JOURNAL OF REGIONAL SCIENCE, VOL. 50, NO. 2, 2010, pp. 592–614

A SPATIAL CLIFF-ORD-TYPE MODEL WITH HETEROSKEDASTIC INNOVATIONS: SMALL AND LARGE SAMPLE RESULTS*

Irani Arraiz
Inter-American Development Bank, 1300 New York Avenue, N.W., Washington, DC 20577. E-mail: arraiz@econ.bsos.umd.edu

David M. Drukker
Stata, 4905 Lakeway Drive, College Station, TX 77845. E-mail: ddrukker@stata.com

Harry H. Kelejian
Department of Economics, University of Maryland, College Park, MD 20742. E-mail: kelejian@econ.umd.edu

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

ABSTRACT. In this paper, we specify a linear Cliff-and-Ord-type spatial model. The model allows for spatial lags in the dependent variable, the exogenous variables, and disturbances. The innovations in the disturbance process are assumed to be heteroskedastic with an unknown form. We formulate multistep GMM/IV-type estimation procedures for the parameters of the model. We also give the limiting distributions for our suggested estimators and consistent estimators for their asymptotic variance-covariance matrices. We conduct a Monte Carlo study to show that the derived large-sample distribution provides a good approximation to the actual small-sample distribution of our estimators.

1. INTRODUCTION

Kelejian and Prucha (1999) suggested a GMM procedure for estimating the autoregressive parameter in the disturbance process in a Cliff-Ord-type spatial model. Although they demonstrated the consistency of their GMM estimator, they did not determine its large-sample distribution and so tests relating to that autoregressive parameter could not be carried out based on results of that paper. Also, Kelejian and Prucha (1998, 1999), as well as subsequent contributions, assumed that the innovations of the disturbance process were homoskedastic. This homoskedasticity assumption restricts the scope of applications of their procedure because cross-sectional spatial units often differ in size and other characteristics which causes one to suspect that the innovations to the disturbance process are heteroskedastic.[1]

*Our thanks for very helpful comments are owed to Peter Egger, Michael Pfaffermayr, and Gianfranco Piras. Also, we gratefully acknowledge financial support from the National Institutes of Health through the SBIR grants R43 AG027622 and R44 AG027622. Ingmar Prucha also thanks the CESifo in Munich for their hospitality and appreciates their support in writing this paper.

Received: May 2008; revised: December 2008; accepted: February 2009.

[1]For example, some units that have been considered are regional incomes, city size, country-industry pairs of foreign direct investment, and a crisis index relating to a country's foreign exchange

 DOI: 10.1111/j.1467-9787.2009.00618.x

Kelejian and Prucha (2007b) extended the results in their earlier papers in a variety of directions. Among other things, they considered a Cliff-Ord-type (Cliff and Ord, 1973, 1981) spatial autoregressive disturbance process with heteroskedastic innovations and suggested a modified GMM estimator, say $\widetilde{\rho}$, for the autoregressive parameter, say ρ. That GMM estimator was assumed to be based on estimated residuals that were formulated in terms of an estimator, say $\hat{\boldsymbol{\delta}}$, of the regression parameters $\boldsymbol{\delta}$. For clarity we write $\widetilde{\rho} = \widetilde{\rho}(\hat{\boldsymbol{\delta}})$ to indicate the dependence of $\widetilde{\rho}$ on $\hat{\boldsymbol{\delta}}$. Kelejian and Prucha (2007b) found that the asymptotic distribution of $\widetilde{\rho}(\hat{\boldsymbol{\delta}})$ depends on the particular choice of $\hat{\boldsymbol{\delta}}$. Because of this, Kelejian and Prucha (2007b) derive a basic theorem regarding the joint large-sample distribution of their GMM estimator $\widetilde{\rho}(\hat{\boldsymbol{\delta}})$ and $\widehat{\boldsymbol{\delta}}$ under general conditions that include various combinations of $\widetilde{\rho}(\hat{\boldsymbol{\delta}})$ and $\widehat{\boldsymbol{\delta}}$. As one example, $\widetilde{\rho}$ could be the GMM estimator of ρ based on residuals that are determined via an initial 2SLS regression parameter estimator, and $\widehat{\boldsymbol{\delta}}$ could be the generalized spatial 2SLS estimator. Results regarding the joint distribution of $\widetilde{\rho}(\hat{\boldsymbol{\delta}})$ and $\hat{\boldsymbol{\delta}}$ should be useful to researchers who wish to test joint hypotheses relating to ρ and $\boldsymbol{\delta}$.

The results in Kelejian and Prucha (2007b) are reasonably general and so they do not provide specific expressions for the large-sample distribution for the various combinations of estimators $\widetilde{\rho}$ and $\hat{\boldsymbol{\delta}}$. Given their level of generality, practitioners may find it challenging and/or tedious to specialize the distributional results in Kelejian and Prucha (2007b) for their particular estimators for ρ and $\boldsymbol{\delta}$. Practitioners may also find it challenging and/or tedious to verify that the general catalogue of assumptions in Kelejian and Prucha (2007b) is satisfied for their particular estimator combination within the context of their particular model.

The purpose of this paper is twofold. First, we specify a linear spatial Cliff-Ord model that might be considered in practice and demonstrate that our suggested estimators of its parameters satisfy the general assumptions in Kelejian and Prucha (2007b). This model allows for spatial lags in the dependent variable, the exogenous variables, and disturbances, and allows for heteroskedasticity of unknown form in the innovations. We also specialize the general distributional results in Kelejian and Prucha (2007b) for various combinations of estimators of its parameters, and provide explicit expressions for the asymptotic variance covariance matrices of the parameter estimators. These results make estimation of and inference about the parameters of this spatial model, and special cases of it, straightforward. Second, we give Monte Carlo results that describe the small-sample properties of our estimators, the estimators of their variances, as well as corresponding Wald-type tests.

Our Monte Carlo results suggest that our estimators behave quite nicely in small samples both under homoskedasticity and under heteroskedasticity of unknown form. They also indicate that the maximum-likelihood estimator of the autoregressive parameters corresponding to spatial lags of the dependent variable and disturbances can be substantially biased if the innovations are heteroskedastic. This result is consistent with the theoretical result that the maximum-likelihood estimator in Cliff-Ord type models with unknown forms of heteroskedasticity in the innovations entering the disturbance process may not be consistent. This is in contrast to the estimators developed in this paper.

market—see, respectively, Rey and Dev (2006), Vigil (1998), Baltagi, Egger, and Pfaffermayr (2007), and Kelejian, Tavlas, and Hondroyiannis (2006). Because the above-named cross-sectional units differ in so many ways, one would strongly suspect heteroskedasticity in such a model's disturbance process.

2. A SPATIAL CLIFF-ORD-TYPE MODEL

Specifications

In this section we specify a linear spatial model that allows for spatial lags in the dependent variable, the exogenous variables, and disturbances. Consistent with the terminology introduced by Anselin (1988), and used elsewhere in the literature, for example, in (Kelejian and Prucha, 2007a), we refer to this model as a spatial ARAR(1,1) model, that is, SARAR(1,1). The specification does not assume homoskedastic innovations, but instead allows for heteroskedasticity of unknown form. Apart from allowing for heteroskedasticity the assumptions are similar to those made in the existing literature. Since those assumptions have been discussed in detail before, our discussion of them will be brief.[2]

Consider the following spatial model relating to n cross-sectional units:

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

and

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

where $\mathbf{Z}_n = [\mathbf{X}_n, \mathbf{W}_n\mathbf{y}_n]$, $\boldsymbol{\delta} = [\boldsymbol{\beta}', \lambda]'$, $\mathbf{y}_n$ is the $n \times 1$ vector of observations of the dependent variable, $\mathbf{X}_n$ is the $n \times k$ matrix of observations on nonstochastic (exogenous) regressors, $\mathbf{W}_n$ and $\mathbf{M}_n$ are $n \times n$ nonstochastic weights matrices, $\mathbf{u}_n$ is the $n \times 1$ vector of regression disturbances, $\boldsymbol{\varepsilon}_n$ is an $n \times 1$ vector of innovations, λ and ρ are scalar parameters, and $\boldsymbol{\beta}$ is a $k \times 1$ vector of parameters. The subscript n denotes dependence on the sample size and so equations (1) and (2) allow for triangular arrays. Consequently, this specification allows some or all of the exogenous variables to be spatial lags of exogenous variables. Thus, the model is fairly general in that it allows for spatial spillovers in the endogenous variables, exogenous variables, and disturbances.

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

$$\mathbf{y}_{n*}(\rho) = \mathbf{Z}_{n*}(\rho)\boldsymbol{\delta} + \boldsymbol{\varepsilon}_n, \tag{3}$$

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

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

ASSUMPTION 1: *(a) All diagonal elements of* $\mathbf{W}_n$ *and* $\mathbf{M}_n$ *are zero. (b)* $\lambda \in (-1, 1)$, $\rho \in (-1, 1)$. *(c) The matrices* $\mathbf{I}_n - \bar{\lambda}\mathbf{W}_n$ *and* $\mathbf{I}_n - \bar{\rho}\mathbf{M}_n$ *are nonsingular for all* $\bar{\lambda} \in (-1, 1)$ *and* $\bar{\rho} \in (-1, 1)$.

ASSUMPTION 2: *The innovations* $\{\boldsymbol{\varepsilon}_{i,n} : 1 \leq i \leq n,\ n \geq 1\}$ *satisfy* $\mathrm{E}\boldsymbol{\varepsilon}_{i,n} = 0$, $\mathrm{E}(\boldsymbol{\varepsilon}_{i,n}^2) = \sigma_{i,n}^2$ *with* $0 < \underline{a}^\sigma \leq \sigma_{i,n}^2 \leq \bar{a}^\sigma < \infty$, *and* $\sup_{1 \leq i \leq n, n \geq 1} \mathrm{E}|\boldsymbol{\varepsilon}_{i,n}|^{4+\eta} < \infty$ *for some* $\eta > 0$. *Furthermore, for each* $n \geq 1$ *the random variables* $\boldsymbol{\varepsilon}_{1,n}, \ldots, \boldsymbol{\varepsilon}_{n,n}$ *are totally independent.*

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* $(\mathbf{I}_n - \lambda\mathbf{W}_n)^{-1}$ *and* $(\mathbf{I}_n - \rho\mathbf{M}_n)^{-1}$ *are bounded uniformly in absolute value by some finite constant.*

[2]Among other studies, see Kelejian and Prucha (1998, 2004, 2007a,b) for a more extensive discussion of these assumptions.

It is evident from (1) and (2) that, under typical specifications, $\mathbf{W}_n\mathbf{y}_n$ will be correlated with the disturbances $\mathbf{u}_n$, which motivates the use of the instrumental variable procedure. The selection of instruments as an approximation to ideal instruments is discussed by Kelejian and Prucha (1998, 2007a,b), and a review of that discussion is given below. At this point let $\mathbf{H}_n$ be an $n \times p$ matrix of nonstochastic instruments where $p \geq k + 1$, and note that in practice $\mathbf{H}_n$ would depend on $\mathbf{X}_n$. Our assumptions concerning $\mathbf{X}_n$ and $\mathbf{H}_n$ are given below.

ASSUMPTION 4: *The regressor matrices $\mathbf{X}_n$ have full column rank (for n large enough). Furthermore, the elements of the matrices $\mathbf{X}_n$ are uniformly bounded in absolute value.*

ASSUMPTION 5: *The instrument matrices $\mathbf{H}_n$ have full column rank $p \geq k + 1$ (for all n large enough). Furthermore, the elements of the matrices $\mathbf{H}_n$ are uniformly bounded in absolute value. Additionally, $\mathbf{H}_n$ is assumed to, at least, contain the linearly independent columns of $(\mathbf{X}_n, \mathbf{M}_n\mathbf{X}_n)$.*

ASSUMPTION 6: *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. Furthermore,* $\mathbf{Q}_{HZ*}(\rho) = \mathbf{Q}_{HZ} - \rho\mathbf{Q}_{HMZ}$ *has full column rank.*

(c) $\mathbf{Q}_{H\Sigma H} = \lim_{n\to\infty} n^{-1}\mathbf{H}'_n\Sigma_n\mathbf{H}_n$ *is finite and nonsingular, where* $\Sigma_n = diag_{i=1}^{n}(\sigma_{i,n}^2)$.

In treating $\mathbf{X}_n$ and $\mathbf{H}_n$ as nonstochastic our analysis should be viewed as conditional on $\mathbf{X}_n$ and $\mathbf{H}_n$.

A Brief Discussion of the Assumptions

Among other things, Assumption 1 implies that the model is complete in that the dependent vector $\mathbf{y}_n$ can be solved for in terms of $\mathbf{X}_n$ and the innovation $\boldsymbol{\varepsilon}_n$. Specifically,

$$\begin{aligned} \mathbf{y}_n &= (\mathbf{I}_n - \lambda\mathbf{W}_n)^{-1}[\mathbf{X}_n\boldsymbol{\beta} + \mathbf{u}_n] \\ \mathbf{u}_n &= (\mathbf{I}_n - \rho\mathbf{M}_n)^{-1}\boldsymbol{\varepsilon}_n. \end{aligned} \tag{4}$$

For a detailed discussion of the specification of the parameter space for the autoregressive parameters and normalizations of the spatial weights matrices, see Kelejian and Prucha (2007b).

Assumption 2 allows the innovations to be heteroskedastic with uniformly bounded variances.

