Papers Reg. Sci. 82, 1–26 (2003)

Papers in
Regional
Science
© RSAI 2003

Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances

Debabrata Das[1], **Harry H. Kelejian**[2], **Ingmar R. Prucha**[2]

[1] Freddie Mac, Reston, VA 20190, USA (e-mail: debabrata_das@freddiemac.com)
[2] Department of Economics, University of Maryland, College Park, MD 20742, USA
(e-mail: {kelejian,prucha}@econ.umd.edu)

Received: 20 January 2001 / Accepted: 31 August 2001

Abstract. The article investigates the finite sample properties of estimators for spatial autoregressive models where the disturbance terms may follow a spatial autoregressive process. In particular we investigate the finite sample behavior of the feasible generalized spatial two-stage least squares (FGS2SLS) estimator introduced by Kelejian and Prucha (1998), the maximum likelihood (ML) estimator, as well as that of several other estimators. We find that the FGS2SLS estimator is virtually as efficient as the ML estimator. This is important because the ML estimator is computationally burdensome, and may even be forbidding in large samples, while the FGS2SLS estimator remains computationally feasible in large samples.

JEL classification: C0, C2

Key words: Spatial autoregressive models, ordinary least squares, two-stage least squares, maximum likelihood, finite sample distribution

1 Introduction[1]

Spatial econometric models are an important research tool in regional science, geography, and economics. For example, in recent years these models have been applied, among others, in studies of county police expenditures, local wages, per capita state and local government expenditures, technology adoption, intra-metropolitan population and employment growth, and expenditures on airports.[2] By far, the most widely used such models are variants of the one suggested in Cliff and

[1] Financial support from the National Science Foundation through grant SES-0001780 is gratefully acknowledged.

[2] See, e.g., Case (1992), Kelejian and Robinson (1992), Case, Hines, and Rosen (1993), Boarnet (1994), Bernat Jr. (1996), Bollinger and Ihlanfeldt (1997), Cohen (1998), and Buettner (1999).

Ord (1973, 1981), which is itself a variant of that considered in Whittle (1954).[3] One method of estimation of these models is maximum likelihood (ML), which is based on the normality assumption. Unfortunately, if the model has a spatially autoregressive error term, the ML procedure is computationally tedious and may even be forbidding in large samples. Denoting the sample size by n, the reason for this is that the likelihood function contains the determinant of an $n \times n$ matrix, which involves a parameter of the spatially autoregressive process determining the error term. If the model also has a spatially lagged dependent variable, the computational problems are compounded in the sense that the likelihood will contain the determinants of two $n \times n$ matrices involving parameters to be estimated. Ord (1975) suggests a computational simplification involving the calculation of the characteristic roots of these matrices. However, results given by Kelejian and Prucha (1999) suggest that even Ord's procedure may encounter computational accuracy problems because the roots themselves are not easily and accurately calculated in large samples. In some cases the computational issues associated with ML estimation are "eased" due to certain specifications, such as symmetry and sparseness of the matrices involved. See, for instance, Pace and Barry (1997) and Bell and Bockstael (2000). However, in practice sparse and/or symmetric matrix specifications cannot always be maintained and so techniques other than ML should be of interest.

Against this backdrop Kelejian and Prucha (1999) suggest a generalized moments (GM) estimator of the autoregressive parameter in a spatially autoregressive disturbance process.[4] This GM estimator remains computationally feasible in large samples regardless of whether or not the weights matrix is sparse or symmetric.[5] They then apply their general results to obtain a feasible, and again a computationally manageable, generalized least squares estimator of the regression parameters in a spatial model containing only exogenous variables. In a later study Kelejian and Prucha (1998) generalize their results to linear spatial models which contain both a spatially autoregressive disturbance term and a spatially lagged dependent variable.[6] The estimator suggested in that study is an instrumental variable estimator which accounts for spatial correlation and so was termed a feasible generalized spatial two-stage least squares (FGS2SLS) estimator. In both studies formal large sample results are given that relate to the consistency of the GM estimator as well as the consistency and asymptotic distribution of the corresponding regression parameter estimators. The established asymptotic distribution can be used to approximate the small sample distribution and to test hypotheses relating to the regression parameters. Kelejian and Prucha (1998) further demonstrate that the autoregressive

[3] Recent theoretical contributions concerning extensions and issues of inference from this class of models include Anselin and Florax (1995), Florax and Rey (1995), Anselin et al. (1996), Pace and Barry (1997), Rey and Boarnet (1998), Driscoll and Kraay (1998), Kelejian and Prucha (1998, 1999, 2000, 2001), Lee (1999a,b, 2002), Das (2000), Baltagi, Song and Koh (2000). Classic references are Anselin (1988) and Cressie (1993).

[4] To the best of our knowledge, the only other alternative to ML would be based on a procedure suggested by Ord (1975). However, this procedure is not seriously considered because of its inefficiency. See, e.g., Ord (1975, p. 122)

[5] For example, we implemented the procedure with a sample of size 100,000.

[6] Due to publication lags, Kelejian and Prucha (1998) was published before Kelejian and Prucha (1999), even though it was written at a later point in time.

disturbance parameter is a "nuisance" parameter in the sense that their FGS2SLS estimator has the same asymptotic distribution whether it is based on the true value of the autoregressive disturbance parameter or on a consistent estimator of it. Since the asymptotic distribution of the GM estimator was not determined, tests for spatial correlation must be based on other statistics, one of which is Moran's I test statistic. See, for instance, Kelejian and Prucha (2001) for general results relating to Moran's I test statistic.

As far as we know, finite sample results relating to the FGS2SLS estimator are not available, and so the purpose of this article is to give such results via Monte Carlo methods. The results we provide relate to linear spatial models containing a spatially lagged dependent variable, as well as a spatially correlated disturbance term. Among others, we consider the ML estimator, the FGS2SLS estimator and certain variations of the FGS2SLS estimator which are based on iterations.[7]

Our results are very encouraging. For instance, for all cases considered, the FGS2SLS estimator is virtually as efficient as the maximum likelihood estimator. We also note that iterations based on the FGS2SLS estimator yield only marginal improvements (in certain cases). The virtual equivalence between the ML and FGS2SLS estimator suggests that there is little "cost" in using the computationally simpler FGS2SLS rather than the more burdensome ML approach. On a somewhat more theoretical plane, in Kelejian and Prucha (1998) there is a reasonably general catalogue of assumptions under which the FGS2SLS estimator was shown to be consistent and asymptotically normal. Therefore, in practice researchers can determine whether or not their specific modelling assumptions are such that "proper" inference can be based on established results relating to FGS2SLS estimation. A corresponding scenario is not available for ML estimation, even if the model disturbances are normally distributed.

Another noteworthy result relates to a comparison of the ML estimator, which accounts for spatial correlation, and the two-stage least squares estimator, which does not account for spatial correlation. Specifically, our results suggest that in small-to-moderately large (49-400) samples, the ML estimator of a model parameter may actually be inefficient relative to the corresponding two-stage least squares (2SLS) estimator under certain conditions. This result may seem to be at odds with commonly accepted notions concerning the efficiency of ML estimation, but it is not! The reason for this is that the ML approach involves the estimation of more parameters than does the 2SLS estimator (e.g., the autoregressive parameter) and so standard arguments concerning relative efficiency do not apply.

The Monte Carlo model is specified in Sect. 2, and the particular estimators considered are described in Sect. 3. The results are discussed in Sect. 4. A summary and certain conclusions are given in Sect. 5. Technical details are relegated to appendices.

[7] Rey and Boarnet (1998) consider, via Monte Carlo methods, the small sample efficiency of estimators relating to a two equation linear spatial model containing spatially lagged dependent variables as well as systems endogenous variables. In their study they also consider efficiency issues relating to the number of instruments. However, they did not consider spatially correlated disturbance terms, nor FGS2SLS estimation.

2 The model

In our Monte Carlo study we investigate the finite sample properties of estimators of the following important spatial model:

$$y_n = \lambda W_n y_n + X_n \beta + u_n, \qquad |\lambda| < 1, \tag{1}$$

$$u_n = \rho W_n u_n + \varepsilon_n, \qquad |\rho| < 1, \tag{2}$$

where y_n is the $n \times 1$ vector of observations on the dependent variable of n spatial units, W_n is an $n \times n$ spatial weights matrix of known constants, X_n denotes the $n \times k$ matrix of non-stochastic exogenous explanatory variables, u_n is the $n \times 1$ disturbance vector, ε_n is the $n \times 1$ stochastic innovation vector, and λ and ρ are scalar parameters – typically referred to as spatial autoregressive parameters – and β is a $k \times 1$ vector of regression parameters.

This model is a generalization of the model introduced by Cliff and Ord (1973, 1981). Consistent with the terminology described in Anselin and Florax (1995, pp. 22–24) we refer to this model as a spatial autoregressive model with autoregressive disturbances of order (1,1), for short SARAR(1,1). For this model Kelejian and Prucha (1998) introduced a feasible generalized spatial 2SLS estimator, for short FGS2SLS estimator, and derived its asymptotic distribution under a set of low level regularity conditions.

The FGS2SLS estimator is an instrumental variable estimator based on an approximation of the ideal instrument $E(W_n y_n)$ for $W_n y_n$, which makes the estimator relatively simple to compute. Lee (1999a) proposed a modification of this estimator which employs an estimate of the ideal instrument and derived the asymptotic distribution of this estimator. He showed that his estimator is an asymptotically optimal instrumental variable estimator and provided an algorithm to facilitate the computation of the estimator even in large samples. We note that the FGS2SLS estimator nevertheless seems simpler to compute.

