ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 735 October 2013

Inference on Higher-Order Spatial Autoregressive Models with Increasingly Many Parameters

Abhimanyu Gupta and Peter M. Robinson

Inference on Higher-Order Spatial Autoregressive Models with Increasingly Many Parameters

Abhimanyu Gupta*
Department of Economics
University of Essex, UK

Peter M. Robinson †‡
Department of Economics
London School of Economics, UK

November 22, 2013

Abstract

This paper develops consistency and asymptotic normality of parameter estimates for a higher-order spatial autoregressive model whose order, and number of regressors, are allowed to approach infinity slowly with sample size. Both least squares and instrumental variables estimates are examined, and the permissible rate of growth of the dimension of the parameter space relative to sample size is studied. Besides allowing the number of estimable parameters to increase with the data, this has the advantage of accommodating explicitly some asymptotic regimes that arise in practice. Illustrations are discussed, in particular settings where the need for such theory arises quite naturally. A Monte Carlo study analyses various implications of the theory in finite samples. For empirical researchers our work has implications for the choice of model. In particular if the structure of the spatial weights matrix assumes a partitioning of the data then spatial parameters should be allowed to vary over clusters.

JEL classifications: C21, C31, C33, C36

Keywords: Spatial autoregression; increasingly many parameters; central limit theorem; rate of convergence; spatial panel data

Email: a.gupta@essex.ac.uk.

Inputs and suggestions by Peter Burridge, Swati Chandna, Javier Hidalgo, Francesca Rossi and Marcia Schafgans are gratefully acknowledged.

† *Email*: p.m.robinson@lse.ac.uk, *Telephone*: +44-20-7955-7516 *Fax*: +44-20-7955-6592.

‡ Research supported by ESRC Grant ES/J007242/1.

1 Introduction

Correlation in cross-sectional data poses considerable challenges, complicating both modelling and statistical inference. When information on geographical locations is available, it may be possible to extend models developed for time series data. However, when locations are irregularly-spaced serious difficulties arise, and frequently only information on economic (not necessarily geographic) distances is available. Spatial autoregressive (SAR) models, due to Cliff and Ord (1973), have become widely used in this setting. Given a sample of size n, these employ a known $n \times n$ spatial weights matrix whose (i, j)-th element is frequently inversely related to some measure of economic distance between units i and j. The elements may also be binary, for instance taking equal values 1 when two units are contiguous and 0 otherwise, but many other other specifications are possible.

To be specific, for an $n \times 1$ vector of observations y_n, an $n \times k$ matrix of regressors X_n and $n \times n$ weight matrices W_{in}, $i = 1, \ldots, p$, it is assumed that there exist unknown scalars $\lambda_1, \lambda_2, \ldots, \lambda_p$ and an unknown $k \times 1$ vector β such that

$$y_n = \sum_{i=1}^{p} \lambda_i W_{in} y_n + X_n \beta + U_n, \tag{1.1}$$

where U_n is an $n \times 1$ vector of disturbances. In this paper we will refer to the above as the SAR model while the SAR model without X_n will be the *pure* SAR model.

Weight matrices need not be symmetric and can contain negative elements, but their diagonal elements are normalised to zero, and they are frequently row-normalised such that each row sums to 1. If W_{in} has non-negative elements, this implies that $\|W_{in}\|_R = 1$, where for any $s \times q$ matrix $A = [a_{ij}]$ we define $\|A\|_R = \max_{i=1,\ldots,s} \sum_{j=1}^{q} |a_{ij}|$, which is the maximum absolute row-sum norm. In this case, taking $p = 1$ for illustrative purposes, the (i, j)-th element of W_{1n} can be interpreted as $w_{ij,1n} = d_{ij,1n} / \sum_{h=1}^{n} d_{ih,1n}$, where $d_{ij,1n}$ measures distance between units i and j. Thus elements of the W_{in} are allowed to depend on n, so those of y_n form a triangular array. Kelejian and Prucha (2010) study some implications of row normalisation for parameter interpretation. Since X_n may also depend on spatial weights, we also allow its elements to depend on n. See e.g. Arbia (2006) for a review of spatial autoregressions.

By far the most popular version of (1.1) takes $p = 1$, when we write

$$y_n = \lambda W_n y_n + X_n \beta + U_n. \tag{1.2}$$

Due to the spatially lagged y_n on the right, ordinary least squares (OLS) estimation of λ and β is problematic, but Lee (2002) showed that under suitable conditions such estimates can be consistent, and asymptotically normal and efficient. In particular, for a divergent positive sequence h_n that is bounded away from zero uniformly in n, consistency follows if $w_{ij,n} = \mathscr{O}\left(h_n^{-1}\right)$ and asymptotic normality if also $n^{\frac{1}{2}}/h_n \to 0$ as $n \to \infty$.

Instrumental variables (IV) estimation (see Kelejian and Prucha (1998)) is $n^{\frac{1}{2}}$-consistent under less restrictive conditions than the OLS estimate, but is inefficient. On the other hand, it is computationally simpler than estimates which may have better statistical properties, such as the generalized method of moments estimates of Kelejian and Prucha (1999) and Lee and Liu (2010), the optimal IV estimate of Lee (2003), the (Gaussian) pseudo maximum likelihood estimate studied by Lee (2004), and the adaptive estimate of Robinson (2010). Additionally, desirable asymptotic properties of OLS and IV estimates require X_n to contain at least one non-intercept regressor.

In this paper we allow the spatial lag order p in (1.1) and the number of regressors k to increase slowly with n, as opposed to being fixed. This scheme reflects the practical reality that the richness of a parametric model often deepens with sample size, and has been explored previously in various settings. In the present context, it can arise from a natural extension of the weight matrix employed by Case (1991, 1992). In her scenario data are recorded in p districts, each of which contains m farmers, implying $n = mp$. Farmers within each district impact each other equally and there is inter-district independence between farmers so that in (1.2)

$$W_n = I_p \otimes B_m, \text{ with } B_m = \frac{1}{m-1}\left(l_m l_m' - I_m\right), \tag{1.3}$$

where prime denotes transposition, l_m is the m-dimensional vector of ones $(1, \ldots, 1)'$, I_m is the m-dimensional identity matrix and $\otimes$ denotes Kronecker product. An extension to (1.1) takes

$$W_{in} = diag\left[0, \ldots, \underbrace{B_m}_{i^{th}\text{diagonal block}}, \ldots, 0\right],\ i = 1, \ldots, p, \tag{1.4}$$

so SAR parameters are allowed to vary across districts, perhaps due to geographic or demographic differences, say. In asymptotic theory for estimates of λ and β in (1.2) with (1.3), p is often allowed to increase, perhaps slowly, with n. Indeed if only m increases and p remains fixed, the undesirable situation of 'infill-asymptotics' arises. Allowing the number of districts to increase with n in (1.1) with (1.4) entails increasingly many λ_i. As in the statistical literature on regression models, we also allow k to increase slowly with n. This theme has been pursued in a variety of models (see e.g. Huber (1973), Berk (1974), Portnoy (1984, 1985), Robinson (1979, 2003)). We know of no such literature dealing with SAR models, however.

The following section introduces some assumptions that are basic to our theoretical results. Further assumptions, and theorems, for the consistency and asymptotic normality of IV and OLS estimates are presented in Sections 3 and 4 respectively. In Section 5 we consider some illustrations and then conduct a Monte Carlo study in Section 6, while Section 7 concludes. Proofs may be found in appendices.

2 Model and basic assumptions

We rewrite (1.1) to stress the possible dependence of the parameter dimension, and the parameters themselves, on n:

$$y_n = \sum_{i=1}^{p_n} \lambda_{in} W_{in} y_n + X_n \beta_{(n)} + U_n, \tag{2.1}$$

where $\lambda_{(n)} = (\lambda_{1n}, \ldots, \lambda_{p_n n})'$ and $\beta_{(n)} = (\beta_{1n}, \ldots, \beta_{k_n n})'$. We may write (2.1) as

$$S_n y_n = X_n \beta_{(n)} + U_n \tag{2.2}$$

where $S_n = I_n - \sum_{i=1}^{p_n} \lambda_{in} W_{in}$, or equivalently $y_n = R_n \lambda_{(n)} + X_n \beta_{(n)} + U_n$ with $R_n = (W_{1n} y_n, \ldots, W_{p_n n} y_n)$. We now introduce some basic assumptions.

Assumption 1. *$U_n = (u_1, \ldots, u_n)'$ has iid elements with zero mean and finite variance σ^2.*

Assumption 2. *For $i = 1, \ldots, p_n$, the diagonal elements of each W_{in} are zero and the off-diagonal elements of W_{in} are uniformly $\mathscr{O}\left(h_n^{-1}\right)$, where h_n is some positive sequence which is bounded away from zero and which may be bounded or divergent, with $n/h_n \to \infty$ as $n \to \infty$.*

Different h_{in} sequences for each of the W_{in} may be used. However for least squares estimation, even for fixed p, Lee (2002) demonstrated that consistency requires divergence so that $\min_{i=1,\ldots,p_n} h_{in} \to \infty$ must be assumed and Assumption 2 entails no loss of generality. He also provides a detailed discussion of this assumption. In IV estimation, any mixture of bounded and divergent h_{in} sequences may be employed. However boundedness away from zero is crucial as even consistency of the error variance estimate based on IV residuals may fail if this does not hold. Indeed, in the 'farmer-district' setting discussed in the previous section, $h_n = m - 1$, i.e. it is the number of neighbours of a unit and it is rather odd to allow this to go to zero as the sample size increases.

Assumption 3. *S_n is non-singular for sufficiently large n.*

This assumption ensures that (2.2) has a solution for y_n. In certain special cases such as the farmer-district setting presented above, a sufficient condition can be provided for Assumption 3. The proof of the following can be found in the appendix.

Proposition 2.1. *A sufficient condition for invertibility of $S_n(\lambda_{(n)})$ when, for each $i = 1, \ldots, p_n$, $\|W_{in}\|_R \leq 1$ and each W_{in} has a single non-zero diagonal block structure is that $|\lambda_{in}| < 1$ for each $i = 1, \ldots, p_n$.*

Assumption 4. *$\|S_n^{-1}\|_R$, $\|S_n'^{-1}\|_R$, $\|W_{in}\|_R$ and $\|W_{in}'\|_R$ are uniformly bounded in n and i, $i = 1, \ldots, p_n$, for sufficiently large n.*

This assumption is standard, the parts pertaining to S_n^{-1} ensuring that the spatial correlation is curtailed to a manageable degree because the covariance matrix of y_n is $\sigma^2 S_n^{-1} S_n'^{-1}$, while those for the W_{in} are satisfied trivially if one unit is assumed to be a 'neighbour' of only a finite number of other units, and more generally satisfied if, for each i, the elements of W_{in} decline fast enough with n, as is natural if they are inverse distances. The final assumption is also standard.