Given (4), Assumption 2 implies that $E(\mathbf{y}_n) = (\mathbf{I}_n - \lambda\mathbf{W}_n)^{-1}\mathbf{X}_n\boldsymbol{\beta}$. Since under Assumptions 1 and 3 the roots of $\mathbf{W}_n$ are all less than one in absolute value,

$$E(\mathbf{y}_n) = \left[\mathbf{I}_n + \lambda\mathbf{W}_n + \lambda^2\mathbf{W}_n^2 + \cdots\right]\mathbf{X}_n\boldsymbol{\beta}. \tag{5}$$

We suggest a multistep estimation procedure below. In the first step instruments are needed for $\mathbf{Z}_n$, and in a later step instruments are needed for $\mathbf{M}_n\mathbf{Z}_n$. The ideal instruments are

$$\begin{aligned} E(\mathbf{Z}_n) &= [\mathbf{X}_n, \mathbf{W}_nE(\mathbf{y}_n)], \\ E(\mathbf{M}_n\mathbf{Z}_n) &= [\mathbf{M}_n\mathbf{X}_n, \mathbf{M}_n\mathbf{W}_nE(\mathbf{y}_n)]. \end{aligned} \tag{6}$$

In light of (5), all of the columns of $E(\mathbf{Z}_n)$ and $E(\mathbf{M}_n\mathbf{Z}_n)$ are linear in

$$\mathbf{X}_n, \mathbf{W}_n\mathbf{X}_n, \mathbf{W}_n^2\mathbf{X}_n, \ldots, \mathbf{M}_n\mathbf{X}_n, \mathbf{M}_n\mathbf{W}_n\mathbf{X}_n, \mathbf{M}_n\mathbf{W}_n^2\mathbf{X}_n, \ldots \tag{7}$$

Let $\mathbf{H}_n$ be a subset of the columns in (7), say

$$\mathbf{H}_n = \left(\mathbf{X}_n, \mathbf{W}_n\mathbf{X}_n, \ldots, \mathbf{W}_n^q\mathbf{X}_n, \mathbf{M}_n\mathbf{X}_n, \mathbf{M}_n\mathbf{W}_n\mathbf{X}_n, \ldots, \mathbf{M}_n\mathbf{W}_n^q\mathbf{X}_n\right), \tag{8}$$

where, typically, $q \leq 2$. Then the evident approximation to the ideal instruments for $\mathbf{Z}_n$ and $\mathbf{M}_n\mathbf{Z}_n$ is $\mathbf{P}_n\mathbf{Z}_n$ and $\mathbf{P}_n\mathbf{M}_n\mathbf{Z}_n$, where $\mathbf{P}_n$ is the projection matrix: $\mathbf{P}_n = \mathbf{H}_n(\mathbf{H}'_n\mathbf{H}_n)^{-1}\mathbf{H}_n$. In passing note that, via Assumption 5, $\mathbf{H}_n$ is assumed to contain at least the linearly independent columns of $\mathbf{X}_n$ and $\mathbf{M}_n\mathbf{X}_n$, and therefore

$$\begin{aligned} \mathbf{P}_n\mathbf{Z}_n &= (\mathbf{X}_n, \mathbf{P}_n\mathbf{W}_n\mathbf{y}_n), \\ \mathbf{P}_n\mathbf{M}_n\mathbf{Z}_n &= (\mathbf{M}_n\mathbf{X}_n, \mathbf{P}_n\mathbf{M}_n\mathbf{W}_n\mathbf{y}_n). \end{aligned} \tag{9}$$

Assumption 3 is a technical assumption, which is used in the large-sample derivation of the regression parameter estimator. Among other things, this assumption limits the extent of spatial autocorrelation.

Assumption 4 rules out multicollinearity problems, as well as unbounded exogenous variables. Among other things, Assumption 5 implies that there are at least as many instruments as there are regression parameters. Assumption 6 rules out redundant instruments and specifies conditions, which ensure the identifiability of the regression parameter estimators.

3. ESTIMATORS

In this section we specify GMM and instrumental variable (IV) estimators for the model parameters ρ and $\boldsymbol{\delta}$. The suggested estimation procedure consists of two steps. Each step consists of substeps involving the estimation of ρ and $\boldsymbol{\delta}$ by GMM and IV methods. 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 replaced by an estimator.

Moment Conditions

Following Kelejian and Prucha (2007b) our estimators for ρ will be GMM estimators corresponding to the following population moment conditions:

$$\begin{aligned} n^{-1}\mathrm{E}\bar{\boldsymbol{\varepsilon}}'_n\bar{\boldsymbol{\varepsilon}}_n &= n^{-1}tr\left\{\mathbf{M}_n\left[diag_{i=1}^n(\mathrm{E}\boldsymbol{\varepsilon}_{i,n}^2)\right]\mathbf{M}'_n\right\}, \\ n^{-1}\mathrm{E}\bar{\boldsymbol{\varepsilon}}'_n\boldsymbol{\varepsilon}_n &= 0, \end{aligned} \tag{10}$$

with $\bar{\boldsymbol{\varepsilon}}_n = \mathbf{M}_n\boldsymbol{\varepsilon}_n$. Let $\mathbf{A}_{1,n} = \mathbf{M}'_n\mathbf{M}_n - diag_{i=1}^n(\mathbf{m}'_{.i,n}\ \mathbf{m}_{.i,n})$ and $\mathbf{A}_{2,n} = \mathbf{M}_n$. It is readily seen that these moment conditions can also be written as

$$\begin{aligned} n^{-1}\mathrm{E}\boldsymbol{\varepsilon}'_n\mathbf{A}_{1,n}\boldsymbol{\varepsilon}_n &= n^{-1}\mathrm{E}[\mathbf{u}_n - \rho\bar{\mathbf{u}}_n]'\mathbf{A}_{1,n}[\mathbf{u}_n - \rho\bar{\mathbf{u}}_n] = 0, \\ n^{-1}\mathrm{E}\boldsymbol{\varepsilon}'_n\mathbf{A}_{2,n}\boldsymbol{\varepsilon}_n &= n^{-1}\mathrm{E}[\mathbf{u}_n - \rho\bar{\mathbf{u}}_n]'\mathbf{A}_{2,n}[\mathbf{u}_n - \rho\bar{\mathbf{u}}_n] = 0, \end{aligned} \tag{11}$$

with $\bar{\mathbf{u}}_n = \mathbf{M}_n\mathbf{u}_n$.

The first condition in equation (10) allows the innovations to be heteroskedastic of unknown form. If the innovations are homoskedastic with finite variance σ^2, this condition simplifies to

$$n^{-1}\mathrm{E}\bar{\boldsymbol{\varepsilon}}'_n\bar{\boldsymbol{\varepsilon}}_n = \sigma^2 n^{-1}tr\left\{\mathbf{M}_n\mathbf{M}'_n\right\}.$$

Under the null hypothesis of homoskedasticity, sample versions of the two conditions will converge to the same quantity.

GMM/IV Estimators, Original Model

Step 1a: 2SLS estimator. In the first step, $\boldsymbol{\delta}$ is estimated by 2SLS applied to model (1) using the instrument matrix $\mathbf{H}_n$ in Assumption 5. Let $\widetilde{\boldsymbol{\delta}}_n$ denote the 2SLS estimator, then

$$\widetilde{\boldsymbol{\delta}}_n = (\widetilde{\mathbf{Z}}_n'\mathbf{Z}_n)^{-1}\widetilde{\mathbf{Z}}_n'\mathbf{y}_n, \tag{12}$$

where $\widetilde{\mathbf{Z}}_n = \mathbf{P}_H\mathbf{Z}_n = (\mathbf{X}_n, \widetilde{\mathbf{W}_n\mathbf{y}_n})$, $\widetilde{\mathbf{W}_n\mathbf{y}_n} = \mathbf{P}_H\mathbf{W}_n\mathbf{y}_n$, and where $\mathbf{P}_H = \mathbf{H}_n\ (\mathbf{H}_n' \times \mathbf{H}_n)^{-1}\ \mathbf{H}_n'$. An instrument matrix such as $\mathbf{H}_n$ was suggested originally in Kelejian and Prucha (1998).

Step 1b: Initial GMM estimator of ρ based on 2SLS residuals. In light of (1) and (12), the 2SLS residuals are $\widetilde{\mathbf{u}}_n = \mathbf{y}_n - \mathbf{Z}_n\widetilde{\boldsymbol{\delta}}_n$. Let $\widetilde{\bar{\mathbf{u}}}_n = \mathbf{M}_n\widetilde{\mathbf{u}}_n$ and $\widetilde{\bar{\bar{\mathbf{u}}}}_n = \mathbf{M}_n^2\widetilde{\mathbf{u}}_n$. Consider the following sample moments corresponding to (11) based on estimated residuals:

$$\begin{aligned} m(\rho, \widetilde{\boldsymbol{\delta}}_n) &= n^{-1}\begin{bmatrix} (\widetilde{\mathbf{u}}_n - \rho\widetilde{\bar{\mathbf{u}}}_n)'\mathbf{A}_1(\widetilde{\mathbf{u}}_n - \rho\widetilde{\bar{\mathbf{u}}}_n) \\ (\widetilde{\mathbf{u}}_n - \rho\widetilde{\bar{\mathbf{u}}}_n)'\mathbf{A}_2(\widetilde{\mathbf{u}}_n - \rho\widetilde{\bar{\mathbf{u}}}_n) \end{bmatrix} \\ &= \mathbf{g}_n(\widetilde{\boldsymbol{\delta}}_n) - \mathbf{G}_n(\widetilde{\boldsymbol{\delta}}_n)\begin{bmatrix} \rho \\ \rho^2 \end{bmatrix}, \end{aligned} \tag{13}$$

where the elements of the 2×1 vector $\mathbf{g}_n$ and the 2×2 matrix $\mathbf{G}_n$ are defined in Appendix B: Definition of $\mathbf{G}$ and $\mathbf{g}$. Equation (13) implies that the elements of $\mathbf{g}_n(\widetilde{\boldsymbol{\delta}}_n)$ and $\mathbf{G}_n(\widetilde{\boldsymbol{\delta}}_n)$ are observable functions of $\widetilde{\mathbf{u}}_n$, $\widetilde{\bar{\mathbf{u}}}_n$, and $\widetilde{\bar{\bar{\mathbf{u}}}}_n$. Our initial GMM estimator for ρ is defined as

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

where $a^\rho \geq 1$. In light of the second expression in (13) the estimator can be viewed as an unweighted nonlinear least squares estimator. Given further assumptions listed below, it is consistent, but not efficient because of this lack of weighting.

Step 1c: Efficient GMM estimator of ρ based on 2SLS residuals. As might be anticipated from the discussion above, our efficient GMM estimator of ρ is a weighted nonlinear least squares estimator. Specifically, this estimator is $\widetilde{\rho}_n$ where

$$\widetilde{\rho}_n = \underset{[\rho\in[-a^\rho, a^\rho]]}{argmin}\left[m(\rho, \widetilde{\boldsymbol{\delta}}_n)'\widetilde{\boldsymbol{\Psi}}_n^{-1}m(\rho, \widetilde{\boldsymbol{\delta}}_n)\right] \tag{15}$$

and where the weighting matrix is $\widetilde{\boldsymbol{\Psi}}_n^{-1}$. The matrix $\widetilde{\boldsymbol{\Psi}}_n = \widetilde{\boldsymbol{\Psi}}_n(\check{\rho}_n)$, defined in Appendix B: Definition of $\widetilde{\boldsymbol{\Psi}}$ and $\widetilde{\boldsymbol{\Omega}}$, is an estimator of the variance-covariance matrix of the limiting distribution of the normalized sample moments $n^{1/2}m(\rho, \widetilde{\boldsymbol{\delta}}_n)$.

GMM/IV Estimators, Transformed Model

Step 2a: GS2SLS estimator. Consider the spatial Cochrane-Orcutt transformed model in (3). Analogous to Kelejian and Prucha (1998) we now define a generalized spatial two-stage least squares (GS2SLS) estimator of $\boldsymbol{\delta}$ as the 2SLS estimator of the transformed model in (3) after replacing the parameter ρ by $\widetilde{\rho}_n$ computed in Step 1c. Specifically, the GS2SLS estimator is defined as

$$\widehat{\boldsymbol{\delta}}_n(\widetilde{\rho}_n) = [\widehat{\mathbf{Z}}_{n*}(\widetilde{\rho}_n)'\mathbf{Z}_{n*}(\widetilde{\rho}_n)]^{-1}\widehat{\mathbf{Z}}_{n*}(\widetilde{\rho}_n)'\mathbf{y}_{n*}(\widetilde{\rho}_n), \tag{16}$$

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

Step 2b: Efficient GMM estimator of ρ using GS2SLS residuals. The GS2SLS residuals are given by $\widehat{\mathbf{u}}_n = \mathbf{y}_n - \mathbf{Z}_n\widehat{\boldsymbol{\delta}}_n(\widetilde{\rho}_n)$. Let $\widehat{\bar{\mathbf{u}}}_n = \mathbf{M}_n\widehat{\mathbf{u}}_n$ and $\widehat{\bar{\bar{\mathbf{u}}}}_n = \mathbf{M}_n^2\widehat{\mathbf{u}}_n$. Now consider the sample moments $m(\rho, \widehat{\boldsymbol{\delta}}_n)$ obtained by replacing the 2SLS residuals in (13) by the GS2SLS residuals $\widehat{\mathbf{u}}_n$, $\widehat{\widehat{\mathbf{u}}}_n$, and $\widehat{\widehat{\widehat{\mathbf{u}}}}_n$. The efficient GMM estimator for ρ based on GS2SLS residuals is now given by

$$\widehat{\rho}_n = \underset{[\rho \in [-a^\rho, a^\rho]]}{argmin} \left[m(\rho, \widehat{\boldsymbol{\delta}}_n)' \widehat{\boldsymbol{\Psi}}_n^{-1} m(\rho, \widehat{\boldsymbol{\delta}}_n) \right], \tag{17}$$

where the weighting matrix is $\widehat{\boldsymbol{\Psi}}_n^{-1}$. The matrix $\widehat{\boldsymbol{\Psi}}_n = \widehat{\boldsymbol{\Psi}}_n(\widetilde{\rho}_n)$, defined in Appendix B: Definition of $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Omega}}$, is an estimator of the variance-covariance matrix of the limiting distribution of the normalized sample moments $n^{1/2} m(\rho, \widehat{\boldsymbol{\delta}}_n)$.[3]

4. LARGE SAMPLE DISTRIBUTION

In this section we give results on the joint limiting distribution of the initial 2SLS estimator, $\widetilde{\boldsymbol{\delta}}_n$, and the efficient GMM estimator of ρ based on 2SLS residuals, namely $\widetilde{\rho}_n$. These estimators relate to the untransformed model. We also give the joint limiting distribution of the GS2SLS estimator $\widehat{\boldsymbol{\delta}}_n$, and the efficient GMM estimator of ρ that is based on GS2SLS residuals, namely $\widehat{\rho}_n$. These estimators correspond to the transformed model. Proofs are given in Appendix C.