In all the Monte Carlo experiments, which are each based on 5000 repetitions, we took $\varepsilon_n \sim N(0, \sigma^2 I_n)$. Also all the specifications of the weights matrix considered, which are described below, are such that $(I - aW_n)^{-1}$ exists for all $|a| < 1$. Therefore, via (1) and (2) it follows that:

$$y_n = (I - \lambda W_n)^{-1} X_n \beta + (I - \lambda W_n)^{-1} (I - \rho W_n)^{-1} \varepsilon_n. \tag{3}$$

For future reference, we note that the log likelihood function corresponding to (3) is:

$$\begin{aligned} \ln(L) = & -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |\Omega_n| \\ & -\frac{1}{2} [y_n - (I - \lambda W_n)^{-1} X_n \beta]' \Omega_n^{-1} [y_n - (I - \lambda W_n)^{-1} X_n \beta] \end{aligned} \tag{4}$$

where $\Omega_n = \sigma^2 (I - \lambda W_n)^{-1} (I - \rho W_n)^{-1} (I - \rho W_n')^{-1} (I - \lambda W_n')^{-1}$ denotes the variance covariance matrix of y_n. We also note for future reference that, given (3) and $u_n = (I - \rho W_n)^{-1} \varepsilon_n$,

$$E[(W_n y_n) u_n'] = \sigma^2 W_n (I - \lambda W_n)^{-1} (I - \rho W_n)^{-1} (I - \rho W_n')^{-1} \neq 0. \tag{5}$$

In light of (5) it is not difficult to show that the least squares estimator of the parameters in (1) is inconsistent for the spatial weights matrices considered in our Monte Carlo study, assuming that regularity conditions such as those in Kelejian and Prucha (1998) hold.[8]

We now describe the experiments in more detail. We consider three values of the sample size n, namely 49, 100, and 400. For each value of the sample size we consider three specifications of the weights matrix which essentially differ in their degree of sparseness. The first is a matrix which relates each element of y_n and u_n, see (1) and (2), to the one immediately before it and to the one immediately after it. Specifically, in this matrix the i-th row has non-zero elements only in positions $i-1$ and $i+1$, for $i=2,\ldots,n-1$. We consider a circular world and so the non-zero elements in the first row are in positions 2 and n, and those in the last row are in positions 1 and $n-1$. Furthermore, the matrix is specified such that all non-zero elements are equal and the respective rows sum to unity. Thus, for this matrix all non-zero elements are equal to $1/2$. For future reference, we will henceforth refer to this matrix as "one ahead and one behind". The second and third matrices are defined in an analogous manner as "three ahead and three behind" and "five ahead and five behind," again in a circular world. The non-zero elements in these matrices are, respectively, $1/6$ and $1/10$. The average number of neighbors per unit, henceforth J, associated with these three matrices are, obviously, $J=2$, 6, and 10. These matrices should be able to provide guidance for many applied situations.[9]

We consider nine values for λ and nine values for ρ, namely -0.8, -0.6, -0.4, -0.2, 0.0, 0.2, 0.4, 0.6, and 0.8. Three values are considered for σ^2, which are 0.25, 0.50, and 1.0. These values of σ^2 correspond to average R^2 values over the experiments considered of, roughly, 0.7, 0.8, and 0.9, where these R^2 values relate to the squared correlation coefficient between y_n and the model explained mean vector, namely $E(y_n)=(I-\lambda W_n)^{-1}X_n\beta$.

Finally, our Monte Carlo model is specified in terms of two regressors, i.e., $X_n = (x_{1n}, x_{2n})$. Values for the $n \times 1$ regression vectors x_{1n} and x_{2n} are based on data given in Kelejian and Robinson (1992) on income per capita in 1982 and on the percent of rental housing in 1980 in 760 counties in the U.S. mid-western states. Specifically, the 760 observations on the income and rental variables are normalized so that their sample means and sample variances are, respectively, zero and one. The values of x_{1n} and x_{2n} are then taken as the first n values of, respectively, the normalized income and rental variables for the considered sample sizes of $n=49$, 100, and 400. The same vectors x_{1n} and x_{2n} are used in all experiments in which the

[8] Note that in general (5) holds even if the disturbance process is i.i.d., i.e., $\rho=0$. In an interesting paper, Lee (2002) considers a spatial autoregressive model of the form (1), but where the disturbance process is assumed to be i.i.d. He shows that for a certain class of weights matrices the least squares estimator of its parameters is consistent. One important characteristic of this class of weights matrices is that all non-zero weights tend to zero as the sample size tends to infinity. A further analysis of some of the issues considered in Lee (2002) is given in Kelejian and Prucha (2000).

[9] These matrices were also considered in Kelejian and Prucha (1999) together with several "real world" matrices. It is important to note that the results obtained in that study from the stylized matrices and from the "real world" matrices were very similar.

Table 1. Design values of σ^2 in the 729 experiments

$n = 49$		$n = 100$		$n = 400$	
λ	σ^2	λ	σ^2	λ	σ^2
−0.8	1.0	−0.8	0.25	−0.8	0.5
−0.6	0.5	−0.6	0.5	−0.6	0.25
−0.4	0.25	−0.4	1.0	−0.4	1.0
−0.2	1.0	−0.2	0.25	−0.2	0.5
0.0	0.5	0.0	0.5	0.0	0.25
0.2	0.25	0.2	1.0	0.2	1.0
0.4	1.0	0.4	0.25	0.4	0.5
0.6	0.5	0.6	0.5	0.6	0.25
0.8	0.25	0.8	1.0	0.8	1.0

Table 2. Experiments in which either λ or ρ equal 0.9 in absolute value

		$J = 2$		$J = 6$		$J = 10$	
λ	ρ	n	σ^2	n	σ^2	n	σ^2
−0.9	0.0	49	0.5	400	1.0	49	1.0
0.0	−0.9	400	0.5	49	1.0	196	1.0
0.0	0.9	400	0.5	49	1.0	196	1.0
0.9	0.0	49	0.5	400	1.0	49	1.0
−0.9	−0.7	49	0.25	400	0.5	49	0.25
0.9	0.7	49	0.25	400	0.5	100	0.25
−0.7	0.9	100	0.5	196	1.0	100	0.5
0.7	−0.9	100	0.5	196	1.0	49	0.5

sample size is n. The elements of the parameter vector $\beta = (\beta_1, \beta_2)$ are specified as $\beta_1 = \beta_2 = 1$.

As a review we consider nine values of λ, nine values of ρ, three values of n, three values of σ^2, and three values of J. Instead of considering all combinations of λ, ρ, n, σ^2 and J, which would have led to 2187 experiments, we only consider all combinations of λ, ρ, n, and J, which result in 729 experiments. The three values of σ^2 are related to the values of λ and n, and hence are woven into the 729 experiments. These values of σ^2 are described in Table 1.

In addition to the experiments described above we also performed 24 experiments in which either λ or ρ is 0.9 in absolute value. These additional experiments were considered because preliminary results suggest a certain sensitivity of the root mean squared errors of the considered estimators to the more extreme values of ρ and λ. These 24 experiments are described in Table 2.

Note from Table 2 that the 24 experimental values of λ and ρ are orthogonal to each other, as well as to those of n, J, and σ^2. Since the values of n, J, and σ^2 are positive they, of course, are not orthogonal. However, the R^2 statistics relating the 24 experimental values of n to those of J, of n to those of σ^2, and of J to those of σ^2, are, respectively, 0.021, 0.026, and 0.12. The experimental values of these parameters are therefore virtually uncorrelated. As described below, we present our results in terms of tables as well as response functions relating a measure of the root mean squared error of a considered estimator to the parameters of the Monte Carlo model. Our design of the experiments should facilitate the estimation of these response functions.

3 The estimators considered

Denote the regressor matrix in (1) as $Z_n = (W_n y_n, X_n)$, and the corresponding parameter vector as $\delta = (\lambda, \beta')'$ so that (1) and (2) can be written as:

$$y_n = Z_n\delta + u_n, \quad u_n = \rho W_n u_n + \varepsilon_n. \tag{6}$$

Let:

$$y_n(a) = y_n - aW_n y_n, \quad Z_n(a) = Z_n - aW_n Z_n,$$

where a is any finite scalar. Applying a Cochrane-Orcutt type transformation to (6) then yields:

$$y_n(\rho) = Z_n(\rho)\delta + \varepsilon_n.$$

Consistent with suggestions in Kelejian and Prucha (1998), let the independent columns of $H_n = (X_n, W_n X_n, W_n^2 X_n)$ be the set of instruments used in conjunction with their FGS2SLS procedure. Note that our model does not contain an intercept term and so the six columns of H_n are linearly independent. Also note that if $W_n^2 X_n$ is computed recursively as $W_n(W_n X_n)$, its computational count is $O(n^2)$; on the other hand, if W_n^2 is first computed and then postmultiplied by X_n, the computational count would be $O(n^3)$. We therefore recommend the former calculation procedure. As a point of comparison we also indicate that the computation of the eigenvalues of W_n involved in the computation of the ML estimator has an operational count of $O(n^3)$.

The ideal instruments are given by $EZ_n(\rho) = (W_n Ey_n - \rho W_n^2 Ey_n, X_n - \rho W_n X_n)$ and so are determined by $E(y_n)$ and X_n, given the autoregressive parameter ρ and the weights matrix W_n. Observe that for the model at hand:

$$Ey_n = (I - \lambda W_n)^{-1} X_n\beta = \sum_{i=0}^{\infty} \lambda^i W_n^i X_n\beta,$$

and hence, these ideal instruments are linear combinations of the matrices X_n, $W_n X_n, \ldots$. Our choice of H_n is motivated by this observation, i.e., H_n contains the first three matrices in this sequence.[10]

[10] In a related study which involved the estimation of a regression model containing a spatially lagged dependent variable as well as a systems endogenous variable, but a non-spatially correlated error term, Rey and Boarnet (1998) also considered the instrument matrix H_n. We note that the discussion above, as well as that in Kelejian and Prucha (1998), suggests that it may be reasonable to include the product of higher powers of the spatial weights matrix and X_n in H_n.

Based on H_n, the instrument matrices used in the first and third steps of the FGS2SLS procedure are:

$$\hat{Z}_n = P_{H_n} Z_n, \quad \hat{Z}_n(\widehat{\rho}) = P_{H_n} Z_n(\widehat{\rho}),$$

where $P_{H_n} = H_n(H_n' H_n)^{-1} H_n'$ and $\widehat{\rho}$ is an estimator for ρ. Let $\hat{\rho}_{GM}$ and $\hat{\rho}_{GM}^{(i)}$ be, respectively, the nonlinear generalized moments (GM) estimator of ρ suggested in Kelejian and Prucha (1998, 1999), and the nonlinear GM estimator of ρ based upon the i^{th} iteration of the FGS2SLS procedure.[11] Let $\hat{\rho}_{ML}$ be the ML estimator of ρ.

