	.054	.056	.857	.599
-.3	3	-0.300	-0.301	.036	.074	.036	.073	.05	.057	.864	.614
-.3	4	-0.300	-0.300	.029	.058	.028	.057	.053	.061	.858	.602
0	1	0.003	-0.001	.043	.085	.042	.085	.057	.054	.87	.624
0	2	0.000	-0.000	.033	.069	.034	.067	.054	.065	.861	.607
0	3	-0.001	-0.002	.035	.07	.035	.07	.044	.051	.869	.624
0	4	-0.001	0.002	.028	.057	.028	.056	.054	.058	.862	.609
.3	1	0.300	0.302	.04	.08	.039	.077	.06	.058	.875	.636
.3	2	0.300	0.301	.03	.063	.031	.061	.049	.054	.865	.617
.3	3	0.301	0.300	.033	.066	.033	.066	.056	.052	.876	.639
.3	4	0.299	0.299	.026	.053	.026	.053	.044	.054	.866	.618
.8	1	0.800	0.800	.032	.064	.032	.064	.054	.05	.89	.668
.8	2	0.799	0.802	.025	.051	.025	.05	.041	.062	.878	.643
.8	3	0.802	0.799	.028	.057	.028	.057	.054	.046	.893	.674
.8	4	0.800	0.799	.023	.046	.023	.045	.053	.055	.878	.643

Table 19: $\widehat{\lambda}$ results when ρ= 0.3

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	-0.799	-0.800	.048	.095	.048	.095	.051	.051	.863	.611
-.8	2	-0.801	-0.799	.038	.076	.038	.076	.046	.054	.853	.593
-.8	3	-0.800	-0.799	.034	.066	.033	.067	.053	.05	.863	.611
-.8	4	-0.800	-0.798	.025	.049	.025	.049	.055	.055	.859	.603
-.3	1	-0.298	-0.296	.05	.101	.049	.098	.06	.068	.865	.615
-.3	2	-0.299	-0.299	.039	.078	.039	.078	.054	.052	.856	.598
-.3	3	-0.298	-0.298	.038	.074	.037	.074	.056	.054	.863	.612
-.3	4	-0.300	-0.295	.028	.058	.029	.057	.048	.051	.857	.601
0	1	0.001	0.002	.049	.095	.048	.096	.058	.057	.867	.62
0	2	-0.002	0.001	.039	.078	.038	.076	.059	.054	.859	.603
0	3	0.001	0.001	.037	.079	.038	.075	.046	.06	.866	.617
0	4	0.001	0.000	.03	.059	.03	.06	.054	.052	.858	.602
.3	1	0.301	0.301	.046	.093	.046	.092	.05	.057	.871	.628
.3	2	0.300	0.303	.036	.072	.036	.072	.049	.057	.862	.609
.3	3	0.300	0.300	.037	.076	.037	.075	.048	.053	.87	.625
.3	4	0.300	0.301	.03	.06	.03	.06	.051	.05	.86	.606
.8	1	0.801	0.802	.04	.081	.039	.079	.058	.056	.883	.653
.8	2	0.801	0.801	.031	.062	.031	.062	.054	.051	.871	.628
.8	3	0.800	0.799	.035	.068	.034	.068	.057	.047	.883	.654
.8	4	0.800	0.800	.029	.055	.028	.055	.056	.052	.867	.62

Table 20: $\widehat{\lambda}$ results when ρ= 0.8

λ	**W**	Median		Std. Dev.		Est. Std. Dev.		Rej. Rate		R2	
		s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1	s=.5	s=1
-.8	1	-0.799	-0.786	.052	.113	.05	.101	.06	.08	.846	.58
-.8	2	-0.798	-0.795	.04	.082	.039	.078	.054	.059	.836	.561
-.8	3	-0.799	-0.793	.033	.068	.031	.063	.056	.072	.841	.568
-.8	4	-0.800	-0.797	.023	.045	.022	.044	.053	.056	.836	.561
-.3	1	-0.293	-0.275	.063	.135	.06	.121	.065	.087	.839	.565
-.3	2	-0.297	-0.284	.049	.098	.047	.094	.058	.068	.829	.549
-.3	3	-0.299	-0.289	.043	.096	.042	.084	.059	.084	.826	.544
-.3	4	-0.299	-0.289	.031	.064	.031	.062	.053	.066	.819	.531
0	1	0.006	0.029	.07	.145	.065	.13	.064	.094	.834	.557
0	2	0.002	0.016	.051	.104	.05	.1	.054	.064	.824	.54
0	3	0.005	0.018	.05	.106	.048	.096	.063	.076	.819	.53
0	4	0.003	0.007	.037	.074	.036	.072	.064	.059	.809	.513
.3	1	0.309	0.337	.075	.145	.069	.137	.077	.082	.829	.548
.3	2	0.306	0.321	.053	.109	.053	.104	.054	.07	.818	.53
.3	3	0.307	0.327	.056	.111	.053	.106	.068	.08	.812	.518
.3	4	0.301	0.308	.04	.083	.04	.08	.051	.06	.798	.496
.8	1	0.809	0.839	.073	.147	.071	.138	.07	.094	.821	.534
.8	2	0.806	0.820	.054	.109	.053	.105	.067	.079	.806	.509
.8	3	0.811	0.824	.059	.116	.058	.114	.066	.073	.802	.503
.8	4	0.803	0.815	.045	.092	.045	.089	.058	.067	.778	.468

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] Arraiz, I., Drukker, D.M., Kelejian, H.H. and Prucha, I.R., A Spatial Cliff-Ord-type Model with Heteroskedastic Innovations: Small and Large Sample Results, *Journal of Regional Science* 50, (2010), 592-614.