GMM/IV Estimators, Original Model

In Appendix C we prove the following theorem concerning the joint limiting distribution of $\widetilde{\rho}_n$ and $\widetilde{\boldsymbol{\delta}}_n$.

THEOREM 1: *Suppose Assumptions 1–6 above and Assumptions A1 and A2 in Appendix A hold. Then, $\widetilde{\rho}_n$ is efficient among the class of GMM estimators based on 2SLS residuals, and*

$$\begin{bmatrix} n^{1/2}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}) \\ n^{1/2}(\widetilde{\rho}_n - \rho) \end{bmatrix} \xrightarrow{\mathbf{D}} N\left[0, \underset{n\to\infty}{\text{plim}}\, \widetilde{\boldsymbol{\Omega}}_n(\widetilde{\rho}_n)\right], \tag{18}$$

where $\text{plim}_{n\to\infty} \widetilde{\boldsymbol{\Omega}}_n(\widetilde{\rho}_n)$ *is a positive definite matrix. For applied purposes, an expression is needed for $\widetilde{\boldsymbol{\Omega}}_n(\widetilde{\rho}_n)$. This expression is given in Appendix B: Definition of $\widetilde{\boldsymbol{\Psi}}$ and $\widetilde{\boldsymbol{\Omega}}$.*

The result in (18) indicates that both $\widetilde{\boldsymbol{\delta}}_n$ and $\widetilde{\rho}_n$ are consistent. It also suggests that small-sample inferences concerning either ρ, $\boldsymbol{\delta}$, or both can be based on the small-sample approximation

$$\begin{bmatrix} \widetilde{\boldsymbol{\delta}}_n \\ \widetilde{\rho}_n \end{bmatrix} \dot{\sim} N\left(\begin{bmatrix} \boldsymbol{\delta} \\ \rho \end{bmatrix}, n^{-1}\widetilde{\boldsymbol{\Omega}}_n \right).$$

GMM/IV Estimators, Transformed Model

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

THEOREM 2: *Suppose Assumptions 1–6 above and Assumptions A1 and A3 in the Appendix A hold. Then, $\widehat{\rho}_n$ is efficient among the class of GMM estimators based on GS2SLS residuals, and*

$$\begin{bmatrix} n^{1/2}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}) \\ n^{1/2}(\widehat{\rho}_n - \rho) \end{bmatrix} \xrightarrow{\mathbf{D}} N\left[0, \underset{n\to\infty}{\text{plim}}\, \widehat{\boldsymbol{\Omega}}_n(\widehat{\rho}_n)\right],$$

[3] $n^{1/2} m(\rho, \widehat{\boldsymbol{\delta}}_n)$ and $n^{1/2} m(\rho, \widetilde{\boldsymbol{\delta}}_n)$ have different limiting distributions.

where $\operatorname{plim}_{n\to\infty}\widehat{\boldsymbol{\Omega}}_n(\widehat{\rho}_n)$ *is a positive definite matrix. For applied purposes, an expression is needed for* $\widehat{\boldsymbol{\Omega}}_n(\widehat{\rho}_n)$. *This expression is given in Appendix B: Definition of* $\widehat{\boldsymbol{\Psi}}$ *and* $\widehat{\boldsymbol{\Omega}}$.

Clearly, Theorem 2 implies that both $\widehat{\boldsymbol{\delta}}_n$ and $\widehat{\rho}_n$ are consistent. It also suggests that small-sample inferences can be based on the approximation

$$\begin{bmatrix}\widehat{\boldsymbol{\delta}}_n\\ \widehat{\rho}_n\end{bmatrix} \dot{\sim} N\left(\begin{bmatrix}\boldsymbol{\delta}\\ \rho\end{bmatrix}, n^{-1}\widehat{\boldsymbol{\Omega}}_n\right).$$

5. MONTE CARLO EXPERIMENTS

In this section we give Monte Carlo results which suggest that our estimators and corresponding test statistics behave well in finite samples. Our Monte Carlo model is a special case of the one specified in (1) and (2). Our experimental design is somewhat similar to those used in the literature by Kelejian and Prucha (1999, 2007a) and by Anselin and Florax (1995).

The Model

The model underlying our Monte Carlo experiments is a special case of the model specified in (1) and (2) with two exogenous regressors, that is, $\mathbf{X}_n = [x_{n,1}, x_{n,2}]$ and $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2)'$, and with $\mathbf{M}_n = \mathbf{W}_n$.

We consider two cases for the innovation vector $\boldsymbol{\varepsilon}_n$. In one of these cases the elements of the innovation vector are i.i.d. $N(0, c^2)$, and so their standard deviation is c. In our second case the elements of the innovation vector are heteroskedastic. In this case we take the i-th element of the innovation vector $\boldsymbol{\varepsilon}_n$ as

$$\text{(19)}\qquad \begin{aligned}\boldsymbol{\varepsilon}_{n,i} &= \sigma_{n,i}\zeta_{n,i},\\ \sigma_{n,i} &= c\frac{d_{n,i}}{\Sigma_{j=1}^{n} d_{n,j}/n},\end{aligned}$$

where $\zeta_{n,i}$ is, for each of our considered sample sizes, i.i.d. $N(0, 1)$, and $d_{n,i}$ is the number of neighbors the i-th unit has, which will be defined by the sample size, and weights matrices described below. At this point note that the average of the standard deviations of the elements of $\boldsymbol{\varepsilon}_n$ is c, and thus the average standard deviation is identical to that in the homoskedastic case. Also note that these standard deviations are related to the number of neighbors each unit has. One example in which units might have different numbers of neighbors is the case in which the units differ in size. If neighbors are defined as units falling within a certain distance, then each unit in a group of smaller units could have many neighbors, while each unit in a group of larger units could have fewer neighbors. This scenario could relate to the northeastern portion of the United States, as compared to western states in the United States.

The parameters of the model that we will estimate are $\boldsymbol{\delta} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \lambda)'$ and ρ. The specifications we use to generate 2,000 repetitions for each Monte Carlo experiment are described below.

The two $n \times 1$ regressors $x_{n,1}$ and $x_{n,2}$ are normalized versions of income per capita and the proportion of housing units that are rental in 1980, in 760 counties in U.S. mid-western states. These data were taken from Kelejian and Robinson (1995). We normalized the 760 observations on these variables by subtracting from each observation the corresponding sample average, and then dividing that result by the sample standard deviation. The first n values of these normalized variables were used in our Monte Carlo experiments of sample size n. For sample sizes larger than 760 the observations were repeated. Finally, the same set of observations on these variables were used in all Monte Carlo repetitions.

We considered five experimental values for λ and for ρ, namely -0.8, -0.3, $0, 0.3., 0.8$. In all of our experiments we took $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = 1$. We consider two values for the (average) standard deviation c, namely 0.5 and 1. However, due to space limitations, we only report results below for the case $c = 1$. Results for the case $c = 0.5$ are consistent with those for the case $c = 1$, and are available on our website in a longer version of this paper.

For each approximate sample size,[4] we consider three weights matrices, but again report results for only two of them that are described below. The description of the third matrix and corresponding results are available on our website. The weights matrices we report results for correspond to a "space" in which units located in the northeast portion of that space are smaller, closer to each other, and have more neighbors than the units corresponding to other quadrants of that space. Again, one might think of the states located in the northeastern portion of the United States, as compared to western states.

To define these matrices we report results for a matrix in terms of a square grid with both the x and y coordinates only taking on the values $1,\ 1.5,\ 2,\ 2.5, \ldots, \bar{m}$. Let the units in the northeast quadrant of this matrix be 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: $x = 1, 2, \ldots, m-1$ and $y = 1, 2, \ldots, m-1$. In this setup it should be clear that the number of units located in the northeast quadrant is inversely related to m.

For this matrix we define a distance measure between any two units, i_1 and i_2, that have coordinates (x_1, y_1) and (x_2, y_2), respectively, as the Euclidean distance between them, namely

$$d(i_1, i_2) = \left[(x_1 - x_2)^2 + (y_1 - y_2)^2\right]^{1/2}.$$

Given this distance measure we define the (i, j)-th element of our row normalized weights matrix $\mathbf{W}$ as

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

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

For our experiments with a sample size of approximately 500, we considered two cases of this matrix, namely $(m = 5, \bar{m} = 15)$ and $(m = 14, \bar{m} = 20)$. These values of m and $\bar{m}$ imply sample sizes of $n = 486$ and $n = 485$, respectively. These values of m and $\bar{m}$ were selected because they correspond to different proportions of units in the northeast quadrant, where each unit has more neighbors than units located in the other quadrants. As indicated, the number of neighbors each unit has is important because it is a determinant of the standard deviation of the innovation—see equation (19). In our first small-sample case, namely $(m = 5, \bar{m} = 15)$, approximately 75 percent of the units are located in the northeast quadrant; in our second case, $(m = 14, \bar{m} = 20)$, approximately 25 percent of the units are located in the northeast quadrant.

For our experiments with a sample size of approximately 1,000, the two variations of this matrix we considered are $(m = 7, \bar{m} = 21)$ and $(m = 20, \bar{m} = 28)$. The implied sample sizes are $n = 974$ and $n = 945$, respectively. In these two cases, the proportion of units located in the northeast quadrant are approximately 75 and 24 percent, respectively.

[4]Our discussion below will clarify this notion of "approximate" sample size.

TABLE 1: North-East Modified Rook Matrices Changed % Changes to Opposite

Matrix R1	Matrix R2	Matrix R3	Matrix R4
$(m = 5, \bar{m} = 15)$	$(m = 7, \bar{m} = 21)$	$(m = 14, \bar{m} = 20)$	$(m = 20, \bar{m} = 28)$
$n = 486$	$n = 974$	$n = 485$	$n = 945$
$\%NE : 75\%$	$\%NE : 75\%$	$\%NE : 25\%$	$\%NE : 24\%$

5.0	*		*		*	*	*	*	*
4.5					*	*	*	*	*
4.0	*		*		*	*	*	*	*
3.5					*	*	*	*	*
3.0	*		*		*	*	*	*	*
2.5									
2.0	*		*		*		*		*
1.5									
1.0	*		*		*		*		*
	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0

FIGURE 1: Example of a North-East Modified Rook Matrix: $m = 2$ and $\bar{m} = 5$.

Below we refer to all of these matrices as north-east modified-rook matrices. For future reference we summarize the characteristics of these four "modified rook" matrices in Table 1. We also illustrate a north-east modified rook matrix, with the units indicated by the stars, in Figure 1 for the case in which $m = 2$ and $\bar{m} = 5$.

Monte Carlo Results

Our Monte Carlo results are given in Tables 2–6. These tables contain results for the generalized spatial 2SLS estimator $\widehat{\boldsymbol{\delta}}(\hat{\rho})$ defined in equation (16), which is based on the instrument matrix $\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}]$, where $\tilde{\rho}$ is replaced by $\hat{\rho}$ that is the efficient GM estimator given in equation (17). Only results for the estimators of λ and ρ, which are denoted in the tables as λ_{GS} and ρ_{GS}, are reported. For purposes of comparison, we also report the quasi-maximum likelihood estimators of these two parameters, denoted in the tables as λ_{ML} and ρ_{ML}.

The results in Tables 2–5 correspond to the case of heteroskedastic innovations, while the results in Table 6 correspond to homoskedastic innovations. The results in all five tables are based on north-east modified rook matrices. For cases involving heteroskedasticity, efficiency issues involving sample size comparisons can be based on comparisons of Tables 2 and 3, and on comparisons of Tables 4 and 5. The results in Table 6 correspond to those in Table 2, the difference being that the former reports on the homoskedasticity case while the latter reports on the heteroskedasticity case.

The results in Tables 2–5 are consistent with our large-sample theory, namely that λ_{GS} and ρ_{GS} are consistent estimators and, in the presence of heteroskedastic innovations, the quasi-maximum-likelihood estimators λ_{ML} and ρ_{ML} are in general not consistent. For example, notice that in all of the tables the biases of λ_{GS} and ρ_{GS} are so small that the root mean square error is approximately equal to the standard deviation. Also note that in all of the tables the rejection rates corresponding to λ_{GS} and to ρ_{GS} are quite close to the theoretical 0.05 level. Indeed, in each of the Tables 2–5 the average of these rejection rates over all of the experiments considered relating to both λ_{GS} and ρ_{GS} are quite close to the theoretical 0.05 level. For future reference we note that these averages do not mask outliers; the largest of these outliers, namely 0.1190, relates λ_{GS} in Table 4 and corresponds to the experiment $\rho = 0.8$ and $\lambda = 0.3$. In the tables there are no rejection rate outliers that relate to the estimator ρ_{GS}.