Then in this study the considered estimators of δ in (6) are:

- Maximum Likelihood based on (4): $\hat{\delta}_{ML}$
- Least squares:
$$\hat{\delta}_{OLS} = (Z_n' Z_n)^{-1} Z_n' y_n.$$
- Two-stage least squares:
$$\hat{\delta}_{2SLS} = (\hat{Z}_n' \hat{Z}_n)^{-1} \hat{Z}_n' y_n.$$
- GS2SLS based on the true value of ρ :
$$\hat{\delta}_{GS2SLS} = (\hat{Z}_n(\rho)' \hat{Z}_n(\rho))^{-1} \hat{Z}_n'(\rho) y_n(\rho).$$
- FGS2SLS based on $\hat{\rho}_{GM}$:
$$\hat{\delta}_{FGS2SLS} = (\hat{Z}_n(\hat{\rho}_{GM})' \hat{Z}_n(\hat{\rho}_{GM}))^{-1} \hat{Z}_n'(\hat{\rho}_{GM}) y_n(\hat{\rho}_{GM}).$$
- Iterated FGS2SLS based on $\hat{\rho}_{GM}^{(i)}$, $i = 1, ..., 5$:
$$\hat{\delta}_{IFi} = (\hat{Z}_n(\hat{\rho}_{GM}^{(i)})' \hat{Z}_n(\hat{\rho}_{GM}^{(i)}))^{-1} \hat{Z}_n'(\hat{\rho}_{GM}^{(i)}) y_n(\hat{\rho}_{GM}^{(i)}).$$

In passing we note that the FGS2SLS estimator of $\delta' = (\lambda, \beta')$ described above does not impose the restriction $|\lambda| < 1$. A nonlinear variant of this estimator which does impose this restriction is discussed in the Appendix B.

4 Monte Carlo results

Due to space limitations, we present our results in two ways. We first give tables of the root mean squared errors ($RMSE$s) of all respective estimators considered for a subset of our experimental parameter values. Afterwards, we give response functions for the $RMSE$s of the ML and FGS2SLS estimators, which are estimated in terms of the entire set of $(729+24 = 753)$ Monte Carlo experiments, and present graphs of those response functions.

[11] These estimators are outlined in Appendix A.

4.1 Root mean squared errors in tabular form

To conserve space we characterize the bias and spread of the finite sample distribution of the respective estimators in terms of a single measure. This measure is closely related to the standard $RMSE$, but unlike that standard measure the components of our measure relating to the bias and spread of our considered estimators are assured to exist. Details about this are described in Appendix C.

Our measure is defined as $[bias^2 + (IQ/1.35)^2]^{1/2}$ where $bias$ is the difference between the median, or 0.5 quantile, and the true parameter value, and IQ is the inter-quantile range. That is, $IQ = c_1 - c_2$ where c_1 is the 0.75 quantile and c_2 is the 0.25 quantile. If the distribution is normal, the median is equal to the mean and $IQ/1.35$, except for a rounding error, is equal to the standard deviation; thus our measure coincides with the standard $RMSE$ measure in this case. In discussing the tables we refer to our measure simply as the $RMSE$. The results in the tables are Monte Carlo estimates of $[bias^2 + (IQ/1.35)^2]^{1/2}$ based on quantiles computed from the empirical distributions corresponding to the 5000 Monte Carlo replications.

In total we performed 753 Monte Carlo experiments. To conserve space we only report $RMSE$s for a subset of those experiments here. As a point of interest the biases of all our considered estimators, with the exception of OLS, which is inconsistent, were "quite small", typically less than 0.01 in absolute value. As a point of comparison, we also calculated the $RMSE$s in the standard way and found the results to be quite similar.

In Tables 3–10 we give results on the $RMSE$ of the considered estimators of the parameters λ, β_1, β_2, and ρ corresponding to 50 experimental parameter values. The set of 50 is based on all combinations of five values of λ, five values of ρ and two values of n. The values of σ^2 are woven into the 25 combinations of λ and ρ. In all tables the value of $J = 6$. The first table corresponding to each parameter contains Monte Carlo estimates of the $RMSE$s when $n = 100$; the second table relates to cases in which $n = 400$.

First note from the tables that results corresponding to iterations on the FGS2SLS estimators for λ, β_1, and β_2 and for the GM estimator for ρ beyond the first iteration are not reported. The reason is that these estimators do not have consistently lower $RMSE$s than the estimators which only involve one, or no iterations. As seen from the tables, even the first iteration itself often does not lead to lower $RMSE$s.

Focusing first on the regression coefficients λ, β_1, and β_2 we see from the tables that the $RMSE$s of the OLS estimators are typically largest, while those of the ML estimators are typically the lowest. This is consistent with theoretical notions since the OLS estimator is not consistent, while the ML estimator is consistent and efficient (assuming standard maximum likelihood theory applies for the model under investigation). Note also that the $RMSE$s of the 2SLS estimators, while typically lower than those of OLS estimators, are typically larger than those of FGS2SLS estimators. Again, this accords with the theoretical notions that, although both are consistent, the 2SLS estimators do not account for the spatial correlation, while the FGS2SLS estimators do. In comparing the ML and FGS2SLS estimators

Table 3. Root mean square errors of the estimators of λ, $N = 100$, $J = 6$

λ	ρ	σ^2	**OLS**	**ML**	**2SLS**	**GS2SLS**	**FGS2SLS**	**IF1**
−0.8	−0.8	0.25	0.181	0.066	0.076	0.065	0.066	0.067
−0.8	−0.4	0.25	0.127	0.068	0.070	0.067	0.068	0.068
−0.8	0	0.25	0.095	0.075	0.075	0.075	0.076	0.076
−0.8	0.4	0.25	0.100	0.091	0.101	0.091	0.093	0.093
−0.8	0.8	0.25	0.394	0.110	0.238	0.115	0.126	0.121
−0.4	−0.8	0.50	0.242	0.082	0.091	0.081	0.083	0.083
−0.4	−0.4	0.50	0.162	0.086	0.089	0.085	0.086	0.086
−0.4	0	0.50	0.113	0.097	0.098	0.098	0.099	0.100
−0.4	0.4	0.50	0.152	0.127	0.135	0.128	0.129	0.130
−0.4	0.8	0.50	0.610	0.168	0.315	0.185	0.210	0.197
0	−0.8	1.00	0.258	0.091	0.098	0.091	0.092	0.091
0	−0.4	1.00	0.161	0.096	0.097	0.097	0.097	0.097
0	0	1.00	0.115	0.113	0.113	0.113	0.114	0.115
0	0.4	1.00	0.222	0.153	0.160	0.153	0.155	0.156
0	0.8	1.00	0.684	0.224	0.346	0.264	0.298	0.284
0.4	−0.8	0.25	0.040	0.030	0.031	0.030	0.030	0.030
0.4	−0.4	0.25	0.034	0.032	0.033	0.032	0.032	0.033
0.4	0	0.25	0.040	0.038	0.038	0.038	0.039	0.039
0.4	0.4	0.25	0.070	0.056	0.057	0.055	0.056	0.056
0.4	0.8	0.25	0.248	0.113	0.145	0.112	0.116	0.113
0.8	−0.8	0.50	0.016	0.015	0.016	0.015	0.015	0.015
0.8	−0.4	0.50	0.017	0.017	0.017	0.017	0.017	0.017
0.8	0	0.50	0.023	0.021	0.021	0.021	0.021	0.021
0.8	0.4	0.50	0.043	0.032	0.031	0.031	0.032	0.032
0.8	0.8	0.50	0.133	0.079	0.075	0.071	0.070	0.072
Column average			0.171	0.083	0.103	0.085	0.089	0.088

we note that the ML estimator for λ, is on average, roughly seven percent more efficient that the FGS2SLS estimator. The $RMSE$s of the ML and FGS2SLS estimators of β_1 and β_2, conversely, are on average roughly the same. As a general observation it seems that for the regression coefficients the loss of efficiency of the FGS2SLS estimator relative to the ML estimator is fairly small. This is important because of the computational advantages of the FGS2SLS estimator, especially in large samples. A related observation is that the instrument matrix H_n seems to yield a reasonable "approximation" to the set of optimal instruments, so any efficiency gain based on use of those optimal instruments should be limited.

The results in the tables also indicate that the relative size of the $RMSE$s of the ML and FGS2SLS estimators depends on the true values of ρ and λ. In particular, in the tables corresponding to λ, this difference is largest when λ is negative and large in absolute value, and ρ is large and positive. When λ and ρ are positive and large the difference is actually negative in that the FGS2SLS estimator has the lower $RMSE$s. Interestingly, in this case the $RMSE$s of the 2SLS estimator are also

Table 4. Root mean square errors of the estimators of λ, $N = 400$, $J = 6$

λ	ρ	σ^2	**OLS**	**ML**	**2SLS**	**GS2SLS**	**FGS2SLS**	**IF1**
−0.8	−0.8	0.50	0.398	0.055	0.064	0.055	0.056	0.056
−0.8	−0.4	0.50	0.266	0.057	0.059	0.057	0.057	0.057
−0.8	0	0.50	0.150	0.062	0.062	0.062	0.062	0.062
−0.8	0.4	0.50	0.090	0.072	0.080	0.074	0.075	0.075
−0.8	0.8	0.50	0.711	0.081	0.196	0.098	0.104	0.100
−0.4	−0.8	1.00	0.511	0.067	0.078	0.068	0.068	0.068
−0.4	−0.4	1.00	0.318	0.072	0.074	0.071	0.071	0.072
−0.4	0	1.00	0.140	0.080	0.080	0.080	0.080	0.080
−0.4	0.4	1.00	0.207	0.099	0.109	0.103	0.102	0.103
−0.4	0.8	1.00	0.918	0.116	0.266	0.154	0.171	0.163
0	−0.8	0.25	0.093	0.026	0.030	0.026	0.026	0.026
0	−0.4	0.25	0.054	0.028	0.029	0.028	0.028	0.028
0	0	0.25	0.033	0.033	0.033	0.033	0.033	0.033
0	0.4	0.25	0.089	0.044	0.047	0.044	0.045	0.045
0	0.8	0.25	0.419	0.073	0.120	0.077	0.081	0.079
0.4	−0.8	0.50	0.064	0.023	0.026	0.024	0.024	0.024
0.4	−0.4	0.50	0.032	0.026	0.027	0.026	0.026	0.026
0.4	0	0.50	0.041	0.031	0.031	0.031	0.031	0.031
0.4	0.4	0.50	0.123	0.044	0.045	0.044	0.044	0.044
0.4	0.8	0.50	0.401	0.085	0.116	0.091	0.094	0.092
0.8	−0.8	1.00	0.014	0.011	0.013	0.012	0.012	0.012
0.8	−0.4	1.00	0.016	0.013	0.014	0.013	0.013	0.013
0.8	0	1.00	0.035	0.016	0.016	0.016	0.016	0.016
0.8	0.4	1.00	0.079	0.024	0.024	0.024	0.024	0.024
0.8	0.8	1.00	0.184	0.063	0.060	0.058	0.056	0.059
Column average			0.215	0.052	0.068	0.055	0.056	0.056

lower than those of the ML estimator.[12] Perhaps the reason for this is that the ML estimator of λ requires the estimation of ρ, whereas the 2SLS estimator does not. That is, the ML procedure estimates more parameters than the 2SLS procedure so the standard results on relative ML efficiency do not apply here.