Assumption 5. *The elements of X_n are constants and are uniformly bounded, in absolute value, for all sufficiently large n.*

3 Instrumental variables estimates

Because of the endogeneity of the $W_{in} y_n$, $i = 1, \ldots, p_n$, IV estimation has been employed for estimation of SAR models. Let Z_n be an $n \times r_n$ matrix of instruments, with $r_n \geq p_n$ for all n, and introduce

Assumption 6. *The elements of Z_n are constants and are uniformly bounded in absolute value.*

For the model (1.2), Kelejian and Prucha (1998) noted that $W_n\mathbb{E}(y_n)$ can be written as an infinite linear combination of the columns of the matrices X_n, W_nX_n, $W_n^2X_n$, $W_n^3X_n$,..., assuming the existence of a convergent power series for $(I_n - \lambda W_n)^{-1}$. The existence of such a series is guaranteed if $\|\lambda W_n\|_{\mathcal{M}} < 1$, where $\|\cdot\|_{\mathcal{M}}$ denotes any matrix norm (see e.g. Kreyszig (2011), pg. 398). It was suggested that the instrument matrix be constructed from linearly independent subsets of the columns of

$$X_n, W_nX_n, W_n^2X_n, \ldots, W_n^sX_n, \text{ for some } s \geq 1.$$

Our theory allows the number of instruments to increase with sample size and provides a new result for the case when p_n is fixed while r_n is allowed to diverge with n.

Write $\overline{\eta}(B)$ (respectively $\underline{\eta}(B)$) for the largest (smallest) eigenvalue of a square matrix B. For a rectangular matrix A, define the spectral norm $\|A\| = \{\overline{\eta}(A'A)\}^{\frac{1}{2}}$. For the specification (2.1),

$$\mathbb{E}(y_n) = S_n^{-1}X_n\beta_{(n)} = \left[\sum_{k=0}^{\infty}\left(\sum_{i=1}^{p_n}\lambda_{in}W_{in}\right)^k\right]X_n\beta_{(n)}, \tag{3.1}$$

assuming that the power series is well-defined, for which a sufficient condition is

$$\left\|\sum_{i=1}^{p_n}\lambda_{in}W_{in}\right\| < 1. \tag{3.2}$$

For example, when the W_{in} take the form (1.4), then

$$\sum_{i=1}^{p_n}\lambda_{in}W_{in} = diag\left(\lambda_{1n}B_m, \lambda_{2n}B_m, \ldots, \lambda_{p_nn}B_m\right), \tag{3.3}$$

implying that $\|\sum_{i=1}^{p_n}\lambda_{in}W_{in}\| \leq \max_{i=1,\ldots,p_n;\, n\geq 1}|\lambda_{in}|\,\|B_m\|$. B_m has one eigenvalue equal to 1 and also $-1/(m-1)$ as an eigenvalue with multiplicity $m-1$. Hence $\|B_m\| = 1$ and $\max_{i=1,\ldots,p_n;\, n\geq 1}|\lambda_{in}| < 1$ suffices for (3.2). See also Proposition 2.1 for a similar result. If the power series in (3.1) is indeed valid then instruments may be constructed as subsets of the linearly independent columns of

$$X_n, W_{1n}X_n, W_{1n}^2X_n, \ldots, W_{2n}X_n, W_{2n}^2X_n, \ldots, W_{p_nn}X_n, W_{p_nn}^2X_n, \ldots \tag{3.4}$$

Columns of X_n pre-multiplied by cross-products of the W_{in} may also be employed in view of (3.1). Other choices of instruments from outside the model may be available to the practitioner.

Denoting $\theta_{(n)} = \left(\lambda'_{(n)}, \beta'_{(n)}\right)'$, define the IV estimate of $\theta_{(n)}$ as

$$\hat{\theta}_{(n)} = \frac{1}{n}\hat{Q}_n^{-1}\hat{K}_n'J_n^{-1}\begin{bmatrix} Z_n' \\ X_n' \end{bmatrix} y_n = \theta_{(n)} + \hat{Q}_n^{-1}\hat{K}_n'J_n^{-1}q_n, \tag{3.5}$$

where $\hat{Q}_n = \hat{K}_n'J_n^{-1}\hat{K}_n$ and

$$\hat{K}_n = \frac{1}{n}\begin{bmatrix} Z_n' \\ X_n' \end{bmatrix}[R_n, X_n],\ J_n = \frac{1}{n}\begin{bmatrix} Z_n' \\ X_n' \end{bmatrix}[Z_n, X_n],\ q_n = \frac{1}{n}\begin{bmatrix} Z_n' \\ X_n' \end{bmatrix}U_n.$$

Since (2.2) and Assumption 3 imply that $y_n = S_n^{-1}X_n\beta_{(n)} + S_n^{-1}U_n$ we can write $R_n = A_n + B_n$ where

$$A_n = (G_{1n}X_n\beta_{(n)}, \ldots, G_{p_nn}X_n\beta_{(n)}),\ B_n = (G_{1n}U_n, \ldots, G_{p_nn}U_n),$$

and $G_{in} = W_{in}S_n^{-1}$ for $i = 1, \ldots, p_n$. Also define

$$K_n = \frac{1}{n}\begin{bmatrix} Z_n' \\ X_n' \end{bmatrix}[A_n, X_n],\ Q_n = K_n'J_n^{-1}K_n,\ L_n = \frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix}[A_n, X_n].$$

Denote limit superior (respectively limit inferior) by $\overline{\lim}_{n\to\infty}$ ($\underline{\lim}_{n\to\infty}$), and introduce

Assumption 7. $\overline{\lim}_{n\to\infty}\overline{\eta}(J_n) < \infty$ *and* $\underline{\lim}_{n\to\infty}\underline{\eta}(K_n'K_n) > 0$.

Assumption 8. $\underline{\lim}_{n\to\infty}\underline{\eta}(J_n) > 0$ *and* $\overline{\lim}_{n\to\infty}\overline{\eta}(K_n'K_n) < \infty$.

These are asymptotic non-multicollinearity and boundedness conditions, which can to some extent be checked. For instance, if X_n contains a column of ones (i.e. the model (2.1) has an intercept) and there exists a row-normalised W_{in} with equal off-diagonal elements (such as (1.4) defined below) then $W_{in}y_n$ is asymptotically collinear with the intercept. In this case $\underline{\lim}_{n\to\infty} \underline{\eta}(K_n'K_n) > 0$ fails, and in fact so does Assumption 10, introduced later. This problem is discussed further in Kelejian and Prucha (2002). A necessary condition for both Assumption 10 and $\underline{\lim}_{n\to\infty} \underline{\eta}(K_n'K_n) > 0$ to hold is that, for all $i = 1, \ldots, p_n$, W_{in} are linearly independent for sufficiently large n, failing which some of the λ_{in} are not identified. It is clear that identification of the λ_{in} is particularly transparent when the W_{in} have a single non-zero block structure, a situation that we have discussed in detail in Section 1.

Lemma 3.1. *Under Assumptions 7 and 8 respectively*

(i) $\underline{\lim}_{n\to\infty} \underline{\eta}(Q_n) > 0.$

(ii) $\overline{\lim}_{n\to\infty} \overline{\eta}(Q_n) < \infty.$

Theorem 3.1. *Let Assumptions 1-7 hold and*

$$\frac{1}{p_n} + \frac{1}{r_n} + \frac{1}{k_n} + \frac{p_n(r_n+k_n)}{n} \to 0 \;\; as\; n\to\infty. \tag{3.6}$$

Then

$$\left\| \hat{\theta}_{(n)} - \theta_{(n)} \right\| \xrightarrow{p} 0, \;\; as\; n\to\infty.$$

Condition (3.6) details the restrictions on the rate of growth of the number of instruments and regressors, and implies a restriction on the rate of growth of the parameter space because $p_n \le r_n$. Slightly weakened conditions yield the same result for the just identified case $p_n = r_n$, where $\hat{K}_n$ and K_n are square matrices so that $\hat{\theta}_{(n)} = \hat{K}_n^{-1}\hat{k}_n$ and $Q_n^{-1} = K_n^{-1}J_nK_n'^{-1}$.

Corollary 3.2. *Let $p_n = r_n$, Assumptions 1-6 hold,*

$$\underline{\lim}_{n\to\infty} \underline{\eta}(K_n'K_n) > 0, \tag{3.7}$$

and

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n(p_n+k_n)}{n} \to 0 \;\; as\; n\to\infty. \tag{3.8}$$

Then

$$\left\| \hat{\theta}_{(n)} - \theta_{(n)} \right\| \xrightarrow{p} 0, \;\; as\; n\to\infty.$$

A natural estimate of σ^2 is

$$\hat{\sigma}^2_{(n)} = \frac{1}{n}\left(y_n - (R_n, X_n)\,\hat{\theta}_{(n)}\right)'\left(y_n - (R_n, X_n)\,\hat{\theta}_{(n)}\right). \tag{3.9}$$

Assumption 9. $\overline{\lim}_{n\to\infty} \overline{\eta}(L_n) < \infty$.

Theorem 3.2. *Let Assumptions 1-7 and 9 hold, and*

$$\frac{1}{p_n} + \frac{1}{r_n} + \frac{1}{k_n} + \frac{(p_n + k_n)(r_n + k_n)}{n} \to 0 \text{ as } n \to \infty. \tag{3.10}$$

Then

$$\hat{\sigma}^2_{(n)} \xrightarrow{p} \sigma^2 \text{ as } n \to \infty.$$

A similar theorem holds in the just identified case $p_n = r_n$ but we omit the statement for brevity. Here the requirement that h_n be bounded away from zero is crucial (see (A.8)), with consistency possibly failing otherwise. We can also record a central limit theorem for finitely many arbitrary linear combinations of $\hat{\theta}_{(n)} - \theta_{(n)}$ under stronger conditions which restrict the growth of p_n and r_n relative to n further.