TABLE 2: Heteroskedasticity with $c = 1$, Modified Rook Matrix R1 ($n = 486$)

		ρ_{GS}				ρ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.7870	0.1242	0.0425	0.1248	−0.6464	0.0634	0.3000	0.1661
−0.8	−0.3	−0.7899	0.1167	0.0420	0.1172	−0.5918	0.0659	0.7100	0.2184
−0.8	0	−0.7902	0.1117	0.0420	0.1121	−0.5847	0.0676	0.7625	0.2256
−0.8	0.3	−0.7923	0.1081	0.0430	0.1083	−0.5863	0.0675	0.7715	0.2241
−0.8	0.8	−0.7922	0.1025	0.0470	0.1028	−0.6036	0.0658	0.7200	0.2071
−0.3	−0.8	−0.2974	0.1285	0.0480	0.1285	−0.3047	0.0949	0.0190	0.0950
−0.3	−0.3	−0.2964	0.1344	0.0485	0.1345	−0.2473	0.0907	0.0410	0.1049
−0.3	0	−0.2960	0.1317	0.0500	0.1317	−0.2281	0.0913	0.0675	0.1162
−0.3	0.3	−0.2949	0.1279	0.0490	0.1280	−0.2185	0.0936	0.0915	0.1241
−0.3	0.8	−0.2944	0.1198	0.0535	0.1199	−0.2269	0.0916	0.0910	0.1172
0	−0.8	0.0013	0.1190	0.0510	0.1190	−0.0685	0.1023	0.0810	0.1232
0	−0.3	−0.0007	0.1262	0.0505	0.1262	−0.0276	0.0961	0.0335	0.1000
0	0	−0.0011	0.1267	0.0480	0.1267	−0.0072	0.0971	0.0265	0.0973
0	0.3	−0.0011	0.1254	0.0465	0.1254	0.0073	0.0977	0.0310	0.0980
0	0.8	0.0013	0.1177	0.0510	0.1177	0.0074	0.0986	0.0380	0.0989
0.3	−0.8	0.3019	0.1012	0.0510	0.1012	0.2013	0.1012	0.1805	0.1414
0.3	−0.3	0.2967	0.1092	0.0560	0.1092	0.2182	0.0974	0.1190	0.1272
0.3	0	0.2958	0.1116	0.0500	0.1117	0.2356	0.0972	0.0855	0.1166
0.3	0.3	0.2947	0.1122	0.0500	0.1123	0.2529	0.0978	0.0665	0.1085
0.3	0.8	0.2971	0.1076	0.0475	0.1077	0.2642	0.0980	0.0670	0.1044
0.8	−0.8	0.7992	0.0511	0.0535	0.0512	0.7364	0.0608	0.2960	0.0880
0.8	−0.3	0.7937	0.0572	0.0560	0.0575	0.7240	0.0659	0.2785	0.1005
0.8	0	0.7885	0.0615	0.0695	0.0625	0.7254	0.0691	0.2210	0.1017
0.8	0.3	0.7822	0.0685	0.0695	0.0708	0.7330	0.0733	0.1700	0.0993
0.8	0.8	0.7814	0.0692	0.0575	0.0717	0.7525	0.0819	0.1500	0.0947
Average		0.0000	0.1068	0.0509	0.1072	0.0207	0.0851	0.2167	0.1279
		λ_{GS}				λ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.8093	0.0627	0.0525	0.0634	−0.7650	0.0384	0.0145	0.0520
−0.8	−0.3	−0.3050	0.0612	0.0495	0.0614	−0.3409	0.0494	0.0525	0.0642
−0.8	0	−0.0031	0.0505	0.0515	0.0506	−0.0510	0.0454	0.1070	0.0683
−0.8	0.3	0.2982	0.0373	0.0505	0.0373	0.2570	0.0357	0.1265	0.0559
−0.8	0.8	0.7996	0.0120	0.0495	0.0120	0.7899	0.0112	0.0710	0.0151
−0.3	−0.8	−0.8014	0.0433	0.0380	0.0434	−0.7307	0.0329	0.1220	0.0767
−0.3	−0.3	−0.3028	0.0574	0.0485	0.0575	−0.2957	0.0476	0.0075	0.0478
−0.3	0	−0.0019	0.0520	0.0490	0.0520	−0.0208	0.0455	0.0230	0.0500
−0.3	0.3	0.2994	0.0410	0.0480	0.0410	0.2722	0.0376	0.0450	0.0468
−0.3	0.8	0.7997	0.0140	0.0480	0.0140	0.7905	0.0130	0.0480	0.0161

Continued

TABLE 2: Continued

		λ_{GS}				λ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
0	−0.8	−0.8005	0.0367	0.0380	0.0367	−0.7279	0.0300	0.1825	0.0781
0	−0.3	−0.2996	0.0552	0.0455	0.0552	−0.2734	0.0473	0.0190	0.0542
0	0	0.0004	0.0543	0.0475	0.0543	−0.0020	0.0481	0.0110	0.0482
0	0.3	0.3002	0.0455	0.0460	0.0455	0.2828	0.0419	0.0260	0.0453
0	0.8	0.8001	0.0167	0.0505	0.0167	0.7910	0.0154	0.0405	0.0178
0.3	−0.8	−0.7991	0.0327	0.0440	0.0327	−0.7307	0.0286	0.2295	0.0750
0.3	−0.3	−0.2972	0.0578	0.0455	0.0578	−0.2523	0.0497	0.0550	0.0689
0.3	0	0.0028	0.0587	0.0465	0.0588	0.0187	0.0530	0.0240	0.0562
0.3	0.3	0.3027	0.0530	0.0510	0.0531	0.2959	0.0487	0.0175	0.0489
0.3	0.8	0.8013	0.0216	0.0485	0.0216	0.7913	0.0204	0.0390	0.0222
0.8	−0.8	−0.7980	0.0288	0.0520	0.0288	−0.7421	0.0243	0.2160	0.0627
0.8	−0.3	−0.2877	0.0659	0.0735	0.0671	−0.2189	0.0570	0.1175	0.0992
0.8	0	0.0186	0.0791	0.1000	0.0813	0.0657	0.0678	0.0820	0.0944
0.8	0.3	0.3260	0.0858	0.1120	0.0897	0.3394	0.0740	0.0670	0.0838
0.8	0.8	0.8158	0.0474	0.0975	0.0500	0.7962	0.0535	0.0920	0.0536
Average		0.0024	0.0468	0.0553	0.0473	0.0136	0.0407	0.0734	0.0561

TABLE 3: Heteroskedasticity with $c = 1$, Modified Rook Matrix R2 ($n = 974$)

		ρ_{GS}				ρ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.7963	0.0945	0.0495	0.0946	−0.6485	0.0471	0.6840	0.1587
−0.8	−0.3	−0.7977	0.0853	0.0400	0.0853	−0.5884	0.0494	0.9760	0.2173
−0.8	0	−0.7979	0.0819	0.0390	0.0820	−0.5796	0.0501	0.9855	0.2260
−0.8	0.3	−0.7980	0.0800	0.0415	0.0800	−0.5806	0.0504	0.9860	0.2251
−0.8	0.8	−0.7985	0.0762	0.0445	0.0763	−0.6005	0.0492	0.9770	0.2055
−0.3	−0.8	−0.3003	0.0983	0.0530	0.0983	−0.3132	0.0678	0.0245	0.0691
−0.3	−0.3	−0.3006	0.0991	0.0525	0.0991	−0.2463	0.0663	0.0600	0.0854
−0.3	0	−0.3001	0.0968	0.0480	0.0968	−0.2238	0.0671	0.1105	0.1015
−0.3	0.3	−0.3003	0.0935	0.0480	0.0935	−0.2126	0.0679	0.1575	0.1107
−0.3	0.8	−0.3003	0.0872	0.0470	0.0872	−0.2215	0.0670	0.1530	0.1032
0	−0.8	−0.0006	0.0914	0.0515	0.0914	−0.0790	0.0735	0.1630	0.1079
0	−0.3	−0.0016	0.0952	0.0540	0.0952	−0.0282	0.0705	0.0425	0.0760
0	0	−0.0021	0.0938	0.0510	0.0938	−0.0050	0.0711	0.0315	0.0713
0	0.3	−0.0018	0.0914	0.0510	0.0914	0.0128	0.0714	0.0360	0.0726
0	0.8	−0.0009	0.0838	0.0530	0.0838	0.0151	0.0697	0.0495	0.0714
0.3	−0.8	0.2981	0.0746	0.0540	0.0747	0.1914	0.0711	0.3455	0.1298
0.3	−0.3	0.2959	0.0814	0.0515	0.0815	0.2163	0.0702	0.2030	0.1093
0.3	0	0.2951	0.0856	0.0530	0.0858	0.2371	0.0709	0.1275	0.0948
0.3	0.3	0.2938	0.0858	0.0530	0.0860	0.2563	0.0715	0.0765	0.0838
0.3	0.8	0.2967	0.0777	0.0505	0.0778	0.2732	0.0704	0.0715	0.0753
0.8	−0.8	0.7994	0.0356	0.0540	0.0356	0.7321	0.0428	0.5045	0.0803
0.8	−0.3	0.7970	0.0393	0.0600	0.0394	0.7240	0.0471	0.4875	0.0894
0.8	0	0.7938	0.0435	0.0645	0.0439	0.7274	0.0507	0.3800	0.0886
0.8	0.3	0.7896	0.0496	0.0660	0.0507	0.7346	0.0547	0.2775	0.0853
0.8	0.8	0.7844	0.0588	0.0660	0.0608	0.7544	0.0625	0.1640	0.0774
Average		−0.0021	0.0792	0.0518	0.0794	0.0219	0.0620	0.3230	0.1126

Continued

TABLE 3: Continued

		λ_{GS}				λ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.8025	0.0538	0.0460	0.0538	−0.7595	0.0303	0.0635	0.0506
−0.8	−0.3	−0.2994	0.0519	0.0505	0.0519	−0.3418	0.0388	0.1100	0.0570
−0.8	0	0.0009	0.0438	0.0475	0.0438	−0.0537	0.0367	0.2250	0.0651
−0.8	0.3	0.3005	0.0326	0.0485	0.0326	0.2530	0.0304	0.2855	0.0560
−0.8	0.8	0.8002	0.0111	0.0495	0.0111	0.7874	0.0106	0.1745	0.0164
−0.3	−0.8	−0.7992	0.0400	0.0490	0.0400	−0.7162	0.0283	0.5305	0.0885
−0.3	−0.3	−0.2984	0.0482	0.0450	0.0483	−0.2923	0.0383	0.0175	0.0391
−0.3	0	0.0016	0.0434	0.0440	0.0434	−0.0202	0.0379	0.0455	0.0429
−0.3	0.3	0.3008	0.0354	0.0465	0.0354	0.2703	0.0324	0.0955	0.0440
−0.3	0.8	0.8007	0.0135	0.0485	0.0136	0.7886	0.0127	0.1140	0.0170
0	−0.8	−0.7989	0.0352	0.0500	0.0352	−0.7103	0.0267	0.6985	0.0936
0	−0.3	−0.2974	0.0489	0.0515	0.0490	−0.2675	0.0395	0.0590	0.0512
0	0	0.0023	0.0471	0.0505	0.0472	0.0007	0.0406	0.0225	0.0406
0	0.3	0.3021	0.0391	0.0470	0.0391	0.2826	0.0361	0.0510	0.0400
0	0.8	0.8011	0.0159	0.0500	0.0159	0.7894	0.0152	0.0910	0.0185
0.3	−0.8	−0.7972	0.0317	0.0560	0.0318	−0.7115	0.0260	0.7645	0.0922
0.3	−0.3	−0.2956	0.0494	0.0560	0.0496	−0.2422	0.0420	0.1520	0.0714
0.3	0	0.0039	0.0517	0.0575	0.0518	0.0235	0.0442	0.0575	0.0500
0.3	0.3	0.3035	0.0451	0.0550	0.0452	0.2979	0.0411	0.0330	0.0412
0.3	0.8	0.8020	0.0206	0.0480	0.0207	0.7906	0.0197	0.0650	0.0218
0.8	−0.8	−0.7980	0.0277	0.0590	0.0278	−0.7279	0.0205	0.7180	0.0750
0.8	−0.3	−0.2909	0.0536	0.0730	0.0543	−0.2094	0.0455	0.3025	0.1013
0.8	0	0.0158	0.0645	0.0885	0.0664	0.0697	0.0545	0.1665	0.0885
0.8	0.3	0.3191	0.0688	0.1010	0.0714	0.3409	0.0584	0.0990	0.0713
0.8	0.8	0.8146	0.0430	0.1030	0.0454	0.7999	0.0442	0.1010	0.0442
Average		0.0037	0.0406	0.0568	0.0410	0.0177	0.0340	0.2017	0.0551

TABLE 4: Heteroskedasticity with $c = 1$, Modified Rook Matrix R3 ($n = 485$)

		ρ_{GS}				ρ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.7813	0.0853	0.0630	0.0873	−0.6909	0.0531	0.2560	0.1213
−0.8	−0.3	−0.7885	0.0668	0.0500	0.0678	−0.6307	0.0521	0.8755	0.1771
−0.8	0	−0.7918	0.0592	0.0485	0.0597	−0.6264	0.0524	0.9240	0.1814
−0.8	0.3	−0.7936	0.0556	0.0485	0.0559	−0.6328	0.0517	0.9240	0.1750
−0.8	0.8	−0.7935	0.0517	0.0510	0.0521	−0.6552	0.0468	0.8845	0.1522
−0.3	−0.8	−0.2858	0.1008	0.0560	0.1018	−0.3133	0.0788	0.0375	0.0799
−0.3	−0.3	−0.2907	0.1052	0.0590	0.1056	−0.2420	0.0696	0.0815	0.0906
−0.3	0	−0.2919	0.1026	0.0590	0.1029	−0.2163	0.0703	0.1300	0.1093
−0.3	0.3	−0.2912	0.0994	0.0580	0.0998	−0.2029	0.0730	0.1860	0.1215
−0.3	0.8	−0.2899	0.0921	0.0610	0.0926	−0.2138	0.0730	0.1810	0.1130
0	−0.8	0.0103	0.1026	0.0570	0.1031	−0.0789	0.0905	0.1490	0.1200
0	−0.3	0.0070	0.1073	0.0595	0.1075	−0.0350	0.0812	0.0610	0.0884
0	0	0.0059	0.1087	0.0620	0.1089	−0.0084	0.0815	0.0490	0.0820
0	0.3	0.0056	0.1070	0.0610	0.1072	0.0108	0.0824	0.0600	0.0831
0	0.8	0.0070	0.1005	0.0580	0.1007	0.0135	0.0829	0.0755	0.0840