$RMSE$ differences between FGS2SLS and GS2SLS (based on the true value of ρ) are quite small. For example, these differences averaged over all six tables corresponding to λ, β_1, and β_2 are only 1.6%. If Table 3 is excluded, these differences average to only 0.7 percent. These results suggest that if the sample is at least moderately sized, the loss in finite sample efficiency, as measured by $RMSE$, due to use of the GM estimator of ρ as compared to the true value of ρ in the Kelejian and Prucha (1998) instrumental variable procedure, is "slight".

Tables 9 and 10 relate to the estimators of ρ. The $RMSE$s of the ML estimators are generally somewhat lower than those of the GM estimators. The difference between the $RMSE$s of the ML and GM estimators of ρ is, averaged over Tables

[12] For further results along these lines when $\rho = \lambda = 0.9$ (See Das 2000).

Table 5. Root mean square errors of the estimators of β_1, $N = 100$, $J = 6$

λ	ρ	σ^2	**OLS**	**ML**	**2SLS**	**GS2SLS**	**FGS2SLS**	**IF1**
−0.8	−0.8	0.25	0.044	0.043	0.047	0.042	0.043	0.043
−0.8	−0.4	0.25	0.042	0.043	0.043	0.042	0.043	0.043
−0.8	0	0.25	0.042	0.043	0.043	0.043	0.043	0.043
−0.8	0.4	0.25	0.045	0.044	0.045	0.044	0.044	0.044
−0.8	0.8	0.25	0.066	0.048	0.074	0.049	0.050	0.049
−0.4	−0.8	0.50	0.065	0.063	0.068	0.062	0.063	0.063
−0.4	−0.4	0.50	0.061	0.062	0.062	0.061	0.061	0.061
−0.4	0	0.50	0.060	0.060	0.060	0.060	0.060	0.060
−0.4	0.4	0.50	0.063	0.061	0.062	0.061	0.061	0.061
−0.4	0.8	0.50	0.090	0.066	0.093	0.067	0.069	0.068
0	−0.8	1.00	0.100	0.090	0.099	0.089	0.091	0.091
0	−0.4	1.00	0.090	0.089	0.090	0.088	0.089	0.090
0	0	1.00	0.086	0.086	0.085	0.085	0.086	0.087
0	0.4	1.00	0.093	0.085	0.088	0.085	0.086	0.085
0	0.8	1.00	0.139	0.088	0.120	0.091	0.091	0.091
0.4	−0.8	0.25	0.051	0.045	0.051	0.045	0.046	0.046
0.4	−0.4	0.25	0.046	0.045	0.046	0.045	0.045	0.045
0.4	0	0.25	0.044	0.044	0.043	0.043	0.044	0.044
0.4	0.4	0.25	0.047	0.042	0.044	0.042	0.042	0.042
0.4	0.8	0.25	0.092	0.043	0.056	0.043	0.043	0.044
0.8	−0.8	0.50	0.072	0.064	0.072	0.063	0.064	0.065
0.8	−0.4	0.50	0.065	0.064	0.065	0.064	0.064	0.064
0.8	0	0.50	0.064	0.063	0.062	0.062	0.064	0.064
0.8	0.4	0.50	0.071	0.060	0.063	0.060	0.061	0.061
0.8	0.8	0.50	0.142	0.059	0.081	0.058	0.059	0.059
Column average			0.071	0.060	0.066	0.060	0.060	0.061

9 and 10, roughly 7%. Returns to iteration on the GM estimator seem marginal. Finally, consistent with prior notions, the $RMSE$s in all the tables corresponding to λ, β_1, β_2 and ρ generally decrease as the sample size increases.

4.2 Root mean squared error response functions

The relationship between the $RMSE$s of the considered estimators and the model parameters cannot easily be determined from Tables 3–10, so we now describe those results in terms of response functions. To conserve space we only consider the ML estimators and the FGS2SLS estimators of λ, β_1, and β_2 and the ML estimator and the GM estimator of ρ. In doing so we consider separate functions for each estimator of each parameter; in total this leads to eight response functions.

Let λ_i, β_{1i}, β_{2i}, ρ_i, J_i, and σ_i^2 be the values, respectively, of λ, β_1, β_2, ρ, J, and σ^2 in the i-th experiment, $i = 1, \ldots, 753$. Let n_i be the corresponding value of the sample size. Given this notation, the eight response functions we consider

Table 6. Root mean square errors of the estimators of β_1, $N = 400$, $J = 6$

λ	ρ	σ^2	**OLS**	**ML**	**2SLS**	**GS2SLS**	**FGS2SLS**	**IF1**
−0.8	−0.8	0.50	0.045	0.036	0.040	0.036	0.036	0.036
−0.8	−0.4	0.50	0.040	0.036	0.037	0.035	0.036	0.036
−0.8	0	0.50	0.038	0.037	0.037	0.037	0.037	0.037
−0.8	0.4	0.50	0.043	0.040	0.042	0.040	0.040	0.040
−0.8	0.8	0.50	0.081	0.043	0.083	0.045	0.047	0.046
−0.4	−0.8	1.00	0.090	0.052	0.058	0.052	0.052	0.052
−0.4	−0.4	1.00	0.068	0.053	0.055	0.053	0.053	0.053
−0.4	0	1.00	0.055	0.053	0.053	0.053	0.053	0.053
−0.4	0.4	1.00	0.065	0.056	0.059	0.055	0.055	0.056
−0.4	0.8	1.00	0.151	0.059	0.106	0.063	0.064	0.064
0	−0.8	0.25	0.037	0.027	0.030	0.027	0.027	0.027
0	−0.4	0.25	0.030	0.027	0.028	0.027	0.027	0.027
0	0	0.25	0.027	0.027	0.027	0.027	0.027	0.027
0	0.4	0.25	0.036	0.028	0.029	0.028	0.028	0.028
0	0.8	0.25	0.115	0.030	0.050	0.030	0.030	0.030
0.4	−0.8	0.50	0.052	0.038	0.044	0.039	0.038	0.039
0.4	−0.4	0.50	0.041	0.039	0.041	0.039	0.039	0.039
0.4	0	0.50	0.042	0.039	0.039	0.039	0.039	0.039
0.4	0.4	0.50	0.071	0.039	0.042	0.039	0.039	0.039
0.4	0.8	0.50	0.198	0.040	0.066	0.040	0.040	0.040
0.8	−0.8	1.00	0.063	0.053	0.064	0.054	0.055	0.055
0.8	−0.4	1.00	0.061	0.056	0.059	0.056	0.056	0.056
0.8	0	1.00	0.074	0.057	0.057	0.057	0.057	0.057
0.8	0.4	1.00	0.128	0.056	0.059	0.056	0.056	0.056
0.8	0.8	1.00	0.274	0.054	0.092	0.055	0.055	0.054
Column average			0.077	0.043	0.052	0.043	0.043	0.043

all have the form:

$$RMSE_i = \frac{\sigma_i^{d_i}}{\sqrt{n_i}} \exp(a_1 + a_2(1/J_i) + a_3(\rho_i/J_i) + a_4(\lambda_i/J_i) \\ + a_5\rho_i + a_6\rho_i^2 + a_7\lambda_i + a_8\lambda_i^2 + a_9(\lambda_i\rho_i) \\ + a_{10}(J_i/n_i) + a_{11}(J_i/n_i)^2 + a_{12}(1/n_i) + a_{13}\sigma_i^2), \\ i = 1, ..., 753, \tag{7}$$

where $RMSE_i$ is the $RMSE$ of an estimator of a given parameter in the i-th experiment, $d_i = 1$ if $RMSE_i$ relates to an estimator of λ, β_1, or β_2 and $d_i = 0$ otherwise (when $RMSE_i$ relates to ρ), and where the parameters $a_1, ..., a_{13}$ are different for each estimator of each parameter. For each case considered, the parameters of (7) are estimated by least squares after taking logs with $\ln(n_i^{1/2}$ $RMSE_i/\ \sigma_i^{d_i})$ as the dependent variable.

Table 7. Root mean square errors of the estimators of β_2, $N = 100$, $J = 6$

λ	ρ	σ^2	**OLS**	**ML**	**2SLS**	**GS2SLS**	**FGS2SLS**	**IF1**
−0.8	−0.8	0.25	0.058	0.055	0.060	0.055	0.055	0.055
−0.8	−0.4	0.25	0.053	0.054	0.054	0.054	0.054	0.054
−0.8	0	0.25	0.053	0.053	0.053	0.053	0.054	0.054
−0.8	0.4	0.25	0.058	0.053	0.058	0.054	0.053	0.053
−0.8	0.8	0.25	0.087	0.057	0.102	0.057	0.059	0.058
−0.4	−0.8	0.50	0.081	0.079	0.087	0.080	0.080	0.080
−0.4	−0.4	0.50	0.076	0.077	0.077	0.077	0.078	0.078
−0.4	0	0.50	0.074	0.075	0.075	0.075	0.075	0.075
−0.4	0.4	0.50	0.080	0.074	0.080	0.073	0.074	0.074
−0.4	0.8	0.50	0.097	0.079	0.126	0.081	0.083	0.082
0	−0.8	1.00	0.121	0.114	0.126	0.114	0.114	0.115
0	−0.4	1.00	0.111	0.111	0.112	0.111	0.112	0.112
0	0	1.00	0.107	0.108	0.107	0.107	0.108	0.108
0	0.4	1.00	0.110	0.104	0.110	0.101	0.104	0.103
0	0.8	1.00	0.124	0.107	0.152	0.110	0.110	0.111
0.4	−0.8	0.25	0.065	0.058	0.065	0.058	0.059	0.059
0.4	−0.4	0.25	0.058	0.057	0.058	0.057	0.057	0.057
0.4	0	0.25	0.054	0.054	0.054	0.054	0.055	0.055
0.4	0.4	0.25	0.056	0.051	0.054	0.051	0.051	0.051
0.4	0.8	0.25	0.091	0.052	0.074	0.051	0.051	0.052
0.8	−0.8	0.50	0.092	0.083	0.093	0.084	0.084	0.084
0.8	−0.4	0.50	0.083	0.081	0.083	0.081	0.081	0.081
0.8	0	0.50	0.079	0.078	0.078	0.078	0.079	0.079
0.8	0.4	0.50	0.084	0.073	0.077	0.073	0.074	0.074
0.8	0.8	0.50	0.150	0.070	0.102	0.070	0.070	0.070
Column average			0.084	0.074	0.087	0.074	0.075	0.075

A few points concerning (7) should be noted. First, the asymptotic variance covariance matrix of the estimators of λ, β_1, and β_2 being considered is proportional to σ^2/n. Corresponding asymptotic standard deviations are therefore proportional to $\sigma/n^{1/2}$. The response functions for the $RMSE$s of λ, β_1, and β_2 in (7) reflect this proportionality. As suggested by (7), we do not expect[13] the variance of the two considered estimators of ρ to be directly proportional to σ^2. One reason for this is that ρ only appears in (2) and σ is just a scale factor in (2). Because of this we have specified the response functions for the $RMSE$s for the estimators for ρ only as proportional to $n^{-1/2}$. Nevertheless, we allow for some possible forms of dependence on σ of the $RMSE$s of the estimators of ρ by including σ in the exponent in (7).