Theorem 3.3. *Let Assumptions 1-9 hold and*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{1}{r_n} + \frac{p_n\left(r_n^2 + k_n^2\right)}{n} + \frac{k_n(r_n + k_n)}{n} \to 0 \text{ as } n \to \infty. \tag{3.11}$$

Then, for any $s \times (p_n + k_n)$ matrix of constants Ψ_n with full row-rank,

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n\left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty} \frac{\sigma^2}{p_n + k_n} \Psi_n Q_n^{-1} \Psi_n'\right), \text{ as } n \to \infty.$$

The asymptotic covariance matrix may be consistently estimated by

$$\frac{\hat{\sigma}^2_{(n)}}{p_n + k_n} \Psi_n \hat{Q}_n^{-1} \Psi_n'.$$

Corollary 3.3. *Let $p_n = r_n$, Assumptions 1-6, 8, 9, (3.7) hold and*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n^3}{n} + \frac{p_n k_n^2}{n} \to 0 \text{ as } n \to \infty. \tag{3.12}$$

Then, for any $s \times (p_n + k_n)$ matrix of constants Ψ_n with full row-rank,

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}} \Psi_n\left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty} \frac{\sigma^2}{p_n + k_n} \Psi_n K_n^{-1} J_n K_n'^{-1} \Psi_n'\right) \text{ as } n \to \infty.$$

The asymptotic covariance matrix may be consistently estimated by

$$\frac{\hat{\sigma}^2_{(n)}}{p_n + k_n} \Psi_n \hat{K}_n^{-1} J_n \hat{K}_n^{\prime -1} \Psi_n^{\prime}.$$

Note that in Theorem 3.3 the condition $p_n r_n^2/n \to 0$ implies $p_n k_n^2/n \to 0$ so long as $k_n = \mathscr{O}(r_n)$ i.e. the number of regressors increases no faster than the number of instruments. In particular if r_n is fixed (implying that p_n is fixed), $k_n = \mathscr{O}(r_n)$ is not satisfied unless k_n is also fixed. Similarly $r_n k_n/n \to 0$ implies $k_n^2/n \to 0$ if $k_n = \mathscr{O}(r_n)$.

The $n^{\frac{1}{2}}/(p_n + k_n)^{\frac{1}{2}}$-norming is needed to ensure a finite asymptotic covariance matrix, and implies a slower than $n^{\frac{1}{2}}$ rate of convergence due to the increasing parameter space dimension, while conditions (3.11) and (3.12) restrict the growth of the parameter space. Indeed, if only $n^{\frac{1}{2}}$-norming was employed the rows of Ψ_n would have to be assumed to have uniformly bounded norm, which implies a similar normalisation as these rows have increasing dimension. The norming can change if the rows of Ψ_n contain many zero elements, indeed the number of non-zero elements can even be allowed to increase at a rate slower than the rate of increase of the parameters. In particular, Theorem 3.3 may be easily rewritten if the interest is in obtaining a central limit theorem for a fixed number of the parameters rather than an increasing number. Suppose without loss of generality that we are interested in, say, the first l elements of $\theta_{(n)}$. In this case we take Ψ_n to be a $1 \times (p_n + k_n)$ non-null row vector of constants with all elements after the l-th entry equal to zero. We then recover a $n^{\frac{1}{2}}$-consistency result which indicates that the definition of simple t-statistics does not change from the fixed-dimension model (1.1) to (2.1).

Corollary 3.4.

(i) *Let Assumptions 1-9 and (3.11) hold. Then*

$$n^{\frac{1}{2}} \left(\hat{\theta}_{(n)} - \theta_{(n)} \right)_l \xrightarrow{d} N\left(0, \sigma^2 \left(\lim_{n \to \infty} Q_n \right)_l^{-1} \right), \text{ as } n \to \infty,$$

where $\left(\hat{\theta}_{(n)} - \theta_{(n)} \right)_l$ *denotes the first* l *elements of* $\hat{\theta}_{(n)} - \theta_{(n)}$ *while the top-left* $l \times l$ *block of* $(\lim_{n\to\infty} Q_n)^{-1}$ *is denoted* $(\lim_{n\to\infty} Q_n)_l^{-1}$.

(ii) *Let* $p_n = r_n$, *Assumptions 1-6, 8, 9, (3.7) and (3.12) hold. Then*

$$n^{\frac{1}{2}} \left(\hat{\theta}_{(n)} - \theta_{(n)} \right)_l \xrightarrow{d} N\left(0, \sigma^2 \left[\left(\lim_{n \to \infty} K_n \right)^{-1} \lim_{n \to \infty} J_n \left(\lim_{n \to \infty} K_n^{\prime} \right)^{-1} \right]_l \right), \text{ as } n \to \infty,$$

where $\left[(\lim_{n\to\infty} K_n)^{-1} \lim_{n\to\infty} J_n \left(\lim_{n\to\infty} K_n'\right)^{-1}\right]_l$ *denotes the top-left* $l \times l$ *block of*

$$\left(\lim_{n\to\infty} K_n\right)^{-1} \lim_{n\to\infty} J_n \left(\lim_{n\to\infty} K_n'\right)^{-1}.$$

The asymptotic covariance matrices are estimated as in Theorem 3.3.

The setting of Case (1991, 1992) was discussed in Section 1 as a natural motivation for the work in this paper. From an applied point of view a parsimonious model may be quite desirable, and so some districts can be allowed to have the same λ_i on the basis of some homogeneity e.g. geography or demographics. There are other reasons to allow for a slower increase of the λ_i than with p. For instance, consider the condition $p_n^3/n \to 0$ (we keep k_n fixed for simplicity). In this setting this translates into requiring that $p^2/m \to 0$. For finite samples an approximation to this would be that the ratio p^2/m be small, but this may not be reasonable if, say, $p = 10$ and $m = 100$. It would be natural then to allow a slower increase of the parameter space than p, and attempts can be made to combine some λ_i to reduce the ratio p^2/m. Section 6 illustrates the behaviour of estimates in this setting.

4 Least squares estimates

Define the OLS estimate of $\theta_{(n)}$ as

$$\tilde{\theta}_{(n)} = \frac{1}{n}\hat{L}_n^{-1} \begin{bmatrix} R_n' \\ X_n' \end{bmatrix} y_n = \theta_{(n)} + \hat{L}_n^{-1} w_n, \tag{4.1}$$

where

$$\hat{L}_n = \frac{1}{n} \begin{bmatrix} R_n' \\ X_n' \end{bmatrix} [R_n, X_n], \quad w_n = \frac{1}{n} \begin{bmatrix} R_n' \\ X_n' \end{bmatrix} U_n.$$

Analogous to the IV case, we also have an asymptotic non-multicollinearity condition given by

Assumption 10. $\varliminf_{n\to\infty} \underline{\eta}(L_n) > 0.$

Theorem 4.1. *Let Assumptions 1-5, 10 hold, and*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n k_n^2 (p_n + k_n)}{n} + \frac{p_n}{h_n} \to 0 \text{ as } n \to \infty. \tag{4.2}$$

Then

$$\left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\| \xrightarrow{p} 0, \text{ as } n \to \infty.$$

Lee (2002) demonstrated consistency of least-squares estimates for the model (1.1), for $p = 1$, when $h_n \to \infty$. This condition ensures that the endogeneity problem caused by the spatially lagged y_n vanishes asymptotically. Our condition (4.2) is suitably strengthened to also account for the increasing p_n and k_n. Let C denote a generic, arbitrarily large but positive constant that is independent of sample size n. To obtain a central limit theorem, we additionally assume

Assumption 11. $\mathbb{E}\left(u_i^4\right) \leq C$ *for* $i = 1, \ldots, n$.

We first introduce the least squares residual based estimate of σ^2, defined as

$$\tilde{\sigma}^2_{(n)} = \frac{1}{n}\left(y_n - (R_n, X_n)\,\tilde{\theta}_{(n)}\right)'\left(y_n - (R_n, X_n)\,\tilde{\theta}_{(n)}\right). \tag{4.3}$$

Theorem 4.2. *Let Assumptions 1-5, 9-11 hold, and*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n k_n^2 (p_n + k_n)}{n} + \frac{p_n}{h_n} \to 0 \text{ as } n \to \infty. \tag{4.4}$$

Then

$$\tilde{\sigma}^2_{(n)} \xrightarrow{p} \sigma^2 \text{ as } n \to \infty.$$

Theorem 4.3. *Let Assumptions 1-5, 9-11 hold, and*

$$\frac{1}{p_n} + \frac{1}{k_n} + \frac{p_n^2 k_n^4 (p_n + k_n)}{n} + n^{\frac{1}{2}}\frac{p_n^{\frac{1}{2}}}{h_n} \to 0 \text{ as } n \to \infty. \tag{4.5}$$

Then, for any $s \times (p_n + k_n)$ *matrix of constants* Ψ_n *with full row-rank,*

$$\frac{n^{\frac{1}{2}}}{(p_n + k_n)^{\frac{1}{2}}}\Psi_n\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right) \xrightarrow{d} N\left(0, \lim_{n\to\infty}\frac{\sigma^2}{p_n + k_n}\Psi_n L_n^{-1}\Psi_n'\right) \text{ as } n \to \infty.$$

The asymptotic covariance matrix may be estimated consistently using

$$\frac{\tilde{\sigma}^2_{(n)}}{p_n + k_n}\Psi_n \hat{L}_n^{-1}\Psi_n'.$$

Corollary 4.1. *Let Assumptions 1-5, 9-11 and (4.5) hold. Then*

$$n^{\frac{1}{2}}\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)_l \xrightarrow{d} N\left(0, \sigma^2\left(\lim_{n\to\infty} L_n\right)_l^{-1}\right), \text{ as } n \to \infty,$$

where $\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)_l$ *denotes the first* l *elements of* $\hat{\theta}_{(n)} - \theta_{(n)}$ *while the top-left* $l \times l$ *block of* $(\lim_{n\to\infty} L_n)^{-1}$ *is denoted* $(\lim_{n\to\infty} L_n)_l^{-1}$ *and the asymptotic covariance matrix is estimated as in Theorem 4.3.*

The theory also supports enrichment of models with further data availability. Kolympiris, Kalaitzandonakes, and Miller (2011) attempt to explain the level of venture capital funding (provided by venture capital firms (VCFs)) for dedicated biotechnology firms (DBFs) with a SAR model. In particular, the hypotheses are that the level of VC funding for a DBF increases with the number of VCFs located in close proximity to the DBF and with the number of other DBFs located in close proximity to the DBF. To model this, (1.1) is employed, where the dependent variable is the natural logarithm of the amount invested by VCFs in each of the $n = 816$ observed DBFs. Weight matrices are defined using a binary neighbourhood criterion and then row-normalised. In particular, three weight matrices are employed (i.e. $p = 3$) with each based on one of 3 sequential 10-mile rings from the origin DBF. The set of DBFs situated less than 10 miles from the origin DBF are considered one set of neighbours, those situated 10.1-20 miles from the origin form the second set and the third set of neighbours is defined in the obvious way. Because the number of neighbours may be taken to increase with sample size, OLS was used. Our theory is relevant here, since if data on more DBFs were to become available it would be attractive to reduce the radius of the rings used in defining neighbours. As discussed earlier, more parsimonious specifications may still be attractive to the practitioner but various models can be employed and relevant statistical tests run to arrive at a more informed choice.