Continued

TABLE 4: Continued

		ρ_{GS}				ρ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
0.3	−0.8	0.3068	0.0960	0.0620	0.0962	0.1839	0.0969	0.3425	0.1512
0.3	−0.3	0.3025	0.0989	0.0600	0.0989	0.1976	0.0871	0.2550	0.1344
0.3	0	0.3025	0.1014	0.0595	0.1014	0.2182	0.0855	0.1625	0.1183
0.3	0.3	0.3016	0.1032	0.0595	0.1032	0.2398	0.0858	0.1085	0.1048
0.3	0.8	0.3027	0.0993	0.0580	0.0993	0.2563	0.0898	0.1165	0.0999
0.8	−0.8	0.7994	0.0541	0.0530	0.0541	0.7207	0.0610	0.5010	0.1001
0.8	−0.3	0.7949	0.0558	0.0595	0.0560	0.7061	0.0661	0.5225	0.1149
0.8	0	0.7887	0.0610	0.0700	0.0621	0.7063	0.0672	0.4675	0.1153
0.8	0.3	0.7828	0.0687	0.0700	0.0708	0.7130	0.0705	0.3630	0.1120
0.8	0.8	0.7794	0.0693	0.0720	0.0723	0.7330	0.0836	0.2655	0.1071
Average		0.0040	0.0861	0.0590	0.0867	0.0061	0.0733	0.3192	0.1175

		λ_{GS}				λ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.8131	0.0593	0.0795	0.0607	−0.7748	0.0365	0.0140	0.0443
−0.8	−0.3	−0.3065	0.0710	0.0640	0.0713	−0.3591	0.0517	0.1480	0.0785
−0.8	0	−0.0024	0.0605	0.0535	0.0606	−0.0756	0.0510	0.2875	0.0912
−0.8	0.3	0.2994	0.0459	0.0535	0.0459	0.2350	0.0426	0.3420	0.0777
−0.8	0.8	0.8001	0.0153	0.0560	0.0153	0.7847	0.0139	0.1975	0.0207
−0.3	−0.8	−0.8019	0.0309	0.0455	0.0310	−0.7364	0.0291	0.2970	0.0699
−0.3	−0.3	−0.3024	0.0537	0.0510	0.0538	−0.2926	0.0416	0.0105	0.0422
−0.3	0	−0.0004	0.0538	0.0485	0.0538	−0.0282	0.0438	0.0490	0.0521
−0.3	0.3	0.3006	0.0452	0.0530	0.0452	0.2583	0.0405	0.1250	0.0581
−0.3	0.8	0.8005	0.0168	0.0535	0.0168	0.7847	0.0159	0.1565	0.0221
0	−0.8	−0.8003	0.0255	0.0430	0.0255	−0.7375	0.0255	0.4520	0.0675
0	−0.3	−0.2997	0.0477	0.0525	0.0477	−0.2655	0.0392	0.0395	0.0522
0	0	0.0001	0.0507	0.0535	0.0507	−0.0007	0.0426	0.0200	0.0426
0	0.3	0.3008	0.0454	0.0540	0.0454	0.2759	0.0399	0.0555	0.0466
0	0.8	0.8009	0.0191	0.0535	0.0192	0.7853	0.0181	0.1275	0.0233
0.3	−0.8	−0.7999	0.0228	0.0495	0.0228	−0.7424	0.0230	0.5150	0.0620
0.3	−0.3	−0.2970	0.0476	0.0540	0.0477	−0.2431	0.0411	0.1240	0.0701
0.3	0	0.0041	0.0535	0.0555	0.0536	0.0257	0.0454	0.0375	0.0521
0.3	0.3	0.3036	0.0498	0.0565	0.0499	0.2950	0.0449	0.0345	0.0451
0.3	0.8	0.8013	0.0235	0.0545	0.0235	0.7868	0.0230	0.0945	0.0265
0.8	−0.8	−0.7993	0.0228	0.0535	0.0229	−0.7542	0.0201	0.4745	0.0500
0.8	−0.3	−0.2925	0.0542	0.0730	0.0547	−0.2156	0.0484	0.2940	0.0973
0.8	0	0.0153	0.0711	0.0910	0.0728	0.0717	0.0590	0.1730	0.0928
0.8	0.3	0.3237	0.0816	0.1190	0.0850	0.3436	0.0660	0.1030	0.0791
0.8	0.8	0.8180	0.0476	0.1100	0.0509	0.7982	0.0547	0.1515	0.0547
Average		0.0021	0.0446	0.0612	0.0451	0.0088	0.0383	0.1729	0.0568

In contrast, the results for λ_{ML} and ρ_{ML} in the heteroskedastic cases show that the biases are typically large and, consequently, the rejection rates, especially for ρ_{ML}, deviate from the theoretical 0.05 level in many of the considered experiments. Indeed, in Tables 3–5 the rejection rates corresponding to ρ_{ML} exceed 0.9 in some experiments. In most of these experiments, the value of $\rho = -0.8$. The rejection rates relating to λ_{ML} are more moderate but still have outlier values, ranging up to 0.8876 in Table 5. In most of these cases either ρ or λ or both are negative. Interestingly, extreme rejection rates, say over 0.8, for ρ_{ML} and λ_{ML} do not always occur for the same set of parameter values—see,

TABLE 5: Heteroskedasticity with $c = 1$, Modified Rook Matrix R4 ($n = 945$)

		ρ_{GS}				ρ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.7895	0.0639	0.0525	0.0647	−0.6909	0.0393	0.5945	0.1160
−0.8	−0.3	−0.7982	0.0495	0.0400	0.0495	−0.6243	0.0390	0.9990	0.1799
−0.8	0	−0.7994	0.0442	0.0415	0.0442	−0.6187	0.0397	1.0000	0.1856
−0.8	0.3	−0.7993	0.0400	0.0425	0.0400	−0.6250	0.0380	1.0000	0.1791
−0.8	0.8	−0.7996	0.0376	0.0470	0.0376	−0.6488	0.0343	0.9975	0.1551
−0.3	−0.8	−0.2930	0.0778	0.0470	0.0781	−0.3106	0.0593	0.0330	0.0602
−0.3	−0.3	−0.2958	0.0770	0.0490	0.0771	−0.2370	0.0518	0.1320	0.0816
−0.3	0	−0.2972	0.0765	0.0520	0.0765	−0.2065	0.0522	0.2940	0.1071
−0.3	0.3	−0.2969	0.0743	0.0475	0.0744	−0.1899	0.0530	0.4330	0.1222
−0.3	0.8	−0.2967	0.0672	0.0505	0.0673	−0.2001	0.0541	0.4100	0.1136
0	−0.8	0.0068	0.0769	0.0520	0.0772	−0.0779	0.0675	0.2420	0.1031
0	−0.3	0.0050	0.0824	0.0505	0.0825	−0.0318	0.0592	0.0625	0.0672
0	0	0.0026	0.0800	0.0545	0.0801	−0.0006	0.0581	0.0380	0.0581
0	0.3	0.0033	0.0802	0.0555	0.0803	0.0240	0.0593	0.0545	0.0640
0	0.8	0.0040	0.0747	0.0520	0.0748	0.0302	0.0621	0.0795	0.0690
0.3	−0.8	0.3064	0.0708	0.0485	0.0711	0.1836	0.0690	0.5340	0.1353
0.3	−0.3	0.3033	0.0758	0.0510	0.0759	0.1974	0.0638	0.4130	0.1208
0.3	0	0.3020	0.0788	0.0485	0.0789	0.2220	0.0620	0.2635	0.0996
0.3	0.3	0.3011	0.0789	0.0505	0.0789	0.2484	0.0620	0.1350	0.0807
0.3	0.8	0.3034	0.0736	0.0525	0.0737	0.2721	0.0656	0.0990	0.0713
0.8	−0.8	0.8019	0.0389	0.0460	0.0389	0.7198	0.0462	0.7115	0.0925
0.8	−0.3	0.7999	0.0402	0.0435	0.0402	0.7036	0.0493	0.7560	0.1083
0.8	0	0.7979	0.0423	0.0485	0.0423	0.7053	0.0500	0.6895	0.1071
0.8	0.3	0.7951	0.0469	0.0535	0.0472	0.7147	0.0509	0.5570	0.0994
0.8	0.8	0.7892	0.0548	0.0545	0.0559	0.7377	0.0612	0.2700	0.0873
Average		0.0023	0.0641	0.0492	0.0643	0.0119	0.0539	0.4319	0.1066

		λ_{GS}				λ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.8104	0.0452	0.0740	0.0464	−0.7735	0.0279	0.0330	0.0385
−0.8	−0.3	−0.3040	0.0556	0.0535	0.0557	−0.3621	0.0387	0.2695	0.0732
−0.8	0	−0.0017	0.0488	0.0525	0.0488	−0.0825	0.0390	0.5250	0.0912
−0.8	0.3	0.2997	0.0372	0.0515	0.0372	0.2270	0.0332	0.6190	0.0802
−0.8	0.8	0.8004	0.0130	0.0515	0.0130	0.7812	0.0120	0.4055	0.0223
−0.3	−0.8	−0.8020	0.0221	0.0550	0.0222	−0.7334	0.0218	0.7245	0.0701
−0.3	−0.3	−0.3016	0.0405	0.0505	0.0405	−0.2898	0.0300	0.0155	0.0317
−0.3	0	−0.0016	0.0412	0.0465	0.0412	−0.0297	0.0319	0.0770	0.0436
−0.3	0.3	0.2994	0.0352	0.0495	0.0352	0.2517	0.0296	0.2605	0.0567
−0.3	0.8	0.8003	0.0141	0.0505	0.0141	0.7802	0.0137	0.3240	0.0241
0	−0.8	−0.8010	0.0184	0.0535	0.0185	−0.7350	0.0185	0.8600	0.0675
0	−0.3	−0.3009	0.0359	0.0505	0.0359	−0.2609	0.0288	0.0815	0.0485
0	0	−0.0004	0.0394	0.0495	0.0394	−0.0013	0.0314	0.0155	0.0314
0	0.3	0.2996	0.0357	0.0440	0.0357	0.2705	0.0303	0.0850	0.0423
0	0.8	0.8004	0.0161	0.0490	0.0161	0.7808	0.0157	0.2290	0.0247
0.3	−0.8	−0.8005	0.0172	0.0535	0.0172	−0.7406	0.0165	0.8870	0.0617
0.3	−0.3	−0.3006	0.0360	0.0590	0.0361	−0.2365	0.0298	0.3075	0.0701
0.3	0	0.0001	0.0394	0.0495	0.0394	0.0280	0.0331	0.0515	0.0434

Continued

TABLE 5: Continued

		λ_{GS}				λ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
0.3	0.3	0.3005	0.0391	0.0480	0.0391	0.2930	0.0335	0.0205	0.0342
0.3	0.8	0.8008	0.0195	0.0470	0.0196	0.7828	0.0198	0.1350	0.0262
0.8	−0.8	−0.8009	0.0162	0.0555	0.0162	−0.7549	0.0150	0.8085	0.0476
0.8	−0.3	−0.2988	0.0389	0.0545	0.0389	−0.2119	0.0354	0.5825	0.0950
0.8	0	0.0051	0.0497	0.0625	0.0499	0.0746	0.0432	0.3110	0.0862
0.8	0.3	0.3101	0.0556	0.0760	0.0565	0.3446	0.0462	0.1120	0.0642
0.8	0.8	0.8093	0.0386	0.0790	0.0397	0.7956	0.0418	0.0945	0.0421
Average		0.0001	0.0339	0.0546	0.0341	0.0079	0.0287	0.3134	0.0527

TABLE 6: Homoskedasticity with $c = 1$, Modified Rook Matrix R1 ($n = 486$)