[13] Recall that Kelejian and Prucha (1999) demonstrated the consistency of the GM estimator of ρ. The large sample distribution of that estimator was not determined.

Table 8. Root mean square errors of the estimators of β_2, $N = 400$, $J = 6$

λ	ρ	σ^2	**OLS**	**ML**	**2SLS**	**GS2SLS**	**FGS2SLS**	**IF1**
−0.8	−0.8	0.50	0.049	0.037	0.042	0.037	0.037	0.037
−0.8	−0.4	0.50	0.043	0.038	0.039	0.037	0.037	0.037
−0.8	0	0.50	0.041	0.039	0.039	0.039	0.039	0.039
−0.8	0.4	0.50	0.043	0.040	0.044	0.040	0.041	0.041
−0.8	0.8	0.50	0.074	0.042	0.085	0.043	0.044	0.044
−0.4	−0.8	1.00	0.056	0.054	0.060	0.054	0.054	0.054
−0.4	−0.4	1.00	0.055	0.053	0.055	0.054	0.053	0.054
−0.4	0	1.00	0.055	0.055	0.055	0.055	0.055	0.055
−0.4	0.4	1.00	0.058	0.056	0.060	0.056	0.056	0.056
−0.4	0.8	1.00	0.064	0.057	0.109	0.060	0.062	0.062
0	−0.8	0.25	0.030	0.027	0.031	0.027	0.027	0.027
0	−0.4	0.25	0.028	0.027	0.028	0.027	0.027	0.027
0	0	0.25	0.027	0.027	0.027	0.027	0.027	0.027
0	0.4	0.25	0.029	0.027	0.030	0.027	0.027	0.027
0	0.8	0.25	0.043	0.029	0.051	0.029	0.030	0.029
0.4	−0.8	0.50	0.044	0.039	0.044	0.039	0.040	0.040
0.4	−0.4	0.50	0.040	0.039	0.040	0.039	0.039	0.039
0.4	0	0.50	0.040	0.039	0.039	0.039	0.039	0.039
0.4	0.4	0.50	0.045	0.038	0.042	0.038	0.038	0.038
0.4	0.8	0.50	0.079	0.039	0.068	0.039	0.039	0.040
0.8	−0.8	1.00	0.063	0.056	0.064	0.057	0.057	0.057
0.8	−0.4	1.00	0.058	0.056	0.058	0.056	0.056	0.056
0.8	0	1.00	0.060	0.055	0.055	0.055	0.056	0.056
0.8	0.4	1.00	0.076	0.054	0.059	0.054	0.055	0.055
0.8	0.8	1.00	0.129	0.053	0.090	0.054	0.054	0.054
Column average			0.053	0.043	0.053	0.043	0.043	0.043

Second, the function in (7) is relatively simple, and yet is non-negative and able to account for certain patterns suggested by time series considerations. For example, consider the time series model:

$$y_t = \lambda y_{t-1} + x_t\beta + u_t,\ |\lambda| < 1, \qquad (8)$$
$$u_t = \rho u_{t-1} + \varepsilon_t,\ |\rho| < 1,$$

where x_t is a scalar exogenous variable and ε_t is $i.i.d.\ N(0, \sigma^2)$. The variance-covariance matrix of the asymptotic distribution of the ML estimators of λ and β is, under typical assumptions, a function of $(1-\rho^2)$, $(1-\lambda^2)$, and $(1-\lambda\rho)$.[14] Although the spatial model we consider is not identical to the model in (8) we might nevertheless expect the exponent in (7) to be nonlinear in ρ and λ, and to involve an interaction between them. The squares and cross product terms in (7) are obvious choices. Interactions between ρ and J, and λ and J are also considered

[14] See Dhrymes (1981, pp. 199–200).

Table 9. Root mean square errors of the estimators of ρ, $N = 100$, $J = 6$

λ	ρ	σ^2	**ML**	**GM**	**GM1**
−0.8	−0.8	0.25	0.229	0.240	0.251
−0.8	−0.4	0.25	0.229	0.225	0.239
−0.8	0	0.25	0.202	0.198	0.208
−0.8	0.4	0.25	0.153	0.156	0.155
−0.8	0.8	0.25	0.067	0.102	0.076
−0.4	−0.8	0.50	0.234	0.247	0.258
−0.4	−0.4	0.50	0.234	0.233	0.247
−0.4	0	0.50	0.213	0.208	0.218
−0.4	0.4	0.50	0.166	0.171	0.166
−0.4	0.8	0.50	0.076	0.130	0.096
0	−0.8	1.00	0.241	0.259	0.271
0	−0.4	1.00	0.244	0.248	0.261
0	0	1.00	0.227	0.228	0.235
0	0.4	1.00	0.183	0.190	0.185
0	0.8	1.00	0.100	0.179	0.143
0.4	−0.8	0.25	0.221	0.237	0.250
0.4	−0.4	0.25	0.223	0.227	0.237
0.4	0	0.25	0.200	0.200	0.206
0.4	0.4	0.25	0.156	0.157	0.158
0.4	0.8	0.25	0.086	0.105	0.089
0.8	−0.8	0.50	0.220	0.238	0.249
0.8	−0.4	0.50	0.224	0.230	0.239
0.8	0	0.50	0.202	0.204	0.209
0.8	0.4	0.50	0.157	0.161	0.161
0.8	0.8	0.50	0.102	0.113	0.098
Column average			0.184	0.195	0.196

because both ρ and λ multiply the weights matrix which contains J, see (1) and (2). The reciprocal form of J was considered because the considered weights matrices involve the reciprocal of J. Finally, the innovation variance, σ^2 is considered for obvious reasons, e.g., ceteris paribus, the larger is σ^2, the more noise the model contains and so the larger should be the $RMSE$s of most estimators.

A final point should be noted before turning to the empirical results is that the response functions were estimated in terms of results based on 5000 iterations for each of 753 Monte Carlo experiments, yielding a total of $5000 * 753 = 3,765,000$ trials. In a manner similar to typical practice in time series analysis, the restrictions $|\rho| < 1$ and $|\lambda| < 1$ were not imposed in any of these trials. The number of trials in which our estimates violated one or both of these parameter space restrictions was approximately 1,500. This is roughly $.04\%$ of the total number of trials and so we expect that our results would be approximately the same had we imposed these restrictions as described in Appendix B.

Table 10. Root mean square errors of the estimators of ρ, $N = 400$, $J = 6$

λ	ρ	σ^2	**ML**	**GM**	**GM1**
-0.8	-0.8	0.50	0.115	0.130	0.127
-0.8	-0.4	0.50	0.115	0.120	0.121
-0.8	0	0.50	0.105	0.105	0.105
-0.8	0.4	0.50	0.079	0.082	0.081
-0.8	0.8	0.50	0.035	0.065	0.043
-0.4	-0.8	1.00	0.122	0.139	0.135
-0.4	-0.4	1.00	0.124	0.128	0.129
-0.4	0	1.00	0.115	0.113	0.115
-0.4	0.4	1.00	0.092	0.096	0.093
-0.4	0.8	1.00	0.041	0.093	0.062
0	-0.8	0.25	0.107	0.119	0.121
0	-0.4	0.25	0.107	0.112	0.113
0	0	0.25	0.096	0.097	0.098
0	0.4	0.25	0.075	0.075	0.075
0	0.8	0.25	0.037	0.054	0.042
0.4	-0.8	0.50	0.107	0.121	0.123
0.4	-0.4	0.50	0.108	0.116	0.117
0.4	0	0.50	0.100	0.101	0.102
0.4	0.4	0.50	0.081	0.082	0.081
0.4	0.8	0.50	0.048	0.068	0.056
0.8	-0.8	1.00	0.107	0.122	0.124
0.8	-0.4	1.00	0.109	0.117	0.118
0.8	0	1.00	0.100	0.104	0.105
0.8	0.4	1.00	0.083	0.085	0.085
0.8	0.8	1.00	0.068	0.077	0.070
Column average			0.091	0.101	0.098

The response function estimation results are given in Table 11. The first line corresponding to each parameter is the OLS estimate of the coefficient of that parameter based on the log transform of (7) indicated above. The second line contains the absolute value of the corresponding t-ratio. The t-ratios are given for descriptive purposes; while suggestive of statistical significance, these t-ratios, of course, cannot be used in the typical way to formally test hypotheses.