5 Illustrations

5.1 Panel data SAR models with fixed effects

Consider a balanced spatial panel data set with N observations in each of T individual panels, so that the sample size is $n = NT$. Let $y_{t,N}$ be the $N \times 1$ vector of observations on the dependent variable for the t-th panel, where t may correspond to a time period or a more general spatial unit like a school, village or district. Also let $X_{t,N}$ and F_N be $N \times k_1$ and $N \times k_2$ matrices of regressors respectively. $X_{t,N}$ contains panel-varying regressors while F_N does not. Let W_{iN}, $i = 1, \ldots, p$, be a set of spatial weight matrices and consider the model

$$y_{t,N} = l_N \alpha_t + X_{t,N}\beta + F_N \gamma_t + \sum_{i=1}^{p} \lambda_i W_{iN} y_{t,N} + U_{t,N}, \quad t = 1, \ldots, T \tag{5.1}$$

where $U_{t,N}$ is the $N \times 1$ vector of disturbances for each panel, which we take to be formed of iid components. The α_t, $t = 1, \ldots, T$, are scalar parameters that control for fixed effects with respect to panels, the λ_i, $i = 1, \ldots, p$, are scalar spatial autoregressive parameters and β is a $k_1 \times 1$ panel-invariant parameter vector. On the other hand γ_t is a $k_2 \times 1$ parameter vector that varies over panels. For this reason, the variables in F_N may be thought of as controlling for 'quasi' fixed-effects. Denote $y_n = \left(y'_{1,n}, \ldots, y'_{T,n}\right)'$, $X_n = \left(X'_{1,n}, \ldots, X'_{T,n}\right)'$, $U_n = \left(U'_{1,n}, \ldots, U'_{T,n}\right)'$, $\alpha = (\alpha_1, \ldots, \alpha_T)'$ and $\gamma = (\gamma_1, \ldots, \gamma_T)'$. We can then stack (5.1) to obtain

$$y_n = (I_T \otimes l_N)\,\alpha + X_n\beta + (I_T \otimes F_N)\,\gamma + \sum_{i=1}^{p} \lambda_i \,(I_T \otimes W_{iN})\, y_n + U_n. \tag{5.2}$$

This model is an extension of that considered in Kelejian, Prucha, and Yuzefovich (2006), and was employed by Yuzefovich (2003). We allow both $T \to \infty$ and $N \to \infty$ for our asymptotic theory. This implies that the number of regression parameters in (5.2) increases asymptotically. Not only this, since the $I_T \otimes W_{iN}$ are block diagonal it would be natural to fear that spatial autoregressive parameters differ for each panel, or at least among subsets of the panels. To illustrate, suppose for the moment that $p = 1$. Allowing a separate spatial parameter for each panel implies the model

$$y_n = (I_T \otimes l_N)\,\alpha + X_n\beta + (I_T \otimes F_N)\,\gamma + \sum_{i=1}^{T} \lambda_i W_N^i y_n + U_n \tag{5.3}$$

where

$$W_N^i = diag\left[0, \ldots, \underbrace{W_N}_{i^{th}\text{diagonal block}}, \ldots, 0\right].$$

The model (5.3) has $k_1 + T(k_2 + 1)$ regression parameters and T spatial parameters, making it fit naturally into the asymptotic regime discussed in Sections 3 and 4. As in the 'farmer-district' setting a point of concern may be that conditions such as $p_n^3 k_n^4$ diverging slower than n (needed for asymptotic normality of least squares estimation in

Section 4) translate here into requiring that

$$\frac{T^6}{N} \to 0 \text{ as } N, T \to \infty. \tag{5.4}$$

In finite samples we would like the above ratio to be somewhat small, but this may be impossible to achieve. Even for $T = 2$, $T^6 = 64$, which may not be small compared to N. A solution is to use a smaller number of spatial parameters in (5.3), consistent with the number of spatial parameters increasing more slowly with T. For example, if t represents monthly observations we may allow the spatial parameters to change on a quarterly basis so that we have $T/4$ spatial parameters, assuming that T is divisible by 4 for simplicity. Then we would need

$$\frac{T^6}{256N} \to 0 \text{ as } N, T \to \infty$$

as opposed to (5.4). The last two displayed conditions are asymptotically the same but in finite samples the last displayed ratio is smaller.

5.2 Models with circulant weight matrices

For any natural number i, define W_{in}^* as the symmetric circulant matrix with first row elements given by

$$w_{1j,in}^* = \begin{cases} 0 & \text{if } j = 1 \text{ or } j = i+2, \ldots, n-i; \\ 1 & \text{if } j = 2, \ldots, i+1 \text{ or } j = n-i+1, \ldots, n. \end{cases} \tag{5.5}$$

Thus the weight matrix W_{in}^* encapsulates a binary neighbourhood criterion for i neighbours on either 'side' of a unit. For instance, with $n = 4$ we will have

$$W_{1n}^* = \begin{bmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}.$$

Now define

$$W_{in} = \frac{1}{\|W_{in}^*\|} W_{in}^*, \tag{5.6}$$

where symmetry of W_{in}^* implies that $\|W_{in}^*\| = \overline{\eta}\left(W_{in}^*\right)$. Because W_{in}^* is a circulant matrix its eigenvalues are given by $\sum_{j=2}^{i+1} \omega_k^{j-1} + \sum_{j=n-i+1}^{n} \omega_k^{j-1}$, where $w_k = \exp(2\pi\imath(k-1)/n)$

with $\imath^2 = -1$ and $k = 1, \ldots, n$ (see e.g. Davis (1979) pg. 73). Thus $\overline{\eta}\left(W_{in}^*\right) = 2i$, implying that the normalisation in (5.6) is equivalent to row-normalisation. Then W_{in} is also a symmetric circulant matrix with first row elements given by

$$w_{1j,in} = \begin{cases} 0 & \text{if } j = 1 \text{ or } j = i+2, \ldots, n-i; \\ \frac{1}{2i} & \text{if } j = 2, \ldots, i+1 \text{ or } j = n-i+1, \ldots, n. \end{cases} \tag{5.7}$$

The maximum number of neighbours is determined by sample size. From (5.7), we must have $i < \lceil n/2 \rceil$ where, for any real number a, $\lceil a \rceil$ denotes the smallest integer greater than or equal to a. The set of weight matrices W_{in} is easily seen to be linearly independent. To prove this it is sufficient to show linear independence of the first rows. Indeed take any finite set $\{W_{in} : i = 1, \ldots, s,\ s \text{ a natural number}\}$, and suppose that $\sum_{i=1}^{s} \xi_i w_{1,in}' = 0$, where $w_{1,in}'$ denotes the first row of W_{in} and ξ_i are scalars. By (5.7) it is immediately obvious that $\xi_s = 0$. This implies $\sum_{i=1}^{s-1} \xi_i w_{1,in}' = 0$, which in turn implies $\xi_{s-1} = 0$. Proceeding in this manner we conclude that $\xi_i = 0$ for each $i = 1, \ldots, s$.

Note that OLS cannot be used to estimate (2.1) with these weight matrices, because as the number of neighbours increases with sample size this is reflected in additional weight matrices rather than more elements in existing weight matrices. It is possible to obtain consistent and asymptotically normal IV estimates, but only if i is restricted further. If this is not the case we have the model $y_n = \sum_{i=1}^{n/2} \lambda_{in} W_{in} y_n + X_n \beta_{(n)} + U_n$ for even n and $y_n = \sum_{i=1}^{(n-1)/2} \lambda_{in} W_{in} y_n + X_n \beta_{(n)} + U_n$ for odd n, implying that p_n increases like n in (2.1). Introduce a positive real valued function g defined on the set of natural numbers such that $g(n/2)$ is either bounded in n or diverges with n and consider the model

$$y_n = \sum_{i=1}^{\lceil g(n/2) \rceil - 1} \lambda_{in} W_{in} y_n + X_n \beta_{(n)} + U_n. \tag{5.8}$$

Then consistent, and asymptotically normal, IV estimates may be obtained (with k_n fixed and $p_n = r_n$ for simplicity) if $\left(g(n/2)\right)^2 / n \to 0$ and $\left(g(n/2)\right)^3 / n \to 0$ as $n \to \infty$, respectively. The function $g(m) = m^{\frac{1}{4}}$ for any natural number m, where we take the positive root, satisfies both conditions, as does $g(m) = (\log m)^{\frac{1}{3}}$. As a guideline for the former case, for $n/2 = 50$ we may use up to $i = 2$ while for $n/2 = 50000$ we can use up to $i = 14$.

6 Monte Carlo experiments

A set of Monte Carlo experiments was carried out with three aims. First, to verify the claim that more data indeed leads to substantial improvement in the precision of estimates despite increasing the dimension of the parameter space, and also to understand the rate at which this happens. Second, to check the relative performance of the OLS and IV estimates for various values of p_n^2/n. Third, to examine the rate of change of the precision of estimates as the ratio p_n^2/n changes.

We employed the spatial scenario of Case (1991, 1992), so that W_{in} given by (1.4) and (1.3) were employed. The number of regressors was kept fixed at $k_n = 2$ for simplicity, and we experimented with three values of p: 2, 6 and 18. For each value of p three different values of m were chosen: 50, 150 and 450. Note that in this setting we have $p_n = p$ and $n = pm$ so that $p_n^2/n = p/m$. The explanatory variables in X_n were generated from a uniform distribution on $(0, 1)$, and kept fixed throughout to adhere to the non-stochastic aspect of Assumption 5. The u_i were generated as iid draws from a standard normal ($\sigma^2 = 1$) distribution, and instruments were constructed as in (3.4) using only first-order spatial lags of the regressors. The vector y_n was generated using (1.1) and (1.4) in each of the 1000 replications. We chose $\beta_1 = 1$ and $\beta_2 = 0.5$ and the following values for the spatial autoregressive parameters:

$$\begin{aligned}
p = 2; \quad & \lambda_1 = 0.7; \lambda_2 = 0.8 \\
p = 6; \quad & \lambda_3 = 0.5; \lambda_4 = 0.8; \lambda_5 = 0.3; \lambda_6 = 0.6 \\
p = 18; \quad & \lambda_7 = 0.7; \lambda_8 = 0.8; \lambda_9 = 0.5; \lambda_{10} = 0.8; \lambda_{11} = 0.3; \lambda_{12} = 0.6; \\
& \lambda_{13} = 0.7; \lambda_{14} = 0.8; \lambda_{15} = 0.5; \lambda_{16} = 0.8; \lambda_{17} = 0.3; \lambda_{18} = 0.6,
\end{aligned}$$

where higher dimensional true parameter vectors inherit the initial entries from the preceding case. For our analysis of finite sample properties we employ a measure called Monte Carlo average mean-squared error (AMSE). For any s-dimensional parameter estimate this is the simple average of the Monte Carlo mean-squared errors, e.g. for the IV estimate we define it as

$$AMSE\left(\hat{\theta}_{(n)}\right) = \frac{1}{s}\sum_{i=1}^{s} MSE\left(\hat{\theta}_{in}\right),$$

with $MSE\left(\hat{\theta}_{in}\right)$ denoting the the Monte Carlo MSE for the IV estimate of the i-th element in the parameter vector. This is a compact way of comparing performance for

m		50	150	450
p				
2	$AMSE\left(\hat{\theta}_{(n)}\right)$	0.0693	0.0219	0.0076
	$AMSE\left(\tilde{\theta}_{(n)}\right)$	0.0714	0.0232	0.0076
	$\frac{AMSE(\hat{\theta}_{(n)})}{AMSE(\tilde{\theta}_{(n)})}$	0.9706	0.9450	1.0000
6	$AMSE\left(\hat{\theta}_{(n)}\right)$	0.0243	0.0071	0.0022
	$AMSE\left(\tilde{\theta}_{(n)}\right)$	0.0283	0.0074	0.0023
	$\frac{AMSE(\hat{\theta}_{(n)})}{AMSE(\tilde{\theta}_{(n)})}$	0.8594	0.9599	0.9663
18	$AMSE\left(\hat{\theta}_{(n)}\right)$	0.0115	0.0033	0.0010
	$AMSE\left(\tilde{\theta}_{(n)}\right)$	0.0184	0.0038	0.0011
	$\frac{AMSE(\hat{\theta}_{(n)})}{AMSE(\tilde{\theta}_{(n)})}$	0.6250	0.8479	0.9218

Table 6.1: Monte Carlo AMSE and RAMSE of IV and OLS estimates

high-dimensional parameters.