		ρ_{GS}				ρ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.7724	0.1019	0.0895	0.1056	−0.7932	0.0855	0.0815	0.0858
−0.8	−0.3	−0.7939	0.0704	0.0555	0.0707	−0.7970	0.0588	0.0455	0.0589
−0.8	0	−0.7979	0.0642	0.0485	0.0643	−0.7966	0.0565	0.0545	0.0566
−0.8	0.3	−0.7994	0.0610	0.0480	0.0610	−0.7970	0.0532	0.0580	0.0533
−0.8	0.8	−0.8003	0.0576	0.0550	0.0576	−0.7978	0.0513	0.0590	0.0514
−0.3	−0.8	−0.2926	0.1124	0.0555	0.1127	−0.3038	0.1108	0.0450	0.1108
−0.3	−0.3	−0.2981	0.1052	0.0500	0.1052	−0.3022	0.1031	0.0475	0.1031
−0.3	0	−0.3010	0.1032	0.0505	0.1032	−0.3026	0.0989	0.0475	0.0990
−0.3	0.3	−0.3026	0.0975	0.0510	0.0975	−0.3026	0.0956	0.0520	0.0956
−0.3	0.8	−0.3033	0.0901	0.0545	0.0901	−0.3019	0.0890	0.0545	0.0890
0	−0.8	0.0015	0.1038	0.0490	0.1038	−0.0048	0.1023	0.0420	0.1024
0	−0.3	−0.0015	0.1069	0.0510	0.1069	−0.0041	0.1048	0.0465	0.1049
0	0	−0.0038	0.1054	0.0490	0.1055	−0.0037	0.1029	0.0485	0.1030
0	0.3	−0.0059	0.1036	0.0500	0.1037	−0.0041	0.1019	0.0480	0.1020
0	0.8	−0.0063	0.0935	0.0535	0.0937	−0.0035	0.0924	0.0515	0.0925
0.3	−0.8	0.2994	0.0885	0.0495	0.0885	0.2975	0.0871	0.0455	0.0872
0.3	−0.3	0.2960	0.0944	0.0545	0.0945	0.2980	0.0940	0.0530	0.0940
0.3	0	0.2938	0.0949	0.0550	0.0951	0.2958	0.0954	0.0465	0.0955
0.3	0.3	0.2924	0.0959	0.0495	0.0962	0.2950	0.0948	0.0460	0.0949
0.3	0.8	0.2933	0.0884	0.0485	0.0887	0.2967	0.0878	0.0470	0.0879
0.8	−0.8	0.7995	0.0427	0.0500	0.0427	0.7972	0.0399	0.0530	0.0400
0.8	−0.3	0.7956	0.0503	0.0580	0.0505	0.7960	0.0448	0.0500	0.0449
0.8	0	0.7906	0.0576	0.0635	0.0583	0.7955	0.0503	0.0525	0.0505
0.8	0.3	0.7841	0.0657	0.0720	0.0676	0.7939	0.0564	0.0560	0.0568
0.8	0.8	0.7791	0.0656	0.0660	0.0688	0.7962	0.0674	0.0900	0.0675
Average		−0.0021	0.0848	0.0551	0.0853	−0.0021	0.0810	0.0528	0.0811
		λ_{GS}				λ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	−0.8	−0.8282	0.0736	0.1220	0.0788	−0.8010	0.0689	0.0835	0.0689
−0.8	−0.3	−0.3105	0.0640	0.0655	0.0649	−0.3053	0.0584	0.0480	0.0587
−0.8	0	−0.0041	0.0510	0.0550	0.0511	−0.0041	0.0484	0.0425	0.0486
−0.8	0.3	0.2990	0.0373	0.0520	0.0373	0.2973	0.0358	0.0455	0.0359

Continued

TABLE 6: Continued

		λ_{GS}				λ_{ML}			
Rho	Lambda	Median	Std. Err	Rej. Rate	RMSE	Median	Std. Err	Rej. Rate	RMSE
−0.8	0.8	0.8001	0.0119	0.0515	0.0119	0.7990	0.0112	0.0425	0.0112
−0.3	−0.8	−0.8081	0.0542	0.0600	0.0548	−0.7975	0.0497	0.0520	0.0498
−0.3	−0.3	−0.3051	0.0619	0.0550	0.0621	−0.3017	0.0618	0.0500	0.0618
−0.3	0	−0.0021	0.0537	0.0505	0.0537	−0.0026	0.0528	0.0465	0.0528
−0.3	0.3	0.2996	0.0417	0.0495	0.0417	0.2981	0.0400	0.0440	0.0400
−0.3	0.8	0.8000	0.0140	0.0495	0.0140	0.7990	0.0132	0.0445	0.0133
0	−0.8	−0.8040	0.0479	0.0545	0.0481	−0.7971	0.0424	0.0450	0.0425
0	−0.3	−0.3019	0.0638	0.0535	0.0639	−0.3011	0.0639	0.0480	0.0639
0	0	−0.0017	0.0587	0.0525	0.0587	−0.0018	0.0578	0.0495	0.0578
0	0.3	0.2997	0.0464	0.0510	0.0464	0.2980	0.0459	0.0460	0.0460
0	0.8	0.8002	0.0165	0.0480	0.0165	0.7984	0.0154	0.0430	0.0155
0.3	−0.8	−0.8019	0.0448	0.0515	0.0449	−0.7977	0.0387	0.0455	0.0388
0.3	−0.3	−0.2987	0.0676	0.0600	0.0676	−0.2988	0.0683	0.0475	0.0683
0.3	0	0.0018	0.0660	0.0575	0.0660	0.0003	0.0666	0.0500	0.0666
0.3	0.3	0.3020	0.0553	0.0595	0.0554	0.2983	0.0551	0.0510	0.0551
0.3	0.8	0.8006	0.0211	0.0485	0.0212	0.7980	0.0202	0.0460	0.0203
0.8	−0.8	−0.7994	0.0403	0.0555	0.0403	−0.7965	0.0321	0.0515	0.0323
0.8	−0.3	−0.2894	0.0820	0.0765	0.0827	−0.2946	0.0698	0.0550	0.0700
0.8	0	0.0188	0.0974	0.0940	0.0992	0.0055	0.0808	0.0575	0.0810
0.8	0.3	0.3276	0.0997	0.1055	0.1034	0.3039	0.0867	0.0695	0.0868
0.8	0.8	0.8176	0.0486	0.1045	0.0517	0.7991	0.0525	0.0915	0.0525
Average		0.0005	0.0528	0.0633	0.0534	−0.0002	0.0495	0.0518	0.0495

for example, Tables 3–5. Of course, when they do occur simultaneously either ρ or λ is negative, but typically not both.

Intuitive explanations of the table results thus far discussed as they relate to the values of ρ and λ are not straightforward. As one example, the reduced form for $\mathbf{y}_n$ from the model (1) and (2) is

$$\mathbf{y}_n = (\mathbf{I}_n - \lambda\mathbf{W}_n)^{-1}\mathbf{X}_n\boldsymbol{\beta} + (\mathbf{I}_n - \lambda\mathbf{W}_n)^{-1}(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}\boldsymbol{\varepsilon}_n. \quad (20)$$

If λ is large in absolute value, say close to 1.0, the variances of the elements of error vector in equation (20), namely $(\mathbf{I}_n - \lambda\mathbf{W}_n)^{-1}(\mathbf{I}_n - \rho\mathbf{W}_n)^{-1}\boldsymbol{\varepsilon}_n$, will, ceteris paribus, tend to be large since $\lambda = 1.0$ is a singular point of the inverse matrix. These larger variances will obviously have a negative effect on estimation precision. On the other hand, increased variation of the vector $\mathbf{y}_n$ will, ceteris paribus, increase the variation in $\mathbf{W}_n\mathbf{y}_n$, which is a right-hand-side variable, and this should increase estimation precision. The net effect on estimation precision of a large value of λ will obviously be the result of these two effects, and it is not clear which of these two effects would dominate in a particular case. Similar concerns relate to the value of ρ since, on the negative side, it also enters the error term in (20) in the same fashion as λ; on the positive side ρ can be viewed as a regression parameter in (2) and the larger the value of ρ the more $\mathbf{W}_n\mathbf{u}_n$ varies and so the more precision is increased. Of course, intuitive interpretations of our results are made still more complex by the interactive effects of ρ and λ as is evident in (20).

Returning to the tables, note from Tables 2–5 that, on average, the root mean square errors relating to ρ_{GS} and to λ_{GS} decrease as the sample size increases in every "comparable" case considered. As an example of "comparable" cases, Tables 2 and 3 both relate to a north-east modified rook matrix in which the north-east quadrant contains 75 percent of the units. The main difference in the design underlying Tables 2 and 3 is the sample size, namely $n = 486$ for Table 2 and $n = 974$ for Table 3. Other comparable tables are Tables 4 and 5. The root mean square errors for ρ_{ML} and λ_{ML} typically decrease in relevant comparisons as the sample size increases. A glance at the tables suggests that the reason for this is that the standard deviations, not the biases, decreases with the sample size.

The "comparable" cases above focused attention on the effects of the sample size in the modified rook matrix cases by holding constant the relative size of the north-east quadrant of those matrices. We now focus attention on comparisons relating to the relative size of the north-east quadrant of those modified rook matrices by holding constant the sample size. For instance, consider the results in Tables 2 and 4. In these tables the sample sizes are 486 and 485, respectively; the proportion of units located in the north-east quadrant are 75 and 25 percent, respectively.

The root mean square errors for ρ_{GS} and λ_{GS} are lower in Table 4 than they are in Table 2, as are the averages of these root mean square errors. The same result holds for the root mean square errors of ρ_{GS} and λ_{GS} in Tables 3 and 5. Thus, the smaller the size of the north-east quadrant, the more precise the estimation is.

Given the complexity of our model and our estimators, there does not seem to be a simple explanation of these results. On an intuitive level, one suspects that the particular values of the instrument matrix and the variances are at least part of the explanation. For example, if $\rho = 0$ the only term in the large sample distribution of $\tilde{\boldsymbol{\delta}}_n$ that would involve the variances would be $\lim_{n\to\infty} n^{-1}\mathbf{H}_n'\boldsymbol{\Sigma}_n\mathbf{H}_n$, which clearly involves the products of the variances and the elements of $\mathbf{H}_n$.

Table 6 contains results for the homoskedastic case in which the weights matrix is a north-east modified rook matrix R1 and $c = 1.0$. Under homoskedasticity the quasi-maximum-likelihood estimator is the maximum-likelihood estimator, and so it is consistent and efficient. Of course, in this case both ρ_{GS} and λ_{GS} are also consistent. Consistent with this, note from Table 6 that the biases are small for all four of the indicated estimators, and the rejection rates are reasonably close to the theoretical 0.05 level. Although the root mean square errors are relatively small for both ρ_{GS} and λ_{GS}, they are typically larger than those of ρ_{ML} and λ_{ML}. On average the root mean square errors of ρ_{GS} and λ_{GS} are 5 and 8 percent, respectively, larger than those of ρ_{ML} and λ_{ML} in Table 6.

6. CONCLUSIONS

In this paper a general Cliff-Ord-type model is considered, which contains spatial lags in both the dependent variable, the disturbance term, and possibly in some of the exogenous variables. In the model the innovations to the disturbance process are assumed to be heteroskedastic of an unknown form. Estimators of the regression parameters and the autoregressive parameter in the disturbance process are suggested. User-friendly expressions are given for asymptotic distributions and for small-sample approximations, which are useful for testing joint hypotheses relating to the regression parameters and/or the autoregressive parameter in the disturbance process. Monte Carlo results are given which suggest that our estimators behave well in small samples; on the other hand, the maximum-likelihood estimator based on the normality assumption may behave quite poorly when the innovations are in fact heteroskedastic in that its biases can be large, which in turn leads to type 1 rejection levels which are large relative to the nominal levels.

REFERENCES

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

Anselin, L. and R. Florax. 1995. "Small Sample Properties of Tests for Spatial Dependence in Regression Models: Some Further Results," in Luc Anselin and Raymond Florax (eds.), *New Directions in Spatial Econometrics*. New York: Springer, pp. 75–95.

Baltagi, B. H., P. Egger, and M. Pfaffermayr. 2007. "Estimating Models of Complex FDI: Are There Third–Country Effects?" *Journal of Econometrics*, 140(1), 260–281.

Cliff, A. D. and J. K. Ord. 1973. *Spatial Autocorrelation*. London: Pion.

——. 1981. *Spatial Processes, Models and Applications*. London: Pion.

Kelejian, H. H. and I. R. Prucha. 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances," *Journal of Real Estate Finance and Economics*, 17(1), 99–121.

——. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model," *International Economic Review*, 40(2), 509–533.

——. 2004. "Estimation of Simultaneous Systems of Spatially Interrelated Cross Sectional Equations," *Journal of Econometrics*, 118(1–2), 27–50.

——. 2007a. "HAC Estimation in a Spatial Framework," *Journal of Econometrics*, 140(1), 131–154.

——. 2007b. "Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances," *Journal of Econometrics*, forthcoming.

Kelejian, H. H. and D. Robinson. 1995. "Spatial Correlation: A Suggested Alternative to the Autoregressive Model," in Luc Anselin and Raymond Florax (eds.), *New Directions in Spatial Econometrics*, New York: Springer, pp. 75–95.

Kelejian, H. H., G. Tavlas, and G. Hondroyiannis. 2006. "A Spatial Modelling Approach Among Emerging Economies," *Open Economies Review*, 17(4), 423–441.

Rey, S. and B. Dev. 2006. "σ-Convergence in the Presence of Spatial Effects," *Papers in Regional Science*, 85(2), 217–234.

Vigil, R. L. 1998. "Interactions Among Municipalities in the Provision of Police Services," Ph.D. dissertation, Department of Economics, University of Maryland.

APPENDIX A: ADDITIONAL ASSUMPTIONS

In this appendix we state the additional assumptions needed to formally establish the limiting distribution of the GMM/IV estimators.