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

[6] Badinger, H., and Egger, P., Estimation of higher-order spatial autoregressive cross-section models with heteroskedastic disturbances. *Papers in Regional Science* 90 (2011), 213-236.

[7] Baltagi, B.H., Egger, P. and Pfaffermayr, M., Estimating Models of Complex FDI: Are There Third-Country Effects? *Journal of Econometrics* 140 (2007), 260-281.

[8] Baltagi, B.H., Egger, P. and Pfaffermayr, M., Estimating Regional Trade Agreement Effects on FDI in an Interdependent World. *Journal of Econometrics* 145 (2008), 194-208.

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

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

[11] Baltagi, B.H., Song, S.H., Jung, B.C. and Koh, W., Testing for Serial Correlation, Spatial Autocorrelation and Random Effects Using Panel Data. *Journal of Econometrics* 140 (2007), 5-51.

[12] 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.

[13] Bao, Y., and Ullah, A., Finite Sample Moments of Maximum Likelihood Estimator in Spatial Models, *Journal of Econometrics* 137 (2007), 396-413.

[14] 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.

[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] Bollinger, C.R. and Ihlanfeldt, K.R., The Impact of Rapid Rail Transit on Economic Development: The Case of Atlanta's MARTA. *Journal of Urban Economics* 42 (1997), 179Ű

[18] Blonigen, B.A., Davies, R.B., Waddell, G.R. and Naughton, H.T., FDI in Space: Spatial Autoregressive Relationships in Foreign Direct Investment, *European Economic Review* 51 (2007), 1303-1325.

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

[20] 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.

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

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

[23] 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.

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

[25] Coughlin, C.C. and Eran Segev, Foreign Direct Investment in China: A Spatial Econometric Study. *The World Economy* 23 (2002), 1-23.

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

[27] 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.

[28] Dowd, M.R. and LeSage, J.P., Analysis of Spatial Contiguity Influences on State Price Level Formation. *International Journal of Forecasting* 13 (1997), 245-253.

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

[30] Kapoor, M., Kelejian, H.H., and Prucha, I.R., Panel Data Models with Spatially Correlated Error Components. *Journal of Econometrics* 140 (2007), 97-130.

[31] 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.

[32] 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.

[33] 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.

[34] 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.

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

[36] Kelejian, H.H. and Prucha, I.R., HAC Estimation in a Spatial Framework. *Journal of Econometrics* 140 (2007), 131-154.

[37] Kelejian, H.H. and Prucha, I.R., Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances. *Journal of Econometrics* 157 (2010), 53-67.

[38] 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.

[39] Kelejian, H.H. and Robinson, D., Infrastructure Productivity Estimation and its Underlying Econometric Specifications. *Papers in Regional Science* 76 (1997), 115-131.

[40] Kelejian, H.H. and Robinson, D., Returns to Investment in Navigation Infrastructure: An Equilibrium Approach. *Annals of Regional Science* 34 (2000), 83-108.

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

[45] Lee, L.F., and Liu, X., Efficient GMM Estimation of Higher Order Spatial Autoregressive Models with Autoregressive Distrurbances, *Econometric Theory* 26 (2010), 187-230.

[46] LeSage, J.P., Bayesian Estimation of Spatial Autoregressive Models. *International Regional Science Review* 20 (1997), 113-129.

[47] LeSage, J.P., A Spatial Econometric Analysis of China's Economic Growth. *Journal of Geographic Information Sciences* 5 (1999), 143–53.

[48] LeSage, J.P., Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. *Geographical Analysis* 32 (2000), 19-35.

[49] Lin, X. and Lee, L.F., GMM Estimation of Spatial Autoregressive Models with Unknown Heteroskedasticity. Department of Economics, *Journal of Econometrics* 157 (2010), 34-52.

[50] Liu, X. and Lee, L.-F.,Two Stage Least Squares Estimation of Spatial Autoregressive Models with Endogenous Regressors and Many Instruments, *Econometric Reviews* (2011), forthcoming.

[51] Pace, R.K. and Barry, R., Sparse Spatial Autoregressions. *Statistics and Probability Letters* 33 (1997), 291-297.

[52] 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.

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

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

[55] Rey, S.J. and Boarnet, M.G., A Taxonomy of Spatial Econometric Models for Simultaneous Equation Systems. In: Anselin, L., Florax, R.J.G.M., Rey, S.J. (Eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*. Springer, New York, NY.

[56] 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.

[57] Yang, Z., A Robust LM Test for Spatial Error Components, *Regional Science and Urban Economics* 40 (2010), 299-310.