Table 6.1 presents AMSE and relative AMSE (RAMSE) of the IV and OLS estimates, the latter defined as the ratio $\mathrm{AMSE}\left(\hat{\theta}_{(n)}\right)/\mathrm{AMSE}\left(\tilde{\theta}_{(n)}\right)$, for various combinations of p and m. The AMSE for both estimators reduces for fixed p and increasing m. Interestingly this happens also for fixed m and increasing p, indicating that the claim of being able to precisely estimate increasingly many parameters is reasonable. The RAMSE indicate that even though the OLS estimator is asymptotically efficient the IV estimator outperforms it in many cases. Additional simulations (not reported here) indicate that the OLS estimate tends to perform better for larger values of m. For instance, when $p = 2$ and $m = 3000$ the RAMSE is 1.0301. Recall that in this setting $h_n = m - 1$, so that this observation supports the theory that $h_n \to \infty$ is required for the consistency

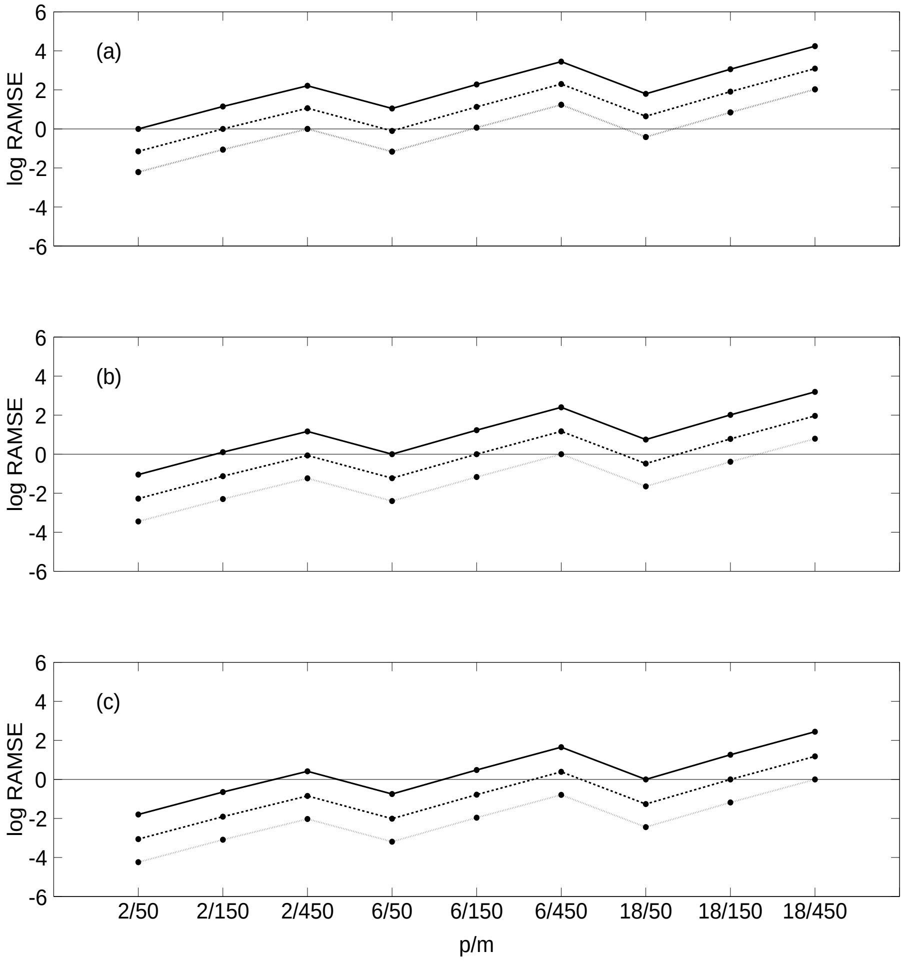

Figure 6.1: Monte Carlo RAMSE for IV estimates $\hat{\theta}_{(n)}$

of the OLS estimate with a fast rate of divergence required for asymptotic normality, while the properties of IV estimate do not depend on the divergence of h_n.

Figures 6.1(a)-(c) and 6.2(a)-(c) plot the RAMSE for $\hat{\theta}_{(n)}$ and $\tilde{\theta}_{(n)}$ for all possible ratios p/m. Figure 6.1(a) illustrates RAMSE for the IV estimate $\hat{\theta}_{(n)}$ when $p = 2$. The points on the horizontal axis correspond to all choices of p/m and the vertical axis measures RAMSE on a logarithmic scale, but for simplicity we will just refer to the vertical values as RAMSE. Each solid dot on the figure marks out the RAMSE of $\hat{\theta}_{(n)}$ for a particular value of m relative to that of $\hat{\theta}_{(n)}$ for the corresponding ratio on the horizontal axis. The solid line joins the RAMSE for $m = 50$, the heavy dotted line for $m = 150$ and the light dotted line for $m = 450$. Figures 6.1(b), (c) repeat the analysis for

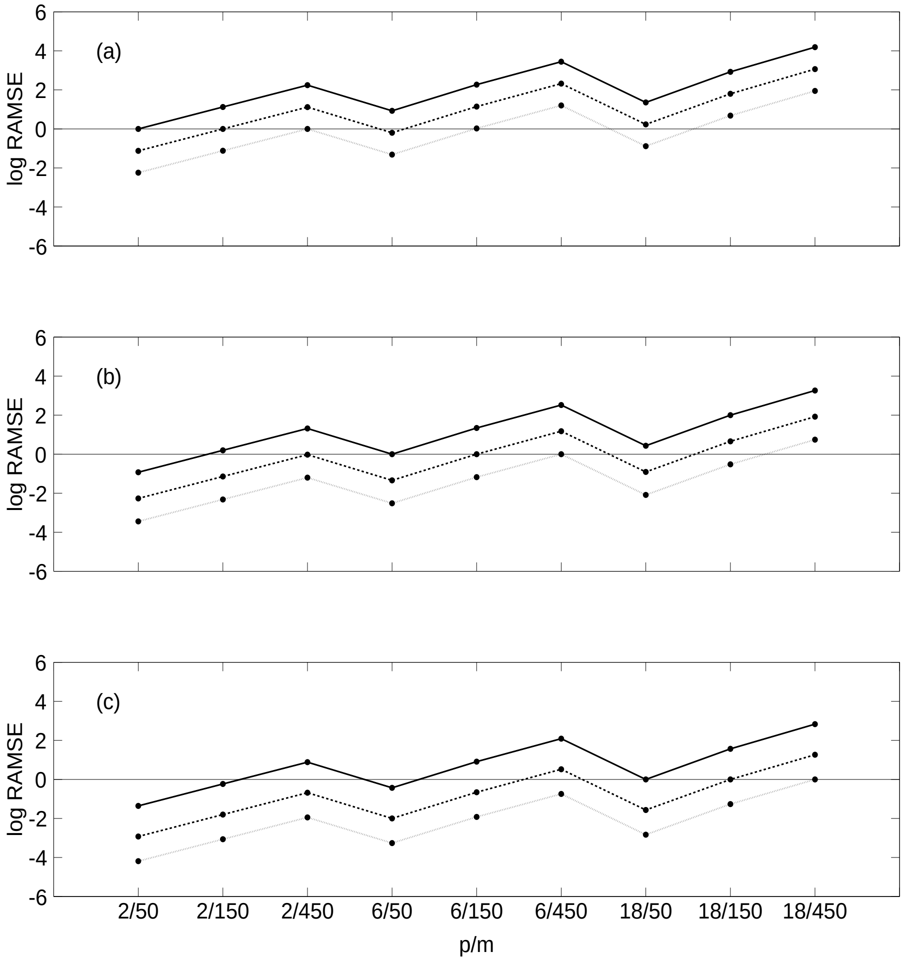

Figure 6.2: Monte Carlo RAMSE for OLS estimates $\tilde{\theta}_{(n)}$

$p = 6$ and $p = 18$ respectively, again the solid lines joining the RAMSE for $m = 50$, the heavy dotted lines for $m = 150$ and the light dotted lines for $m = 450$. Figures 6.2(a)-(c) do exactly the same for the OLS estimate $\tilde{\theta}_{(n)}$. Because the vertical axes in each case measure the logarithm of RAMSE, the RAMSE of an estimate with itself corresponds to a value of zero. Negative RAMSE values indicate that the estimate for which a line is plotted outperforms the comparator indicated on the horizontal axis, while the opposite holds true for positive values. For ease of interpretation a thin solid horizontal line is passed through zero.

We now analyse Figure 6.1(a) in detail. The solid line increases through RAMSE comparisons for estimates with $p = 2$ but higher values of m. While the value of RAMSE then dips for $p = 6$ and $m = 50$, it is still positive. This indicates that 8 parameters are more accurately estimated with 300 observations as opposed to 4 parameters with 100 observations. The RAMSE then increase in the $p = 6$ range, before dropping and then increasing again in the $p = 18$ range. This indicates two findings, reflective of the discussion of Table 6.1 above. First, estimates become more precise for given p with increasing m. Second, many more parameters can be estimated much more accurately with increasing data. The first point implies that in practice, when a dataset of a given size is available, researchers should try to reduce the number of parameters to obtain more precise estimates. This may be achieved in a variety of ways, as discussed earlier. The heavy dotted ($m = 150$) and light dotted ($m = 450$) line behave in the same way, the difference being that the initial values of RAMSE are negative. This is because, for instance, the first point on the heavy dotted line corresponds to the log RAMSE of the estimate with $p = 2$ and $m = 150$ to the estimate with $p = 2$ and $m = 50$. This is simply the log of the reciprocal of the value whose log is the second point on the solid line. The analysis of all the other figures indicates exactly the same pattern in the rate of change of the RAMSE.