ASSUMPTION A1: *Let* $\mathbf{\Gamma}_n = [\boldsymbol{\gamma}_{rs,n}]_{r,s=1,2}$ *and* $\boldsymbol{\gamma}_n = [\boldsymbol{\gamma}_{1,n}, \boldsymbol{\gamma}_{2,n}]'$, *where dropping the subscript n temporarily for notational convenience,*

$$\text{(A1)}\qquad \begin{aligned}
\boldsymbol{\gamma}_{11} &= 2n^{-1}\mathrm{E}\{\bar{\bar{\mathbf{u}}}'\bar{\mathbf{u}} - Tr[\mathbf{M}[diag_{i=1}^{n}(\bar{\mathbf{u}}_i\mathbf{u}_i)]\mathbf{M}']\} = 2n^{-1}\mathrm{E}\mathbf{u}'\mathbf{M}'\mathbf{A}_1\mathbf{u},\\
\boldsymbol{\gamma}_{12} &= -n^{-1}\mathrm{E}\{\bar{\bar{\mathbf{u}}}'\bar{\bar{\mathbf{u}}} + Tr[\mathbf{M}[diag_{i=1}^{n}(\bar{\mathbf{u}}_i^2)]\mathbf{M}']\} = -n^{-1}\mathrm{E}\mathbf{u}'\mathbf{M}'\mathbf{A}_1\mathbf{M}\mathbf{u},\\
\boldsymbol{\gamma}_{21} &= n^{-1}\mathrm{E}(\mathbf{u}'\bar{\bar{\mathbf{u}}} + \bar{\mathbf{u}}'\bar{\mathbf{u}}) = n^{-1}\mathrm{E}\mathbf{u}'\mathbf{M}'(\mathbf{A}_2 + \mathbf{A}_2')\mathbf{u},\\
\boldsymbol{\gamma}_{22} &= -n^{-1}\mathrm{E}\bar{\mathbf{u}}'\bar{\bar{\mathbf{u}}} = -n^{-1}\mathrm{E}\mathbf{u}'\mathbf{M}'\mathbf{A}_2\mathbf{M}\mathbf{u},\\
\boldsymbol{\gamma}_{1} &= n^{-1}\mathrm{E}\{\bar{\mathbf{u}}'\bar{\mathbf{u}} - Tr[\mathbf{M}[diag_{i=1}^{n}(\mathbf{u}_i^2)]\mathbf{M}']\} = n^{-1}\mathrm{E}\mathbf{u}'\mathbf{A}_1\mathbf{u},\\
\boldsymbol{\gamma}_{2} &= n^{-1}\mathrm{E}\mathbf{u}'\bar{\mathbf{u}} = n^{-1}\mathrm{E}\mathbf{u}'\mathbf{A}_2\mathbf{u},
\end{aligned}$$

with $\bar{\mathbf{u}} = \mathbf{M}\mathbf{u}$, *and* $\bar{\bar{\mathbf{u}}} = \mathbf{M}\bar{\mathbf{u}} = \mathbf{M}^2\mathbf{u}$. *Then* $\mathbf{\Gamma}_n$ *is nonsingular for all n sufficiently large and* $\lim_{n\to\infty}\mathbf{\Gamma}_n = \mathbf{\Gamma}$ *is finite and nonsingular.*

ASSUMPTION A2: *Let* $\mathbf{\Psi}_n = (\psi_{rs,n})$ *where for* $r, s = 1, 2$

$$\text{(A2)}\qquad \psi_{rs,n} = (2n)^{-1}tr\left[\left(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}'\right)\Sigma_n\left(\mathbf{A}_{s,n} + \mathbf{A}_{s,n}'\right)\Sigma_n\right] + n^{-1}\mathbf{a}_{r,n}'\Sigma_n\mathbf{a}_{s,n}$$

with $\mathbf{a}_{r,n} = (\mathbf{I}_n - \rho\mathbf{M}_n')^{-1}\mathbf{H}_n\mathbf{P}\boldsymbol{\alpha}_{r,n}$ *where*

$$\text{(A3)}\qquad \begin{aligned}
\boldsymbol{\alpha}_{r,n} &= -n^{-1}\mathrm{E}\left[\mathbf{Z}_n'\left(\mathbf{I}_n - \rho\mathbf{M}_n'\right)(\mathbf{A}_{r,n} + \mathbf{A}_{r,n}')(\mathbf{I}_n - \rho\mathbf{M}_n)\mathbf{u}_n\right],\\
\mathbf{P} &= \mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ}[\mathbf{Q}_{HZ}'\mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ}]^{-1},
\end{aligned}$$

and $\boldsymbol{\Sigma}_n = diag(\sigma_{i,n}^2)$*, where* $\sigma_{i,n}^2 = \mathrm{E}\boldsymbol{\varepsilon}_{i,n}^2$*. Furthermore, let*

$$\boldsymbol{\Psi}_{\circ,n} = \begin{bmatrix} \boldsymbol{\Psi}_{\Delta\Delta,n} & \boldsymbol{\Psi}_{\Delta\rho,n} \\ \boldsymbol{\Psi}'_{\Delta\rho,n} & \boldsymbol{\Psi}_n \end{bmatrix} \tag{A4}$$

with

$$\begin{aligned} \boldsymbol{\Psi}_{\Delta\Delta,n} &= n^{-1}\mathbf{H}'_n(\mathbf{I}_n - \rho\mathbf{M}_n)^{-1}\boldsymbol{\Sigma}_n\left(\mathbf{I}_n - \rho\mathbf{M}'_n\right)^{-1}\mathbf{H}_n, \\ \boldsymbol{\Psi}_{\Delta\rho,n} &= n^{-1}\mathbf{H}'_n(\mathbf{I}_n - \rho\mathbf{M}_n)^{-1}\boldsymbol{\Sigma}_n[\mathbf{a}_{1,n}, \mathbf{a}_{2,n}]. \end{aligned}$$

Then $\boldsymbol{\Psi}_n$ *and* $\boldsymbol{\Psi}_{\circ,n}$ *are nonsingular for all n sufficiently large and* $\lim_{n\to\infty}\boldsymbol{\Psi}_n = \boldsymbol{\Psi}$ *and* $\lim_{n\to\infty}\boldsymbol{\Psi}_{\circ,n} = \boldsymbol{\Psi}_\circ$ *are finite and nonsingular.*

ASSUMPTION A3: *Let* $\boldsymbol{\Psi}_n = (\psi_{rs,n})$ *where for* $r, s = 1, 2$

$$\psi_{rs,n} = (2n)^{-1}tr\left[\left(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n}\right)\boldsymbol{\Sigma}_n\left(\mathbf{A}_{s,n} + \mathbf{A}'_{s,n}\right)\boldsymbol{\Sigma}_n\right] + n^{-1}\mathbf{a}'_{r,n}\boldsymbol{\Sigma}_n\mathbf{a}_{s,n} \tag{A5}$$

with

$$\begin{aligned} \mathbf{a}_{r,n} &= \mathbf{H}_n\mathbf{P}\boldsymbol{\alpha}_{r,n}, \\ \boldsymbol{\alpha}_{r,n} &= -n^{-1}\mathrm{E}\left[\mathbf{Z}'_n\left(\mathbf{I}_n - \rho\mathbf{M}'_n\right)(\mathbf{A}_{r,n} + \mathbf{A}'_{r,n})(\mathbf{I}_n - \rho\mathbf{M}_n)\mathbf{u}_n\right], \\ \mathbf{P} &= \mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ*}\left[\mathbf{Q}'_{HZ*}\mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ*}\right]^{-1}, \end{aligned} \tag{A6}$$

and $\boldsymbol{\Sigma}_n = diag(\sigma_{i,n}^2)$*, where* $\sigma_{i,n}^2 = \mathrm{E}\boldsymbol{\varepsilon}_{i,n}^2$*. Furthermore, let*

$$\boldsymbol{\Psi}_{\circ,n} = \begin{bmatrix} \boldsymbol{\Psi}_{\boldsymbol{\delta}\boldsymbol{\delta},n} & \boldsymbol{\Psi}_{\boldsymbol{\delta}\rho,n} \\ \boldsymbol{\Psi}'_{\boldsymbol{\delta}\rho,n} & \boldsymbol{\Psi}_n \end{bmatrix}, \tag{A7}$$

with

$$\begin{aligned} \boldsymbol{\Psi}_{\boldsymbol{\delta}\boldsymbol{\delta},n} &= n^{-1}\mathbf{H}'_n\boldsymbol{\Sigma}_n\mathbf{H}_n, \\ \boldsymbol{\Psi}_{\boldsymbol{\delta}\rho,n} &= n^{-1}\mathbf{H}'_n\boldsymbol{\Sigma}_n\left[\mathbf{a}_{1,n}, \mathbf{a}_{2,n}\right]. \end{aligned}$$

Then $\boldsymbol{\Psi}_n$ *and* $\boldsymbol{\Psi}_{\circ,n}$ *are nonsingular for all n sufficiently large and* $\lim_{n\to\infty}\boldsymbol{\Psi}_n = \boldsymbol{\Psi}$ *and* $\lim_{n\to\infty}\boldsymbol{\Psi}_{\circ,n} = \boldsymbol{\Psi}_\circ$ *are finite and nonsingular.*

APPENDIX B: ESTIMATORS FOR $\boldsymbol{\Psi}$ AND $\boldsymbol{\Omega}$

For simplicity of notation we drop subscript n in the following.

Definition of $\mathbf{G}$ *and* $\mathbf{g}$

Let $\tilde{\boldsymbol{\delta}}$ be some estimator for $\boldsymbol{\delta}$, let $\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\tilde{\boldsymbol{\delta}}$ be the corresponding estimated residuals, and let $\tilde{\bar{\mathbf{u}}}_n = \mathbf{M}_n\tilde{\mathbf{u}}_n$, $\tilde{\bar{\bar{\mathbf{u}}}}_n = \mathbf{M}_n^2\tilde{\mathbf{u}}_n$. Then, $\mathbf{G}(\tilde{\boldsymbol{\delta}}) = [\mathbf{g}_{rs}(\tilde{\boldsymbol{\delta}})]_{r,s=1,2}$ and $\mathbf{g}(\tilde{\boldsymbol{\delta}}) = [\mathbf{g}_1(\tilde{\boldsymbol{\delta}}), \mathbf{g}_2(\tilde{\boldsymbol{\delta}})]'$ are obtained from the expressions for the elements of $\boldsymbol{\Gamma} = [\boldsymbol{\gamma}_{rs}]_{r,s=1,2}$ and $\boldsymbol{\gamma} = [\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2]'$ in (A1) by suppressing the expectations operator, and replacing the disturbance vectors $\mathbf{u}$, $\bar{\mathbf{u}}$, and $\bar{\bar{\mathbf{u}}}$ by their predictors $\tilde{\mathbf{u}}$, $\tilde{\bar{\mathbf{u}}}$, and $\tilde{\bar{\bar{\mathbf{u}}}}$.

Definition of $\tilde{\boldsymbol{\Psi}}$ *and* $\tilde{\boldsymbol{\Omega}}$

Let $\tilde{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\tilde{\boldsymbol{\delta}}$ denote the 2SLS residuals and let $\bar{\rho}$ be some estimator for ρ. Then $\tilde{\boldsymbol{\Psi}} = [\tilde{\psi}_{rs}]_{r,s=1,2}$ with

$$\tilde{\psi}_{rs} = (2n)^{-1}tr\left[\left(\mathbf{A}_r + \mathbf{A}'_r\right)\tilde{\boldsymbol{\Sigma}}\left(\mathbf{A}_s + \mathbf{A}'_s\right)\tilde{\boldsymbol{\Sigma}}\right] + n^{-1}\tilde{\mathbf{a}}'_r\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{a}}_s, \tag{B1}$$

where

$$\widetilde{\boldsymbol{\Sigma}} = diag_{i=1,\ldots,n}\left(\widetilde{\boldsymbol{\varepsilon}}_i^2\right) \qquad \widetilde{\boldsymbol{\varepsilon}} = (\mathbf{I} - \bar{\rho}\mathbf{M})\,\widetilde{\mathbf{u}}$$
$$\widetilde{\mathbf{a}}_r = (\mathbf{I} - \bar{\rho}\mathbf{M}')^{-1}\,\mathbf{H}\widetilde{\mathbf{P}}\widetilde{\boldsymbol{\alpha}}_r \qquad \widetilde{\boldsymbol{\alpha}}_r = -n^{-1}\left[\mathbf{Z}'(\mathbf{I} - \bar{\rho}\mathbf{M}')\left(\mathbf{A}_r + \mathbf{A}_r'\right)(\mathbf{I} - \bar{\rho}\mathbf{M})\,\widetilde{\mathbf{u}}\right]$$

and

$$\widetilde{\mathbf{P}} = (n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z})\left[(n^{-1}\mathbf{Z}'\mathbf{H})(n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z})\right]^{-1}.$$

Let $\widetilde{\boldsymbol{\Gamma}} = \mathbf{G}(\widetilde{\boldsymbol{\delta}})$, where $\mathbf{G}$ (.) is defined in Appendix B: Definition of $\mathbf{G}$ and $\mathbf{g}$, and let $\widetilde{\mathbf{J}} = \widetilde{\boldsymbol{\Gamma}}[1, 2\bar{\rho}]'$, then the estimator $\widetilde{\boldsymbol{\Omega}}$ is given by

(B2)
$$\widetilde{\boldsymbol{\Omega}} = \begin{bmatrix} \widetilde{\mathbf{P}}' & 0 \\ 0 & (\widetilde{\mathbf{J}}'\widetilde{\boldsymbol{\Psi}}^{-1}\widetilde{\mathbf{J}})^{-1}\widetilde{\mathbf{J}}'\widetilde{\boldsymbol{\Psi}}^{-1} \end{bmatrix} \widetilde{\boldsymbol{\Psi}}_{\circ} \begin{bmatrix} \widetilde{\mathbf{P}} & 0 \\ 0 & \widetilde{\boldsymbol{\Psi}}^{-1}\widetilde{\mathbf{J}}(\widetilde{\mathbf{J}}'\widetilde{\boldsymbol{\Psi}}^{-1}\widetilde{\mathbf{J}})^{-1} \end{bmatrix},$$
$$\widetilde{\boldsymbol{\Psi}}_{\circ} = \begin{bmatrix} \widetilde{\boldsymbol{\Psi}}_{\delta\delta} & \widetilde{\boldsymbol{\Psi}}_{\delta\rho} \\ \widetilde{\boldsymbol{\Psi}}_{\delta\rho}' & \widetilde{\boldsymbol{\Psi}} \end{bmatrix}, \quad \widetilde{\boldsymbol{\Psi}}_{\delta\delta} = n^{-1}\mathbf{H}'(\mathbf{I} - \bar{\rho}\mathbf{M})\,\widetilde{\boldsymbol{\Sigma}}\,(\mathbf{I} - \bar{\rho}\mathbf{M}')\,\mathbf{H},$$
$$\widetilde{\boldsymbol{\Psi}}_{\delta\rho} = n^{-1}\mathbf{H}'(\mathbf{I} - \bar{\rho}\mathbf{M})\,\widetilde{\boldsymbol{\Sigma}}\,[\widetilde{\mathbf{a}}_1, \widetilde{\mathbf{a}}_2].$$

We will also write $\widetilde{\boldsymbol{\Psi}}(\bar{\rho})$, $\widetilde{\boldsymbol{\Omega}}(\bar{\rho})$, and $\widetilde{\boldsymbol{\Psi}}_{\circ}(\bar{\rho})$ for $\widetilde{\boldsymbol{\Psi}}$, $\widetilde{\boldsymbol{\Omega}}$, and $\widetilde{\boldsymbol{\Psi}}_{\circ}$ to explicitly denote the dependence on $\bar{\rho}$.