We first note that the fit of the functions to the actual data is quite good in that the R^2 values between the actual $RMSE$s and the corresponding predicted values based on (7) are at least 0.95. The interpretations of the individual coefficients corresponding to λ, ρ, J, and n are not straightforward because each of these parameters appears in more than one form and/or interacts with other parameters. The only exception is the coefficient corresponding to σ^2. This coefficient is "seemingly significant" only in the response function for the ML estimator of λ. However, even in this case its coefficient is "small" in absolute value and suggests that, once σ^2 is accounted for in the proportionality factor, σ^2 is only a minor component in the

Table 11. The estimated response functions

Response function parameters	Estimator							
	λ_{ML}	$\lambda_{FGS2SLS}$	β_{1ML}	$\beta_{1FGS2SLS}$	β_{2ML}	$\beta_{2FGS2SLS}$	ρ_{ML}	ρ_{GM}
a_1	0.6116	0.5600	0.1529	0.1306	0.0569	0.0394	1.1030	0.9856
	(24.86)	(23.95)	(17.74)	(12.68)	(8.23)	(5.07)	(49.39)	(53.06)
a_2	−1.4363	−1.3661	−0.0004	0.0451	0.2676	0.2933	−2.1488	−1.9116
	(25.37)	(25.40)	(0.02)	(1.90)	(16.81)	(16.39)	(41.82)	(44.72)
a_3	−1.6746	−1.7975	0.1697	0.1552	0.1916	0.1660	1.5720	1.2926
	(32.44)	(36.65)	(9.39)	(7.19)	(13.20)	(10.17)	(33.55)	(33.16)
a_4	2.0485	2.1546	0.0545	0.0703	0.0292	0.0435	−0.1278	−0.0995
	(39.69)	(43.93)	(3.010	(3.25)	(2.01)	(2.67)	(2.73)	(2.55)
a_5	0.8736	0.9425	−0.0141	−0.0043	−0.0358	−0.0202	−0.7717	−0.6061
	(54.65)	(62.04)	(2.52)	(0.64)	(7.96)	(4.00)	(53.17)	(50.21)
a_6	0.4854	0.6288	0.0855	0.1347	0.0324	0.0809	−0.8394	−0.2670
	(24.80)	(33.80)	(12.46)	(16.45)	(5.89)	(13.08)	(47.24)	(18.06)
a_7	−1.1637	−1.2206	0.0036	−0.0006	−0.0048	−0.0129	0.0368	0.0069
	(72.80)	(80.35)	(0.65)	(0.09)	(1.08)	(2.55)	(2.54)	(0.57)
a_8	−1.0654	−1.0538	−0.0224	−0.0139	−0.0061	−0.0037	−0.0170	−0.0730
	(54.40)	(56.62)	(3.26)	(1.69)	(1.11)	(0.59)	(0.96)	(4.94)
a_9	0.5067	0.3768	−0.1304	−0.1704	−0.0887	−0.1205	0.2050	0.0646
	(29.30)	(22.93)	(21.52)	(23.54)	(18.24)	(22.05)	(13.06)	(4.94)
a_{10}	1.2177	1.4693	−1.8048	−1.7499	3.2218	3.2518	1.1810	1.4397
	(2.24)	(2.84)	(9.48)	(7.69)	(21.08)	(18.92)	(2.39)	(3.51)
a_{11}	−4.0727	−5.1348	6.4308	6.0978	−10.6989	−10.8124	2.9813	3.0671
	(2.23)	(2.96)	(10.05)	(7.98)	(20.83)	(18.73)	(1.80)	(2.22)
a_{12}	−20.1843	−21.2148	−19.0589	−19.2992	−21.8814	−22.0361	−6.5053	−9.6071
	(12.28)	(13.58)	(33.08)	(28.05)	(47.33)	(42.40)	(4.36)	(7.74)
a_{13}	−0.0388	0.0088	−0.0080	0.0035	−0.0044	0.0056	0.1648	0.2327
	(2.56)	(0.61)	(1.51)	(0.56)	(1.02)	(1.16)	(11.94)	(20.27)
R^2	0.972	0.975	0.987	0.982	0.994	0.992	0.974	0.978

The figures in parentheses are the absolute values of the corresponding t ratios.
The R^2 values relate to the actual $RMSE$s and their predicted values via (7).

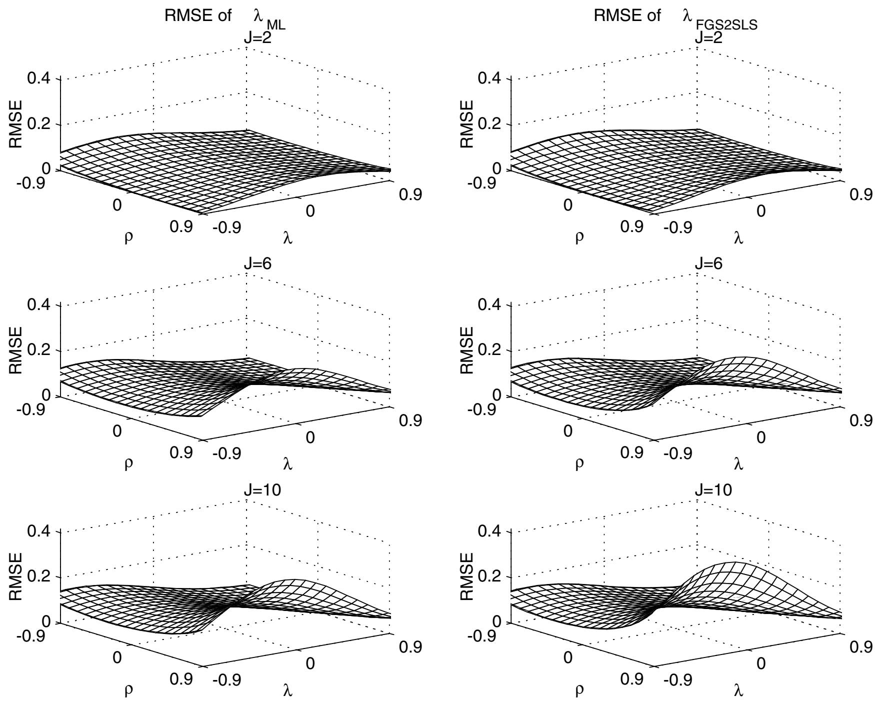

Fig. 1. RMSEs of the ML and FGS2LS estimators of λ $(N = 100, \sigma^2 = 0.5)$

response function for the range of values of σ considered. For example, the absolute value of the elasticity of $(n^{1/2}/\sigma) * RMSE$ with respect to σ^2 is $.04\sigma^2 \leq 0.04$ since the largest value of σ^2 considered is 1.0.

Because the interpretations of most of the coefficients in Table 11 are not straightforward, we describe these eight functions in terms of figures. Figure 1 describes the response functions of the $RMSE$s of the ML and FGS2SLS estimators of λ. This is done in terms of six sub-figures. The three on the left hand side of the page relate to the ML estimator of λ and correspond to $J = 2, 6, 10$. The three on the right hand side of the page relate to the FGS2SLS estimator and have the same format. In all the figures $\sigma^2 = 0.5$ and $n = 100$. Figures 2, 3 and 4 have the same format but correspond to the considered estimators of, respectively, β_1, β_2, and ρ.

Consider Fig. 1. Note first that for both the ML and the FGS2SLS estimators the $RMSE$s increase as J increases. One reason for this may be that λ is the coefficient of $W_n y_n$ in (1) and, since W_n is row normalized, $W_n y_n$ is a vector of averages. The number of terms in these averages is J. Clearly, as J increases the variation of the elements of $W_n y_n$ decreases and hence the precision of estimation concerning λ decreases.

From the figures we also note that the relationship between the $RMSE$s of both the ML and FGS2SLS estimators depends upon ρ and λ in a complex fashion. For instance, these $RMSE$s generally increase as $\rho \to 1$ for $J = 6$ and 10, but decrease

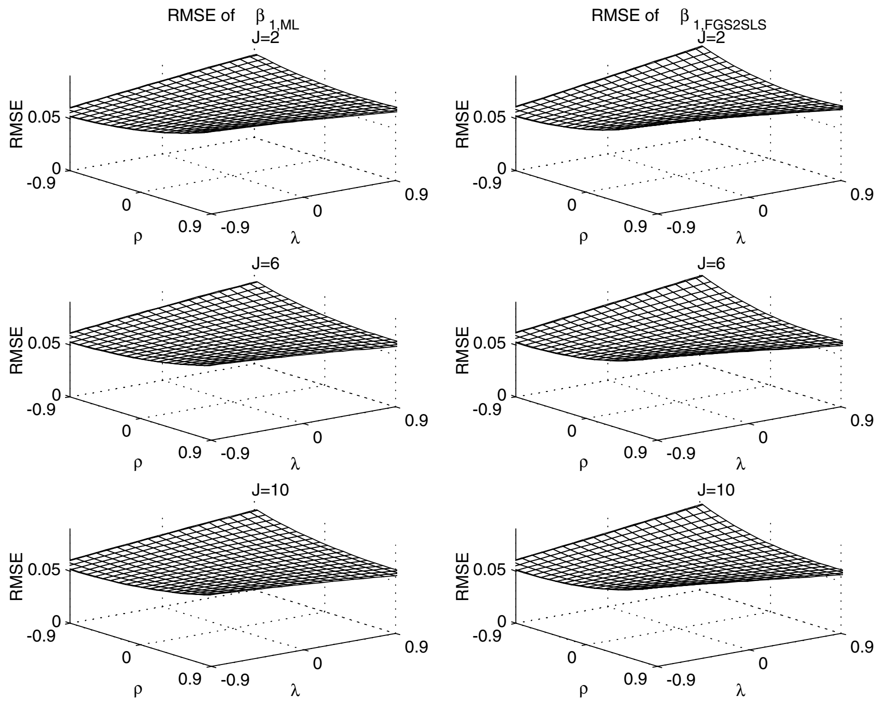

Fig. 2. RMSEs of the ML and FGS2LS estimators of β_1 $(N = 100, \sigma^2 = 0.5)$

as $\lambda \to 1$. For a given value of λ the $RMSE$s relate to ρ in a convex fashion; on the other hand, for a given value of ρ the $RMSE$s relate to λ in a concave fashion. One reason for such a seemingly complex dependence of the $RMSE$s on ρ and λ may relate to the variance of some of the model components. For example, as $\rho \to 1$ the average variance of the error terms u_n in (1) increases. Ceteris paribus, this should make the estimation of the model parameters less precise. However, the ceteris paribus condition is not a relevant one in this case because as the average variance of the error terms u_n increases, so does the variation of the elements of $W_n y_n$ and this would lead to a more precise estimate of λ. Therefore, one would expect that the effect of the experimental value of ρ on the $RMSE$ of an estimator of λ depends, among other things, upon the net effect of these two conditions. Somewhat similar issues relate to λ because as $\lambda \to 1$ the average variance of the elements of y_n, and hence those of $W_n y_n$, increases. Furthermore, since the model solution for y_n in terms of X_n involves the error term $(I - \lambda W_n)^{-1}(I - \rho W_n)^{-1}\varepsilon_n$ interactions between ρ and λ, and of course J, should be evident.