7 Conclusion

We have presented asymptotic theory for a general class of spatial autoregressive models with increasingly many parameters, and with non-intercept regressors. The theory is thus not applicable to the pure SAR model, though analogous theory for this can be developed using other parameter estimates. Another natural question that arises as a result of this work is that of testing for the equality of spatial parameters. The implication of failing to reject equality is a more parsimonious model, potentially one of finite dimension. These two questions are addressed in Gupta (2013), which also provides an empirical example illustrating directions that the applied literature can follow.

Appendices

We denote $a_n = p_n + k_n$, $b_n = r_n + k_n$, $c_n = p_n k_n^2 + k_n$ and $\tau_n = n^{\frac{1}{2}}/a_n^{\frac{1}{2}}$ to conserve space.

A Proofs of results in Sections 3 and 4

Proof of Proposition 2.1: Since $S_n(\lambda_{(n)})$ is block-diagonal, invertibility can be proved by showing that each block is invertible. Let B_{in} denote the i^{th} diagonal block in W_{in}, i.e. this is the only non-zero block in W_{in}. Then $S_n(\lambda_{(n)}) = I_n - diag(\lambda_{1n}B_{1n}, ..., \lambda_{p_n n}B_{pn})$.

By the normalization of diagonal elements of each W_{in} in Assumption 2, the diagonal elements of $S_n(\lambda_{(n)})$ are 1. Consider the i^{th} block in $S_n(\lambda)$. Then

$$\sum_{s \neq r} |\lambda_{in}||w_{rs,in}| \sum_{m \neq l} |\lambda_{in}||w_{lm,in}| < 1$$

if $\lambda_{in}^2 < 1$, due to absolute row-sum norms being uniformly bounded by 1. The claim follows from Horn and Johnson (1985), page 381, Corollary 6.4.11 (b)). □

Proof of Lemma 3.1.

(i) For any $a_n \times 1$ vector x_n satisfying $\|x_n\| = 1$,

$$x_n' K_n' J_n^{-1} K_n x_n \geq \underline{\eta}\left(J_n^{-1}\right) x_n' K_n' K_n x_n \geq \frac{\underline{\eta}\left(K_n' K_n\right)}{\overline{\eta}\left(J_n\right)} \geq c,$$

for large enough n by Assumption 7, where c denotes a positive but arbitrarily small real number that does not depend on n.

(ii) Similar.

□

Proof of Theorem 3.1. Write

$$\begin{aligned}
\hat{\theta}_{(n)} - \theta_{(n)} &= \left(\hat{Q}_n^{-1} - Q_n^{-1}\right) \hat{K}_n' J_n^{-1} q_n + Q_n^{-1} \hat{K}_n' J_n^{-1} q_n \\
&= Q_n^{-1}\left(Q_n - \hat{Q}_n\right) \hat{Q}_n^{-1} \hat{K}_n' J_n^{-1} q_n + Q_n^{-1}\left(\hat{K}_n - K_n\right)' J_n^{-1} q_n \\
&+ Q_n^{-1} K_n' J_n^{-1} q_n \\
&= Q_n^{-1}\left(Q_n - \hat{Q}_n\right)\left(\hat{\theta}_{(n)} - \theta_{(n)}\right) + Q_n^{-1}\left(\hat{K}_n - K_n\right)' J_n^{-1} q_n
\end{aligned}$$

$$+ \quad Q_n^{-1} K_n' J_n^{-1} q_n. \tag{A.1}$$

By elementary norm inequalities

$$\left\| \hat{Q}_n - Q_n \right\| \leq \left\| \hat{K}_n - K_n \right\| \, \| J_n^{-1} \| \left(\left\| \hat{K}_n - K_n \right\| + 2 \|K_n\| \right), \tag{A.2}$$

where $\mathbb{E} \left\| \hat{K}_n - K_n \right\|^2$ is bounded by

$$\frac{\sigma^2}{n^2} \sum_{i=1}^{b_n} \sum_{j=1}^{p_n} \left| p_{in}' G_{jn} G_{jn}' p_{in} \right| \leq \frac{\sigma^2}{n^2} \sum_{i=1}^{b_n} \|p_{in}\|^2 \sum_{j=1}^{p_n} \|G_{jn}\|^2 \leq C \frac{p_n b_n}{n}$$

by Assumptions 5 and 6 and Lemma B.1, denoting by p_{in} the $i-th$ column of (Z_n, X_n). We conclude that

$$\left\| \hat{K}_n - K_n \right\| = \mathscr{O}_p \left(\frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}} \right) \tag{A.3}$$

by Markov's inequality. Then

$$\left\| \hat{Q}_n - Q_n \right\| = \mathscr{O}_p \left(\max \left\{ \frac{p_n b_n}{n}, \frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}} \right\} \right) = \mathscr{O}_p \left(\frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}} \right), \tag{A.4}$$

by Assumption (8) because $\| J_n^{-1} \| = \left(\underline{\eta} \left(J_n \right) \right)^{-1}$ and $\|K_n\|^2 = \overline{\eta} \left(K_n' K_n \right)$. Likewise

$$\mathbb{E} \|q_n\|^2 = E \left\| \frac{1}{n} \sum_{i=1}^{n} a_{in} u_i \right\|^2 = \frac{\sigma^2}{n^2} \sum_{i=1}^{n} \|a_{in}\|^2 = \mathscr{O} \left(\frac{b_n}{n} \right),$$

where a_{in}' is the $i-th$ row of (Z_n, X_n), since the elements of a_{in}' are uniformly bounded by Assumptions 5 and 6. By Markov's inequality

$$\|q_n\| = \mathscr{O}_p \left(\frac{b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}} \right). \tag{A.5}$$

From (A.1),

$$\begin{aligned} \left(1 - \|Q_n^{-1}\| \left\| \hat{Q}_n - Q_n \right\| \right) \left\| \hat{\theta}_{(n)} - \theta_{(n)} \right\| \quad &\leq \quad \|Q_n^{-1}\| \left\| \hat{K}_n - K_n \right\| \|J_n^{-1}\| \, \|q_n\| \\ &+ \quad \|Q_n^{-1}\| \, \|K_n\| \, \|J_n^{-1}\| \, \|q_n\| \, . \end{aligned} \tag{A.6}$$

By (A.4) the first factor on the LHS converges in probability to one by (3.6) and Lemma 3.1 (i), and because $b_n = r_n + k_n$. This also ensures that the first factor in the first term on the RHS of (A.6) is bounded, as well as the third factor by Assumption 8. The second and fourth factors have orders given in (A.3) and (A.5) respectively, implying that the first term is $\mathscr{O}_p\left(p_n^{\frac{1}{2}} b_n/n\right)$. The order of the second term on the RHS is determined similarly to be $\mathscr{O}_p\left(b_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right)$ so that

$$\left\|\hat{\theta}_{(n)} - \theta_{(n)}\right\| = \mathscr{O}_p\left(\max\left\{\frac{p_n^{\frac{1}{2}} b_n}{n}, \frac{b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right\}\right) = \mathscr{O}_p\left(\frac{b_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right). \tag{A.7}$$

This is negligible by Assumption 3.6. The proof of Corollary 3.2 is similar. □

Proof of Theorem 3.2. Write

$$\begin{aligned}
\hat{\sigma}^2_{(n)} &= \frac{1}{n}\left(U_n - (R_n, X_n)\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)\right)'\left(U_n - (R_n, X_n)\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)\right) \\
&= \frac{1}{n}U_n'U_n - \frac{2}{n}\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)'\begin{bmatrix} R_n' \\ X_n' \end{bmatrix} U_n \\
&\quad + \frac{1}{n}\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)'\begin{bmatrix} R_n' \\ X_n' \end{bmatrix}[R_n, X_n]\left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \\
&= \frac{1}{n}U_n'U_n - 2\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)' w_n + \left(\hat{\theta}_{(n)} - \theta_{(n)}\right)'\hat{L}_n\left(\hat{\theta}_{(n)} - \theta_{(n)}\right).
\end{aligned}$$

From Assumption 1, $U_n'U_n/n = \sigma^2 + o_p(1)$. Also by (A.7) and (A.16) the modulus of the second term is bounded by

$$\left\|\hat{\theta}_{(n)} - \theta_{(n)}\right\| \|w_n\| = \mathscr{O}_p\left(\max\left\{\frac{b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}\right\}\right)$$

while the third term has modulus bounded by

$$\begin{aligned}
&\left|\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)' L_n\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)\right| \\
&\quad + \left|\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)'\left(\hat{L}_n - L_n\right)\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)\right| \\
\leq &\left\|\hat{\theta}_{(n)} - \theta_{(n)}\right\|^2 \|L_n\| + \left\|\hat{\theta}_{(n)} - \theta_{(n)}\right\|^2 \left\|\hat{L}_n - L_n\right\|
\end{aligned}$$

$$= \mathscr{O}_p\left(\max\left\{\frac{b_n}{n}, \frac{p_n^{\frac{1}{2}} k_n a_n^{\frac{1}{2}} b_n}{n^{\frac{3}{2}}}, \frac{p_n b_n}{n h_n}\right\}\right)$$
$$= \mathscr{O}_p\left(\max\left\{\frac{b_n}{n}, \frac{p_n b_n}{n h_n}\right\}\right)$$

using (A.7), (A.20) and Assumption 9. Thus, noting that $p_n b_n / n h_n$ and $p_n^{\frac{1}{2}} k_n a_n^{\frac{1}{2}} b_n n^{-\frac{3}{2}}$ are dominated by $p_n^{\frac{1}{2}} b_n^{\frac{1}{2}} n^{-\frac{1}{2}} h_n^{-1}$ under (3.10), we have

$$\hat{\sigma}^2_{(n)} - \sigma^2 = \mathscr{O}_p\left(\max\left\{\frac{b_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n}, \frac{p_n^{\frac{1}{2}} b_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}, \frac{b_n}{n}\right\}\right), \tag{A.8}$$

which is negligible by (3.10) and because h_n is bounded away from zero, noting that $b_n c_n \le C\left(p_n r_n k_n^2 + p_n k_n^3\right)$. □

Proof of Theorem 3.3. Let α be any non-null $s \times 1$ vector of constants and write

$$\begin{aligned} \tau_n \alpha' \Psi_n \left(\hat{\theta}_{(n)} - \theta_{(n)}\right) &= \tau_n \alpha' \Psi_n Q_n^{-1}\left(\hat{Q}_n - Q_n\right)\left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \\ &+ \tau_n \alpha' \Psi_n Q_n^{-1}\left(\hat{K}_n - K_n\right)' J_n^{-1} q_n \\ &+ \tau_n \alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} q_n. \end{aligned} \tag{A.9}$$

We first show that first term on the RHS of (A.9) is negligible in probability. It has modulus bounded by

$$\tau_n \|\alpha\| \|\Psi_n\| \left\|\hat{\theta}_{(n)} - \theta_{(n)}\right\| \|Q_n^{-1}\| \left\|\hat{Q}_n - Q_n\right\| = \mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}} b_n}{n^{\frac{1}{2}}}\right),$$

from (A.4), (A.7) and Assumption 8. This is negligible by (3.11) because $b_n^2 \le 2\left(r_n^2 + k_n^2\right)$. Similarly the second term on the right side of (A.9) is bounded in absolute value by

$$\tau_n \|\alpha\| \|\Psi_n\| \|Q_n^{-1}\| \left\|\hat{K}_n - K_n\right\| \|q_n\| = \mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}} b_n}{n^{\frac{1}{2}}}\right)$$

so we have to prove asymptotic normality only for the third term on the RHS of (A.9).