Definition of $\widehat{\boldsymbol{\Psi}}$ and $\widehat{\boldsymbol{\Omega}}$

Let $\widehat{\mathbf{u}} = \mathbf{y} - \mathbf{Z}\widehat{\boldsymbol{\delta}}$ denote the GS2SLS residuals and let $\bar{\rho}$ be some estimator for ρ. Then $\widehat{\boldsymbol{\Psi}} = [\widehat{\psi}_{rs}]_{r,s=1,2}$ with

$$\widehat{\psi}_{rs} = (2n)^{-1}tr\left[\left(\mathbf{A}_r + \mathbf{A}_r'\right)\widehat{\boldsymbol{\Sigma}}\left(\mathbf{A}_s + \mathbf{A}_s'\right)\widehat{\boldsymbol{\Sigma}}\right] + n^{-1}\widehat{\mathbf{a}}_r'\widehat{\boldsymbol{\Sigma}}\widehat{\mathbf{a}}_s,$$

where

$$\widehat{\boldsymbol{\Sigma}} = diag_{i=1,\ldots,n}\left(\widehat{\boldsymbol{\varepsilon}}_i^2\right) \qquad \widehat{\boldsymbol{\varepsilon}} = (\mathbf{I} - \bar{\rho}\mathbf{M})\,\widehat{\mathbf{u}}$$
$$\widehat{\mathbf{a}}_r = \mathbf{H}\widehat{\mathbf{P}}\widehat{\boldsymbol{\alpha}}_r \qquad \widehat{\boldsymbol{\alpha}}_r = -n^{-1}\left[\mathbf{Z}'(\mathbf{I} - \bar{\rho}\mathbf{M}')\left(\mathbf{A}_r + \mathbf{A}_r'\right)(\mathbf{I} - \bar{\rho}\mathbf{M})\,\widehat{\mathbf{u}}\right]$$

and

$$\widehat{\mathbf{P}} = (n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z}_*(\bar{\rho}))\left[(n^{-1}\mathbf{Z}_*'(\bar{\rho})\mathbf{H})(n^{-1}\mathbf{H}'\mathbf{H})^{-1}(n^{-1}\mathbf{H}'\mathbf{Z}_*(\bar{\rho}))\right]^{-1},$$

and $\mathbf{Z}_*(\bar{\rho}) = \mathbf{Z} - \bar{\rho}\mathbf{MZ}$. Let $\widehat{\boldsymbol{\Gamma}} = \mathbf{G}(\widehat{\boldsymbol{\delta}})$, where $\mathbf{G}$ (.) is defined in Appendix B: Definition of $\mathbf{G}$ and $\mathbf{g}$, then the estimator $\widehat{\boldsymbol{\Omega}}$ is given by

$$\widehat{\boldsymbol{\Omega}} = \begin{bmatrix} \widehat{\mathbf{P}}' & 0 \\ 0 & (\widehat{\mathbf{J}}'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\mathbf{J}})^{-1}\widehat{\mathbf{J}}'\widehat{\boldsymbol{\Psi}}^{-1} \end{bmatrix} \widehat{\boldsymbol{\Psi}}_{\circ} \begin{bmatrix} \widehat{\mathbf{P}} & 0 \\ 0 & \widehat{\boldsymbol{\Psi}}^{-1}\widehat{\mathbf{J}}(\widehat{\mathbf{J}}'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\mathbf{J}})^{-1} \end{bmatrix},$$
$$\widehat{\boldsymbol{\Psi}}_{\circ} = \begin{bmatrix} \widehat{\boldsymbol{\Psi}}_{\delta\delta} & \widehat{\boldsymbol{\Psi}}_{\delta\rho} \\ \widehat{\boldsymbol{\Psi}}_{\delta\rho}' & \widehat{\boldsymbol{\Psi}} \end{bmatrix}, \qquad \widehat{\boldsymbol{\Psi}}_{\delta\delta} = n^{-1}\mathbf{H}'\widehat{\boldsymbol{\Sigma}}\mathbf{H}, \qquad \widehat{\boldsymbol{\Psi}}_{\delta\rho} = n^{-1}\mathbf{H}'\widehat{\boldsymbol{\Sigma}}\,[\widehat{\mathbf{a}}_1, \widehat{\mathbf{a}}_2].$$

We will also write $\widehat{\boldsymbol{\Psi}}(\bar{\rho})$, $\widehat{\boldsymbol{\Omega}}(\bar{\rho})$, and $\widehat{\boldsymbol{\Psi}}_{\circ}(\bar{\rho})$ for $\widehat{\boldsymbol{\Psi}}$, $\widehat{\boldsymbol{\Omega}}$, and $\widehat{\boldsymbol{\Psi}}_{\circ}$ to explicitly denote the dependence on $\bar{\rho}$.

APPENDIX C: PROOFS

Proof of Theorem 1: Consider the 2SLS residuals $\widetilde{\mathbf{u}}_n = \mathbf{y}_n - \mathbf{Z}_n\widetilde{\boldsymbol{\delta}}_n$. Then clearly $\widetilde{\mathbf{u}}_n - \mathbf{u}_n = \mathbf{D}_n\boldsymbol{\Delta}_n$ with $\mathbf{D}_n = -\mathbf{Z}_n$ and $\boldsymbol{\Delta}_n = \widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta}$. Next observe that under our Assumptions 1–3 and 4–6, Assumptions 1–3 and 8–10 in Kelejian and Prucha (2007b) clearly hold. Since $\boldsymbol{\beta}$ does not vary with n it now follows directly from Lemma 3 in Kelejian and Prucha

(2007b) that the fourth moments of the elements of $\mathbf{D}_n = -\mathbf{Z}_n$ are uniformly bounded, that Assumption 6 in Kelejian and Prucha (2007b) holds, and

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

$$\mathbf{P} = \mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ}\left[\mathbf{Q}_{HZ}'\mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ}\right]^{-1},$$
$$\mathbf{F}_n = \left(\mathbf{I}_n - \rho\mathbf{M}_n'\right)^{-1}\mathbf{H}_n.$$

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

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

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

From this we see that also Assumptions 4 and 7 in Kelejian and Prucha (2007b) are satisfied.

By Assumption A1 we have $\boldsymbol{\Gamma}_n$ is nonsingular for all n sufficiently large and $\lim_{n\to\infty}\boldsymbol{\Gamma}_n = \boldsymbol{\Gamma}$ is finite and nonsingular. Consequently, the $\lambda_{\min}(\boldsymbol{\Gamma}_n'\boldsymbol{\Gamma}_n) \geq \text{const} > 0$ for n sufficiently large and thus also Assumption 5(a) in Kelejian and Prucha (2007b) holds. Furthermore, observe that for $\widetilde{\Upsilon}_n = \Upsilon_n = \mathbf{I}_2$ also Assumptions 5(b),(c) in Kelejian and Prucha (2007b) are trivially satisfied.

By Assumption A2 $\boldsymbol{\Psi}_n$ and $\boldsymbol{\Psi}_{\circ,n}$ are nonsingular for all n sufficiently large and $\lim_{n\to\infty}\boldsymbol{\Psi}_n = \boldsymbol{\Psi}$ and $\lim_{n\to\infty}\boldsymbol{\Psi}_{\circ,n} = \boldsymbol{\Psi}_\circ$ are finite and nonsingular, and thus the smallest [largest] eigenvalues of $\boldsymbol{\Psi}_n$, $\boldsymbol{\Psi}_n^{-1}$, $\boldsymbol{\Psi}_{\circ,n}$, and $\boldsymbol{\Psi}_{\circ,n}^{-1}$ are bounded away from zero [bounded from above] for sufficiently large n.

It now follows immediately from Theorems 1–3 in Kelejian and Prucha (2007b) that the initial GMM estimator for ρ, $\check{\rho}_n$, is $n^{1/2}$-consistent and that $\operatorname{plim}_{n\to\infty}\widetilde{\boldsymbol{\Psi}}_n(\check{\rho}_n) = \boldsymbol{\Psi}$ and $\operatorname{plim}_{n\to\infty}\widetilde{\boldsymbol{\Psi}}_n^{-1}(\check{\rho}_n) = \boldsymbol{\Psi}^{-1}$.

The estimator $\widetilde{\rho}_n$ is a special case of the GMM estimators for ρ defined in equation (9) in Kelejian and Prucha (2007b) with $\widetilde{\Upsilon}_n = \widetilde{\boldsymbol{\Psi}}_n^{-1}(\widetilde{\rho}_n)$ and $\Upsilon_n = \boldsymbol{\Psi}_n^{-1}$. Recalling that the smallest [largest] eigenvalues of $\boldsymbol{\Psi}_n^{-1}$ bounded away from zero [bounded from above] for sufficiently large n we see from Theorem 3 in Kelejian and Prucha (2007b) that also in this case Assumption 5(b),(c) in that paper are satisfied. All other assumptions maintained by Theorems 1–3 in Kelejian and Prucha (2007b) have already been verified, which establishes $n^{1/2}$-consistency of $\widetilde{\rho}_n$ and its asymptotic efficiency.

The joint limiting distribution of $n^{1/2}(\widetilde{\boldsymbol{\delta}}_n - \boldsymbol{\delta})$ and $n^{1/2}(\widetilde{\rho}_n - \rho)$ given by the theorem now follows immediately from Theorem 4 in Kelejian and Prucha (2007b).

Proof of Theorem 2: Consider the GS2SLS residuals $\widehat{\mathbf{u}}_n = \mathbf{y}_n - \mathbf{Z}_n\widehat{\boldsymbol{\delta}}_n$. Then clearly $\widehat{\mathbf{u}}_n - \mathbf{u}_n = \mathbf{D}_n\boldsymbol{\Delta}_n$ with $\mathbf{D}_n = -\mathbf{Z}_n$ and $\boldsymbol{\Delta}_n = \widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta}$. As in the proof of Theorem 1, observe that under our Assumptions 1–3 and 4–6, Assumptions 1–3 and 8–10 in Kelejian and Prucha (2007b) clearly hold. Also, recall that in the proof of Theorem 1 we have established that the fourth moments of the elements of $\mathbf{D}_n = -\mathbf{Z}_n$ are uniformly bounded, and that Assumption 6 in Kelejian and Prucha (2007b) holds. It now follows from Lemma 4 in Kelejian and Prucha (2007b) that[5]

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

$$\mathbf{P} = \mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ^*}(\rho)\left[\mathbf{Q}_{HZ^*}'(\rho)\mathbf{Q}_{HH}^{-1}\mathbf{Q}_{HZ^*}(\rho)\right]^{-1},$$
$$\mathbf{F}_n = \mathbf{H}_n.$$

[5]The argument, and hence the Theorem, also holds if $\widetilde{\rho}_n$ is replaced by any other $n^{1/2}$-consistent estimator for ρ.

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

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

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

From this we see that Assumptions 4 and 7 in Kelejian and Prucha (2007b) are also satisfied.

By Assumption A1 we have $\boldsymbol{\Gamma}_n$ is nonsingular for all n sufficiently large and $\lim_{n\to\infty}\boldsymbol{\Gamma}_n = \boldsymbol{\Gamma}$ is finite and nonsingular. Consequently, the $\lambda_{\min}(\boldsymbol{\Gamma}_n'\boldsymbol{\Gamma}_n) \geq \text{const} > 0$ for n sufficiently large and thus Assumption 5(a) in Kelejian and Prucha (2007b) also holds.

By Assumption A3 $\boldsymbol{\Psi}_n$ and $\boldsymbol{\Psi}_{\circ,n}$ are nonsingular for all n sufficiently large and $\lim_{n\to\infty}\boldsymbol{\Psi}_n = \boldsymbol{\Psi}$ and $\lim_{n\to\infty}\boldsymbol{\Psi}_{\circ,n} = \boldsymbol{\Psi}_\circ$ are finite and nonsingular, and thus the smallest [largest] eigenvalues of $\boldsymbol{\Psi}_n$, $\boldsymbol{\Psi}_n^{-1}$, $\boldsymbol{\Psi}_{\circ,n}$, and $\boldsymbol{\Psi}_{\circ,n}^{-1}$ are bounded away from zero [bounded from above] for sufficiently large n.

The estimator $\widehat{\rho}_n$ is a special case of the GMM estimators for ρ defined in equation (9) in Kelejian and Prucha (2007b) with $\widetilde{\Upsilon}_n = \widehat{\boldsymbol{\Psi}}_n^{-1}(\widetilde{\rho}_n)$ and $\Upsilon_n = \boldsymbol{\Psi}_n^{-1}$. As remarked above, the smallest [largest] eigenvalues of $\boldsymbol{\Psi}_n^{-1}$ bounded away from zero [bounded from above] for sufficiently large n, and thus we see from Theorem 3 in Kelejian and Prucha (2007b) that in this case Assumption 5(b),(c) in that paper are also satisfied. All other assumptions maintained by Theorems 1–3 in Kelejian and Prucha (2007b) have already been verified, which establishes $n^{1/2}$-consistency of $\widehat{\rho}_n$ and its asymptotic efficiency.

The joint limiting distribution of $n^{1/2}(\widehat{\boldsymbol{\delta}}_n - \boldsymbol{\delta})$ and $n^{1/2}(\widehat{\rho}_n - \rho)$ given by the theorem now follows immediately from Theorem 4 in Kelejian and Prucha (2007b).