Consider now Figs. 2 and 3 which relate to the remaining regression coefficients β_1 and β_2. Clearly, the $RMSE$s of both the ML and FGS2SLS estimators of β_1 and β_2 slightly decrease as J increases. This is the opposite of the results for the estimators of λ. The reason for this may be that as J increases, the average variation in the elements of $W_n y_n$ decreases, and hence the correlation between

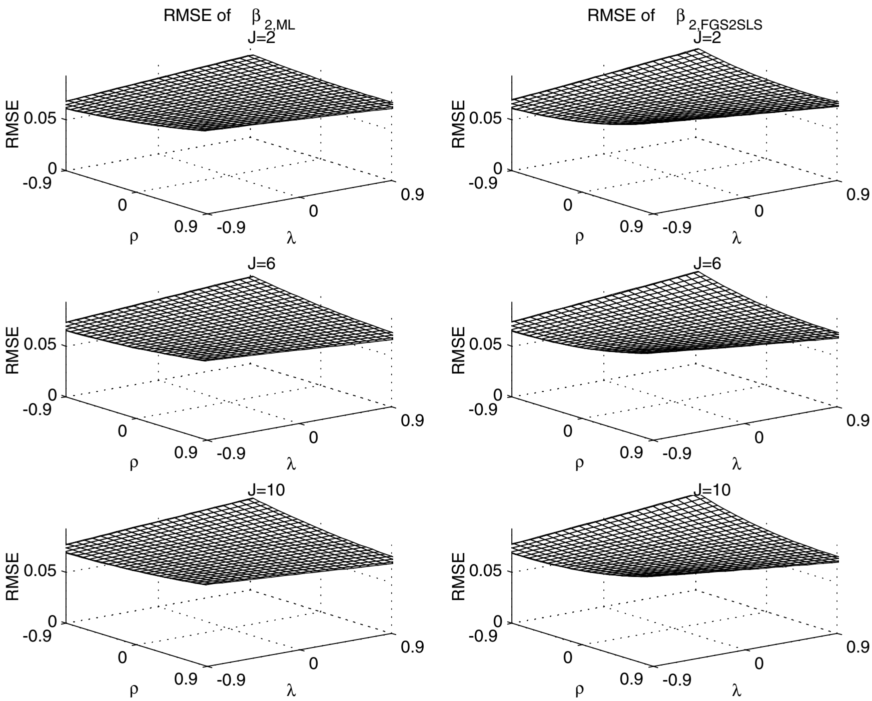

Fig. 3. RMSEs of the ML and FGS2LS estimators of β_2 $(N = 100, \sigma^2 = 0.5)$

x_{1n} and $W_n y_n$, and x_{2n} and $W_n y_n$ decreases.[15] Thus, the precision of estimation concerning β_1 and β_2 increases.

Again the dependence between the $RMSE$s of both the ML and FGS2SLS estimators of β_1 and β_2 on ρ and λ, while slight, is complex. Specifically, for $\lambda < 0$ these $RMSE$s decrease as ρ decreases, while for $\lambda > 0$ these $RMSE$s increase as ρ decreases, and the $RMSE$s appear to minimize as ρ and λ approach -1.

Finally consider Fig. 4. In a manner similar to the cases for λ, the $RMSE$s of both the ML and GM estimators of ρ generally increase as J increases. On an intuitive level the reason for this is that ρ is the coefficient of $W_n u_n$ in (2), and the average variance of the elements of $W_n u_n$ decreases as J increases; hence the precision of estimation decreases. Conversely, unlike in the case for λ, the $RMSE$s of the considered estimators of ρ do not seem to depend much on the value of λ. However, they do depend on ρ. Specifically, these $RMSE$s decrease as $\rho \to 1$;

[15] Observe that in light of the above discussion:

$$E(W_n y_n) = W_n E(y_n) = W_n (I - \lambda W_n)^{-1} X_n \beta$$
$$= \sum_{i=0}^{\infty} \lambda^i W_n^{i+1} X_n \beta.$$

As J increases $E(W_n y_n)$ becomes a vector of averages involving more terms of the elements of x_{1n} and x_{2n}, and so the correlation between between x_{1n} and x_{2n}, and $W_n y_n$ would be expected to decrease.

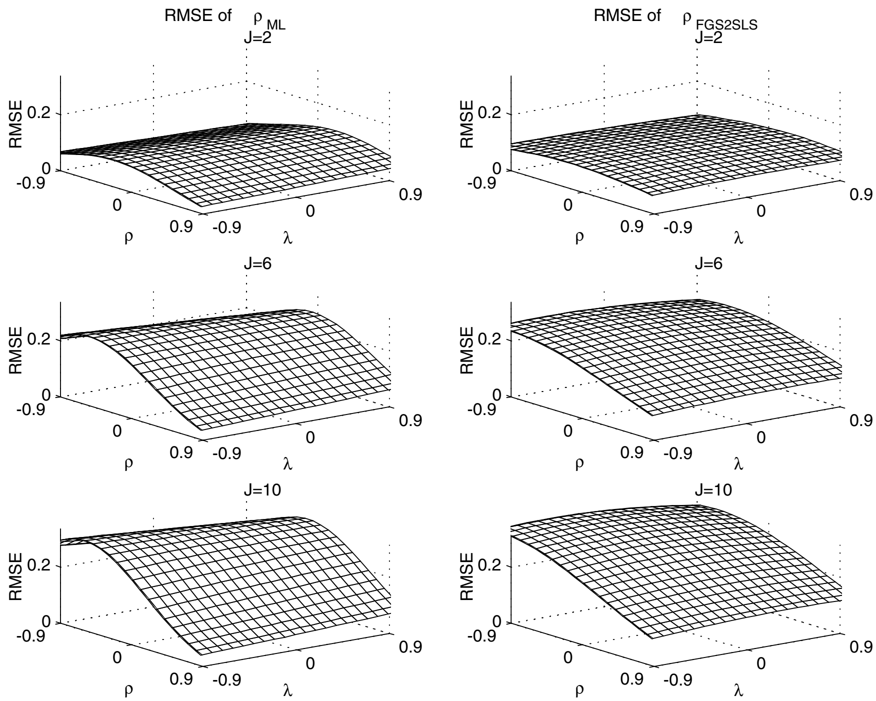

Fig. 4. RMSEs of the ML and FGS2SLS estimators of ρ $(N = 100, \sigma^2 = 0.5)$

they also decrease as $\rho \to -1$ when $J = 2$. For a given value of λ the $RMSE$s seem to have a maximum at $\rho = 0$ when $J = 2$; this maximum occurs at a negative value of ρ for $J = 6$ or 10. The reason for this shift, as explained in Kelejian and Prucha (1999), is that the values of a for which $I - aW_n$ is singular are less than -1 when $J = 6$, 10.

5 Summary and suggestions for further research

Among other things, we have given Monte Carlo results relating to the finite sample properties of the ML and FGS2SLS estimators of the regression parameters of a linear spatial model containing a spatially lagged dependent variable as well as a spatially autocorrelated disturbance term. Our results suggest that the difference in finite sample efficiency between the ML and FGS2SLS estimators is fairly small, so the FGS2SLS estimator can be considered with little penalty. This is important because the FGS2SLS estimator is computationally feasible even in cases involving large samples with non-sparse and non-symmetric weights matrices; in such cases the computation of the ML estimator may not be feasible, or is at least “challenging”.

We have also given finite sample results relating to the ML and GM estimator of the autoregressive parameter ρ in the disturbance process. Again, the results suggest minor differences in their finite sample efficiencies. Thus, the computationally simple GM estimator can be considered with little penalty.

An obvious suggestion for future research relates to nonlinear models. Specifically, to the best of our knowledge neither theoretical nor Monte Carlo results are available for nonlinear spatial models which contain both a spatially correlated error term and a spatially lagged dependent variable. Although certain estimation procedures may seem evident for such models, use of Monte Carlo as well as formal estimation results concerning estimation would be interesting.

Appendix A: GM estimator

Consider (6) in the text and let $\hat{u}_n = y_n - Z_n\hat{\delta}_{2SLS}$. Also let:

$$\nu_{n1}(\underline{\rho}, \underline{\sigma_\varepsilon^2}) = n^{-1}(\hat{u}_n - \underline{\rho} W_n \hat{u}_n)'(\hat{u}_n - \underline{\rho} W_n \hat{u}_n) - \underline{\sigma_\varepsilon^2} \; ,$$
$$\nu_{n2}(\underline{\rho}, \underline{\sigma_\varepsilon^2}) = n^{-1}(W_n\hat{u}_n - \underline{\rho} W_n^2 \hat{u}_n)'(W_n\hat{u}_n - \underline{\rho} W_n^2 \hat{u}_n) - \underline{\sigma_\varepsilon^2} \, n^{-1} tr(W_n' W_n) \; ,$$
$$\nu_{n3}(\underline{\rho}, \underline{\sigma_\varepsilon^2}) = n^{-1}(W_n\hat{u}_n - \underline{\rho} W_n^2 \hat{u}_n)'(\hat{u}_n - \underline{\rho} W_n \hat{u}_n) \; .$$

Let $\nu_n'(\underline{\rho}, \underline{\sigma_\varepsilon^2}) = (\nu_{n1}(\underline{\rho}, \underline{\sigma_\varepsilon^2}), \nu_{n2}(\underline{\rho}, \underline{\sigma_\varepsilon^2}), \nu_{n3}(\underline{\rho}, \underline{\sigma_\varepsilon^2}))$. Then, the GM estimators of ρ and σ_ε^2, say $\hat{\rho}$ and $\hat{\sigma}_\varepsilon^2$, suggested by Kelejian and Prucha (1999) are:

$$(\hat{\rho}, \hat{\sigma}_\varepsilon^2) = \arg\min(\nu_n'(\underline{\rho}, \underline{\sigma_\varepsilon^2})\nu_n(\underline{\rho}, \underline{\sigma_\varepsilon^2}) : \; \underline{\rho} \in [-a, a]; \underline{\sigma_\varepsilon^2} \in [0, s^2]$$

where $a \geq 1$ and s^2 is the upper limit considered for σ_ε^2. The first iterated GM estimator is identical to that defined above except that $\hat{u}_n$ is replaced by $\hat{u}_n^{(1)} = y_n - Z_n\hat{\delta}_{FGS2SLS}$, etc.