Now

$$\tau_n \alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} q_n = \frac{1}{n^{\frac{1}{2}} a_n^{\frac{1}{2}}} \sum_{i=1}^{n} \alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in} u_i$$

has mean zero and variance

$$\frac{\sigma^2}{n a_n} \sum_{i=1}^{n} (\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in})^2.$$

Thus consider

$$\frac{n \alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} q_n}{\sigma \left\{ \sum_{i=1}^{n} (\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in})^2 \right\}^{\frac{1}{2}}} = \sum_{i=1}^{n} c_{in} u_i,$$

where

$$c_{in} = \frac{\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in}}{\sigma \left\{ \sum_{i=1}^{n} (\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in})^2 \right\}^{\frac{1}{2}}}.$$

We now verify the Lindeberg condition for $c_{in} u_i$. With $\mathbf{1}(\cdot)$ denoting indicator function,

$$\sum_{i=1}^{n} \mathbb{E} \left\{ (c_{in} u_i)^2 \mathbf{1}(|c_{in} u_i| > \epsilon) \right\} \leq \max_{1 \leq i \leq n} \mathbb{E} \left\{ u_i^2 \mathbf{1} \left(u_i^2 > \frac{\epsilon^2}{\max_{1 \leq i \leq n} c_{in}^2} \right) \right\} \sum_{i=1}^{n} c_{in}^2$$

Since the u_i^2 are uniformly integrable, it suffices to show that $\max_{1 \leq i \leq n} c_{in}^2 \to 0$ as $n \to \infty$, as the last factor on the RHS of the above displayed inequality equals $1/\sigma^2$. Consider

$$\begin{aligned} \max_{1 \leq i \leq n} c_{in}^2 &= \max_{1 \leq i \leq n} \frac{(\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in})^2}{\sigma^2 \sum_{i=1}^{n} (\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in})^2} \\ &\leq \frac{\left\| Q_n^{-1} K_n' J_n^{-1} \right\|^2 \left\| \Psi_n' \alpha \right\|^2 \max_{1 \leq i \leq n} \left\| a_{in} \right\|^2}{\sigma^2 \sum_{i=1}^{n} (\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} a_{in})^2}. \qquad \text{(A.10)} \end{aligned}$$

The denominator of (A.10) equals σ^2 times

$$\alpha' \Psi_n Q_n^{-1} K_n' J_n^{-1} \sum_{i=1}^{n} a_{in} a_{in}' J_n^{-1} K_n Q_n^{-1} \Psi_n' \alpha$$

$$
\begin{aligned}
&\geq \left\|\Psi_n'\alpha\right\|^2 \underline{\eta}\left(Q_n^{-1}K_n'J_n^{-1}\sum_{i=1}^{n}a_{in}a_{in}'J_n^{-1}K_nQ_n^{-1}\right) \\
&= n\left\|\Psi_n'\alpha\right\|^2\ \underline{\eta}\left(Q_n^{-1}K_n'J_n^{-1}K_nQ_n^{-1}\right) \\
&= n\left\|\Psi_n'\alpha\right\|^2\ \underline{\eta}\left(Q_n^{-1}\right) \\
&\geq nc\left\|\Psi_n'\alpha\right\|^2
\end{aligned}
$$

for sufficiently large n by Lemma 3.1 (ii), noting that $\sum_{i=1}^{n} a_{in}a_{in}' = nJ_n$, so (A.10) is $\mathscr{O}\left(b_n/n\right)$ by Assumptions 5 and 6, which is negligible by (3.6). The Lindeberg condition is then satisfied.

The asymptotic covariance matrix exists, and is positive definite, by Lemma 3.1. The proof of the consistency of the covariance matrix estimate is omitted, while the proof of Corollary 3.3 is similar. For the latter the existence and positive definiteness of the asymptotic covariance matrix is guaranteed by Assumptions 7 and 8. □

Proof of Theorem 4.1. We can write

$$
\begin{aligned}
\tilde{\theta}_{(n)} - \theta_{(n)} &= \left(\hat{L}_n^{-1} - L_n^{-1}\right)w_n + L_n^{-1}w_n \\
&= L_n^{-1}\left(\hat{L}_n - L_n\right)\hat{L}_n^{-1}w_n + L_n^{-1}w_n \\
&= L_n^{-1}\left(\hat{L}_n - L_n\right)\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right) + L_n^{-1}w_n. \qquad \text{(A.11)}
\end{aligned}
$$

It is clear that

$$
\|w_n\| \leq \left\|\frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix} U_n\right\| + \left\|\frac{1}{n}\begin{bmatrix} B_n' \\ 0 \end{bmatrix} U_n\right\|. \qquad \text{(A.12)}
$$

Now

$$
\mathbb{E}\left\|\frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix} U_n\right\|^2 = \mathscr{O}\left(\frac{c_n}{n}\right), \qquad \text{(A.13)}
$$

as in the proof Theorem 3.1, since the elements of A_n are uniformly $\mathscr{O}\left(k_n\right)$ (Lemma B.4). Under Assumption 11, the square of the second term on the RHS of (A.12) has expectation

$$
\frac{1}{n^2}\sum_{i=1}^{p_n}\mathbb{E}\left(U_n'G_{in}'U_n\right)^2, \qquad \text{(A.14)}
$$

which, by the iid property of the u_i, equals $\sum_{i=1}^{4} \Delta_{in}$ where

$$\begin{aligned}
\Delta_{1n} &= \frac{\mu_4}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{n} g_{jj,in}^2 = \mathscr{O}\left(\frac{p_n}{nh_n^2}\right)\\
\Delta_{2n} &= \frac{\sigma^4}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{jj,in}g_{kk,in} = \mathscr{O}\left(\frac{p_n}{h_n^2}\right)\\
\Delta_{3n} &= \frac{\sigma^4}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{jk,in}g_{kj,in} = \mathscr{O}\left(\frac{p_n}{h_n^2}\right)\\
\Delta_{4n} &= \frac{\sigma^4}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{n}\sum_{k=1}^{n} g_{jk,in}^2 = \mathscr{O}\left(\frac{p_n}{h_n^2}\right),
\end{aligned}$$

by Lemma B.2, where $g_{rs,in}$ denotes the (r,s)-th element of G_{in} and denoting $\mathbb{E}\left(u_i^4\right) = \mu_4$. Hence

$$\left\| \frac{1}{n}\begin{bmatrix} B_n' \\ 0 \end{bmatrix} U_n \right\| = \mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}}{h_n}\right) \tag{A.15}$$

so that

$$\|w_n\| = \mathscr{O}_p\left\{\max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n}\right)\right\}. \tag{A.16}$$

However, under Assumption 1 we have

$$\left\| \frac{1}{n}\begin{bmatrix} B_n' \\ 0 \end{bmatrix} U_n \right\| \le \frac{1}{n}\left\|[B_n, 0]\right\|\left\|U_n\right\| = \mathscr{O}_p\left(\frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}}\right) \tag{A.17}$$

by calculations used for bounding the first term on the RHS of (A.19) and so

$$\|w_n\| = \mathscr{O}_p\left\{\max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}}\right)\right\}. \tag{A.18}$$

Also

$$\begin{aligned}
\hat{L}_n - L_n &= \frac{1}{n}\begin{bmatrix} R_n' \\ X_n' \end{bmatrix}[R_n, X_n] - \frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix}[A_n, X_n]\\
&= \frac{1}{n}\left(\begin{bmatrix} R_n' \\ X_n' \end{bmatrix} - \begin{bmatrix} A_n' \\ X_n' \end{bmatrix}\right)\left([R_n, X_n] - [A_n, X_n]\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \quad \frac{1}{n}\left[\begin{array}{c} A_n' \\ X_n' \end{array}\right]([R_n, X_n] - [A_n, X_n]) \\
&+ \quad \frac{1}{n}\left(\left[\begin{array}{c} R_n' \\ X_n' \end{array}\right] - \left[\begin{array}{c} A_n' \\ X_n' \end{array}\right]\right)[A_n, X_n] \\
&= \quad \frac{1}{n}\left[\begin{array}{c} B_n' \\ 0 \end{array}\right][B_n, 0] + \frac{1}{n}\left[\begin{array}{c} A_n' \\ X_n' \end{array}\right][B_n, 0] + \frac{1}{n}\left[\begin{array}{c} B_n' \\ 0 \end{array}\right][A_n, X_n]
\end{aligned}
$$

so we have

$$
\left\| \hat{L}_n - L_n \right\| \leq \frac{1}{n}\left\|[B_n, 0]\right\|^2 + \frac{2}{n}\left\| \left[\begin{array}{c} A_n' \\ X_n' \end{array}\right][B_n, 0] \right\|. \tag{A.19}
$$

The first term on the RHS has expectation bounded by

$$
\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{p_n}\mathbb{E}\left(e_{i,n}' G_{jn} U_n\right)^2 &= \frac{1}{n}\sum_{j=1}^{p_n}\mathbb{E}\left(U_n' G_{jn}\sum_{i=1}^{n} e_{i,n} e_{i,n}' G_{jn}' U_n\right) \\
&\leq \frac{\sigma^2}{n}\sum_{j=1}^{p_n} tr\left(G_{jn} G_{jn}'\right) \leq C\frac{p_n}{h_n},
\end{aligned}
$$

using Lemmas B.2 and B.3. For the second term in (A.19) note that

$$
\left\| \frac{1}{n}\left[\begin{array}{c} A_n' \\ X_n' \end{array}\right][B_n, 0] \right\|^2 \leq \frac{1}{n^2}\sum_{i=1}^{a_n}\sum_{j=1}^{p_n} h_{in}' G_{jn} U_n U_n' G_{jn}' h_{in},
$$

where h_{in} is the $i-th$ column of (A_n, X_n). Then by Lemma B.4, Assumption 5 and Lemma B.1 we have

$$
\begin{aligned}
\mathbb{E}\left(\frac{1}{n^2}\sum_{i=1}^{a_n}\sum_{j=1}^{p_n} h_{in}' G_{jn} U_n U_n' G_{jn}' h_{in}\right) &= \frac{\sigma^2}{n^2}\sum_{i=1}^{a_n}\sum_{j=1}^{p_n} h_{in}' G_{jn}' G_{jn}' h_{in} \\
&\leq \frac{\sigma^2}{n^2}\sum_{i=1}^{a_n}\left\|h_{in}\right\|^2 \sum_{j=1}^{p_n}\left\|G_{jn}\right\|^2 \\
&\leq C\frac{p_n k_n^2 a_n}{n}
\end{aligned}
$$

so that

$$
\left\| \hat{L}_n - L_n \right\| = \mathscr{O}_p\left(\max\left\{\frac{p_n}{h_n}, \frac{p_n^{\frac{1}{2}} k_n a_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}\right\}\right). \tag{A.20}
$$