Appendix B: Nonlinear FGS2SLS estimator

Consider the 2SLS objective function:

$$Q_n(\overline{\lambda}, \overline{\beta}) = \left[y_n(\hat{\rho}_{GM}) - \hat{Z}_n(\hat{\rho}_{GM})\overline{\delta}\right]' P_{H_n} \left[y_n(\hat{\rho}_{GM}) - \hat{Z}_n(\hat{\rho}_{GM})\overline{\delta}\right] \quad \text{(B.1)}$$

with $\overline{\delta} = (\overline{\lambda}, \overline{\beta}')'$ and $P_{H_n} = H_n(H_n'H_n)^{-1}H_n'$. The (linear) FGS2SLS based on $\hat{\rho}_{GM}$ considered in our Monte Carlo study, that is $\hat{\delta}_{FGS2SLS} = (\widehat{\lambda}_{FGS2SLS}, \widehat{\beta}'_{FGS2SLS})'$, is then readily seen to be the unconstrained minimizer of the objective function (B.1). That is, it is obtained by solving the problem:

$$\min_{\overline{\lambda} \in \mathbf{R}, \overline{\beta} \in \mathbf{R}^k} Q_n(\overline{\lambda}, \overline{\beta})$$

where $\mathbf{R}$ and $\mathbf{R}^k$ denote the real line and the k-dimensional real space, respectively. Now let $\hat{\delta}_{NFGS2SLS} = (\widehat{\lambda}_{NFGS2SLS}, \widehat{\beta}'_{NFGS2SLS})'$ denote the estimator obtained by solving the corresponding constrained problem:

$$\min_{|\overline{\lambda}| < 1, \overline{\beta} \in \mathbf{R}^k} Q_n(\overline{\lambda}, \overline{\beta}).$$

Consistent with usual terminology we refer to this estimator as the nonlinear FGS2SLS estimator. Clearly, by construction this estimator satisfies the constraint

$\left|\widehat{\lambda}_{NFGS2SLS}\right| < 1$. Also note that whenever the unconstrained (linear) FGS2SLS estimator for λ is less than one in absolute value, i.e., whenever $\left|\widehat{\lambda}_{FGS2SLS}\right| < 1$, then the linear and nonlinear FGS2SLS coincide, i.e., $\hat{\delta}_{FGS2SLS} = \hat{\delta}_{NFGS2SLS}$.

The asymptotic distribution of $\hat{\delta}_{FGS2SLS}$ was derived in Kelejian and Prucha (1998) under a general set of assumptions. We next show that $\hat{\delta}_{FGS2SLS}$ and $\hat{\delta}_{NFGS2SLS}$ have the same asymptotic distribution, and thus the results obtained for $\hat{\delta}_{FGS2SLS}$ carry over to $\hat{\delta}_{NFGS2SLS}$. A sufficient condition for $\hat{\delta}_{FGS2SLS}$ and $\hat{\delta}_{NFGS2SLS}$ to have the same asymptotic distribution is that:

$$p \lim_{n\to\infty} n^{1/2}(\hat{\delta}_{FGS2SLS} - \hat{\delta}_{NFGS2SLS}) = 0.$$

This clearly holds if the probability of the event where $\hat{\delta}_{FGS2SLS} = \hat{\delta}_{NFGS2SLS}$ goes to one as $n \to \infty$. Since the true parameter vector $\delta = (\lambda, \beta')'$ is an interior point of the open parameter space $\{(\overline{\lambda}, \overline{\beta}')' : \left|\overline{\lambda}\right| < 1, \overline{\beta} \in \mathbf{R}^k\}$ it follows from the consistency of $\hat{\delta}_{FGS2SLS}$ that the probability of the event where $\hat{\delta}_{FGS2SLS}$ falls inside the parameter space, or equivalently where $\left|\widehat{\lambda}_{FGS2SLS}\right| < 1$, goes to one as $n \to \infty$. However, as remarked above, whenever $\left|\widehat{\lambda}_{FGS2SLS}\right| < 1$, then $\hat{\delta}_{FGS2SLS} = \hat{\delta}_{NFGS2SLS}$, and thus the probability of the event where $\hat{\delta}_{FGS2SLS} = \hat{\delta}_{NFGS2SLS}$ goes to one as $n \to \infty$, which proves the claim.

Appendix C: Discussion of the RMSE measure

Let $\widehat{\theta}_n$ be an estimator of θ based on a sample of size n. Then the standard measure of the root mean squared error of $\widehat{\theta}_n$ is $RMSE_n^* = [E(\widehat{\theta}_n - \theta)^2]^{1/2}$. Assuming that the mean and variance of $\widehat{\theta}_n$ exist,

$$RMSE_n^* = \left[[E(\widehat{\theta}_n) - \theta]^2 + E[\widehat{\theta}_n - E(\widehat{\theta}_n)]^2\right]^{1/2}. \tag{C.1}$$

The expression in (C.1) provides a decomposition of $RMSE_n^*$ in terms of the bias and variance of the estimator.

The measure used in this study, which is described in the text, is:

$$RMSE_n = \left[[med(\widehat{\theta}_n) - \theta]^2 + [IQ(\widehat{\theta}_n)/1.35]^2\right]^{1/2} \tag{C.2}$$

where $med(\widehat{\theta}_n)$ is the median (or the 0.5 quantile) and $IQ(\widehat{\theta}_n)$ is the inter-quantile range of the distribution of $\widehat{\theta}_n$. As described in the text, IQ is the difference between the 0.75 and 0.25 quantiles.

If the distribution of $\widehat{\theta}_n$ is normal, $med(\widehat{\theta}_n) = E(\widehat{\theta}_n)$ and, except for a slight rounding error, $[IQ(\widehat{\theta}_n)/1.35]^2 = E[\widehat{\theta}_n - E(\widehat{\theta}_n)]^2$. Therefore $RMSE_n^* = RMSE_n$. If the distribution of $\widehat{\theta}_n$ is not normal $RMSE_n^*$ and $RMSE_n$ will not generally be equal.

The decomposition in (C.1) is defined in terms of the first two (population) moments of the distribution of the estimator $\widehat{\theta}_n$. That decomposition will therefore exist, i.e., be meaningful, only if the first two (population) moments exist. As is well known, in certain cases the (population) mean and variance of the finite sample distribution of an estimator may not exist. See, e.g., Phillips (1983). This existence problem is relevant for all estimators considered in this study, even if the distribution of the error vector in (2) is normal. The reason for this is that all of these estimators are nonlinear in the endogenous vector y_n. In other words, these nonlinearities not only make the determination of the (population) moments "difficult", but raise the possibility that at least some of them may not even exist! Clearly, if these moments do not exist, Monte Carlo estimates of them are meaningless. In contrast, we note that in all cases the $0.25, 0.5$, and 0.75 quantiles of a distribution do exist; and hence our measure $RMSE_n$, as well as its components, always exist. Corresponding Monte Carlo estimates are therefore meaningful. It is for this reason that we have adopted the measure $RMSE_n$.

References

Anselin L (1988) *Spatial econometrics: methods and models*. Kluwer Academic Publishers, Boston

Anselin L, Bera A, Florax R, Yoon M (1996) Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26: 77–104

Anselin L, Florax RJGM (1995) Small sample properties of tests for spatial dependence in regression models: Some further results. In: Anselin L, Florax RJGM (eds) *New directions in spatial econometrics*. Springer, Berlin Heidelberg New York

Baltagi BH, Song SH, Koh W (2000) Testing panel data regression models with spatial error correlation. Department of Economics, Texas A&M University

Bell K, Bockstael N (2000) Applying the generalized moments estimation approach to spatial problems involving micro-level data. *The Review of Economics and Statistics* 82: 72–82

Bernat Jr G (1996) Does manufacturing matter? A spatial econometric view of Kaldor's laws. *Journal of Regional Science* 36: 463–477

Boarnet M (1994) An empirical model of intrametropolitan population and employment growth. *Papers in Regional Science* 73: 135–152

Bollinger C, Ihlanfeldt K (1997) The impact of rapid rail transit on economic development: The case of Atlanta's Marta. *Journal of Urban Economics* 42: 179–204

Buettner T (1999) The effect of unemployment, aggregate wages, and spatial contiguity on local wages: An investigation with German district level data. *Papers in Regional Science* 78: 47–67

Case A (1992) Neighborhood influence and technological change. *Regional Science* 22: 491–508

Case A, Hines Jr J, Rosen H (1993) Budget spillovers and fiscal policy independence: Evidence from the states. *Journal of Public Economics* 52: 285–307

Cliff A, Ord J (1973) *Spatial autocorrelation.* Pion, London

Cliff A, Ord J (1981) *Spatial processes, models and applications.* Pion, London

Cohen J (1998) Essays on the public provision of airport infrastructure. University of Maryland, Ph.D. Thesis

Cressie NAC (1993) *Statistics of spatial data*. Wiley, New York

Das D (2000) The small sample distribution of parameter estimators in a spatial ARAR(1,1) model: A Monte Carlo study. University of Maryland, Ph.D. Thesis

Dhrymes P (1981) *Distributed lags, problems of estimation and formulation*. North Holland, New York

Driscoll J, Kraay A (1998) Consistent covariance matrix estimation with spatially dependent panel data. *The Review of Economics and Statistics* LXXX: 549–560

Florax RJGM, Rey S (1995) Spatial effects in linear regression models, spatial data and model transformations. In: Anselin L, Florax RJGM (eds) *New directions in spatial econometrics*. Springer, Berlin Heidelberg New York

Kelejian HH, Prucha IR (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17: 99–121

Kelejian HH, Prucha IR (1999) A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40: 509-533

Kelejian HH, Prucha IR (2000) 2SLS and OLS in a spatial autoregressive model with equal spatial weights. *Regional Science and Urban Economics* (forthcoming)

Kelejian HH, Prucha IR (2001) On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics*

Kelejian HH, Robinson D (1992) Spatial autocorrelation: A new computationally simple test with an application to per capita county police expenditures. *Regional Science and Urban Economics* 22: 317–331

Lee LF (1999a) Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Department of Economics, HKUST, Hong Kong

Lee LF (1999b) Asymptotic distributions of maximum likelihood estimators for spatial autoregressive models. Department of Economics, HKUST, Hong Kong

Lee LF (2002) Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregessive models. *Econometric Theory* 18: 252–277

Ord J (1975) Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70: 120–126

Pace R, Barry R (1997) Sparse spatial autoregressions. *Statistics and Probability Letters* 33: 291–297

Phillips, PCP (1983) Exact small sample theory in the simultaneous equations model. In: Griliches Z, Intriligator M (eds) *Handbook of Econometrics, Vol. 1*. North Holland, New York

Rey SJ, Boarnet MG (1998) A taxonomy of spatial econometric models for simultaneous equation systems. Department of Geography, San Diego State University

Whittle P (1954) On stationary processes in the plane. *Biometrica* 41: 434–449