Note that the bound derived above required only second order moments for the u_i and using fourth order moments (Assumption 11) will not improve the bound because

$$\frac{1}{n^2}\mathbb{E}\left\|\begin{bmatrix} B_n' \\ 0 \end{bmatrix}[B_n, 0]\right\|^2 \leq \frac{1}{n^2}\sum_{i=1}^{p_n}\sum_{j=1}^{p_n}\mathbb{E}\left(U_n' G_{in}' G_{jn} U_n\right)^2$$

which is $\mathscr{O}\left(p_n^2/h_n^2\right)$ in exactly the same way as we bounded (A.14) since the elements of $G_{in}' G_{jn}$ are $\mathscr{O}\left(h_n^{-1}\right)$ uniformly in i, j and n by Lemma B.3. From (A.11),

$$\left(1 - \left\|L_n^{-1}\right\| \left\|\hat{L}_n - L_n\right\|\right)\left\|\tilde{\theta}_{(n)} - \theta_{(n)}\right\| \leq \left\|L_n^{-1}\right\| \left\|w_n\right\|. \tag{A.21}$$

By (A.20) the first factor on the LHS above converges in probability to one by (4.2) and Assumption 10, the last being useful since $\left\|L_n^{-1}\right\| = (\underline{\eta}(L_n))^{-1}$. Again, the first factor on the RHS of (A.21) is bounded by Assumption 10 for sufficiently large n and so we have

$$\left\|\tilde{\theta}_{(n)} - \theta_{(n)}\right\| = \mathscr{O}_p\left\{\max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n}\right)\right\} \tag{A.22}$$

by (A.16) under Assumptions 1 and 11 but

$$\left\|\tilde{\theta}_{(n)} - \theta_{(n)}\right\| = \mathscr{O}_p\left\{\max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n^{\frac{1}{2}}}\right)\right\} \tag{A.23}$$

by (A.18) under Assumption 1 only. These are both negligible by (4.2). □

Proof of Theorem 4.2. As in the IV case, we write

$$\tilde{\sigma}_{(n)}^2 = \frac{1}{n}U_n' U_n - 2\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)' w_n + \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)' \hat{L}_n \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right).$$

From (A.22) and (A.16) the second term has modulus bounded by

$$\begin{aligned}
\left\|\tilde{\theta}_{(n)} - \theta_{(n)}\right\| \left\|w_n\right\| &= \mathscr{O}_p\left\{\max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n}\right)\right\}\mathscr{O}_p\left\{\max\left(\frac{c_n^{\frac{1}{2}}}{n^{\frac{1}{2}}}, \frac{p_n^{\frac{1}{2}}}{h_n}\right)\right\} \\
&= \mathscr{O}_p\left(\max\left\{\frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n}\right\}\right)
\end{aligned}$$

while the modulus of the third term is bounded by

$$\begin{aligned}
& \left| \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)' \left(\hat{L}_n - L_n\right) \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right) \right| \\
& + \left| \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)' L_n \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right) \right| \\
\leq & \left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\|^2 \left\| \hat{L}_n - L_n \right\| + \left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\|^2 \|L_n\| \\
= & \mathscr{O}_p \left(\max\left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right) \mathscr{O}_p \left(\max\left\{ \frac{p_n}{h_n}, \frac{k_n p_n^{\frac{1}{2}} a_n^{\frac{1}{2}}}{n^{\frac{1}{2}}} \right\} \right) \\
& + \mathscr{O}_p \left(\max\left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right) \\
= & \mathscr{O}_p \left(\max\left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right),
\end{aligned}$$

using (A.22), (A.20) and Assumption 9. We conclude that

$$\tilde{\sigma}_{(n)}^2 - \sigma^2 = \mathscr{O}_p \left(\max\left\{ \frac{c_n}{n}, \frac{p_n}{h_n^2}, \frac{p_n^{\frac{1}{2}} c_n^{\frac{1}{2}}}{n^{\frac{1}{2}} h_n} \right\} \right). \tag{A.24}$$

This is negligible by (4.4). □

Proof of Theorem 4.3. With α any non-null $s \times 1$ vector, write

$$\tau_n \alpha' \Psi_n \left(\tilde{\theta}_{(n)} - \theta_{(n)}\right) = \tau_n \alpha' \Psi_n \left(\hat{L}_n^{-1} - L_n^{-1}\right) w_n + \tau_n \alpha' \Psi_n L_n^{-1} w_n. \tag{A.25}$$

We first show that first term on the RHS of (A.25) is negligible in probability. This term has modulus bounded by τ_n times

$$\|\alpha\| \|\Psi_n\| \left\| \hat{L}_n^{-1} w_n \right\| \left\| L_n^{-1} \right\| \left\| \hat{L}_n - L_n \right\| = \|\alpha\| \|\Psi_n\| \left\| \tilde{\theta}_{(n)} - \theta_{(n)} \right\| \left\| L_n^{-1} \right\| \left\| \hat{L}_n - L_n \right\|.$$

The second factor on the RHS is $\mathscr{O}\left(a_n^{\frac{1}{2}}\right)$, the third is $\mathscr{O}_p\left(\max\left\{c_n^{\frac{1}{2}}/n^{\frac{1}{2}},\ p_n^{\frac{1}{2}}/h_n\right\}\right)$ by (A.22), the fourth is bounded for sufficiently large n by Assumption 10 and the fifth is $\mathscr{O}_p\left(\max\left\{p_n/h_n, p_n^{\frac{1}{2}} k_n a_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right\}\right)$ by (A.20). The total order of the first term on the RHS of (A.25) is the order of the last displayed expression times τ_n, which is

$\mathscr{O}_p\left(\max\left\{p_n^{\frac{1}{2}}k_n a_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n^{\frac{1}{2}}, p_n c_n^{\frac{1}{2}}/h_n, p_n k_n a_n^{\frac{1}{2}}/h_n, n^{\frac{1}{2}}p_n^{\frac{3}{2}}/h_n^2\right\}\right)$, all of which are negligible by (4.5) because

$$\begin{aligned} p_n k_n^2 a_n c_n &\leq C\left(p_n^3 k_n^4 + p_n^2 k_n^5\right)\ ,\ p_n^2 c_n \leq C p_n^3 k_n^2\ ,\ n p_n^3 = n^2 p_n^2 \frac{p_n}{n} \\ p_n^2 k_n^2 a_n &= p_n^3 k_n^2 + p_n^2 k_n^3 = n p_n\left(\frac{p_n^2 k_n^2}{n} + \frac{p_n k_n^3}{n}\right). \end{aligned}$$

The second term on the RHS of (A.25) is

$$\tau_n \alpha' \Psi_n L_n^{-1} w_n = \tau_n \alpha' \Psi_n L_n^{-1} \frac{1}{n}\begin{bmatrix} A_n' \\ X_n' \end{bmatrix} U_n + \tau_n \alpha' \Psi_n L_n^{-1} \frac{1}{n}\begin{bmatrix} B_n' \\ 0 \end{bmatrix} U_n. \tag{A.26}$$

The modulus of the second term on the RHS of (A.26) is bounded by τ_n times

$$\|\alpha\|\,\|\Psi_n\|\ \left\|L_n^{-1}\right\|\ \left\|\frac{1}{n}\begin{bmatrix} B_n' \\ 0 \end{bmatrix} U_n\right\|. \tag{A.27}$$

The second factor on the RHS above is $\mathscr{O}\left(a_n^{\frac{1}{2}}\right)$, the third is bounded for sufficiently large n by Assumption 10, and the fourth is $\mathscr{O}_p\left(p_n^{\frac{1}{2}}/h_n\right)$ by (A.15). Therefore (A.27) is $\mathscr{O}_p\left(p_n^{\frac{1}{2}}a_n^{\frac{1}{2}}/h_n\right)$ and so the modulus of the second term on the RHS of (A.26) is $\mathscr{O}_p\left(n^{\frac{1}{2}}p_n^{\frac{1}{2}}/h_n\right)$. Under (4.5) this is negligible in probability and so we need to compute only the asymptotic distribution of the first term in (A.26). Now

$$\tau_n \alpha' \Psi_n L_n^{-1} t_n = \frac{1}{n^{\frac{1}{2}}a_n^{\frac{1}{2}}} \sum_{i=1}^n \alpha' \Psi_n L_n^{-1} t_{in} u_i$$

has mean zero and variance

$$\frac{\sigma^2}{n a_n} \sum_{i=1}^n (\alpha' \Psi_n L_n^{-1} t_{in})^2,$$

where t_{in}' is the $i-th$ row of (A_n, X_n). Thus consider

$$\frac{n \alpha' \Psi_n L_n^{-1} t_n}{\sigma\left\{\sum_{i=1}^n (\alpha' \Psi_n L_n^{-1} t_{in})^2\right\}^{\frac{1}{2}}} = \sum_{i=1}^n f_{in} u_i,$$

where

$$f_{in} = \frac{\alpha'\Psi_n L_n^{-1} t_{in}}{\sigma\left\{\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1} t_{in})^2\right\}^{\frac{1}{2}}}.$$

We now verify the Lindeberg condition for $f_{in}u_i$. We have

$$\sum_{i=1}^{n}\mathbb{E}\left\{(f_{in}u_i)^2\mathbf{1}(|f_{in}u_i| > \epsilon)\right\} \le \max_{1\le i\le n}\mathbb{E}\left\{u_i^2\mathbf{1}\left(u_i^2 > \frac{\epsilon^2}{\max_{1\le i\le n} f_{in}^2}\right)\right\}\sum_{i=1}^{n} f_{in}^2$$

Since u_i^2 are uniformly integrable it suffices to show that $\max_{1\le i\le n} f_{in}^2 \to 0$ as $n \to \infty$, as the last factor on the RHS of the above displayed inequality equals $1/\sigma^2$. Consider

$$\begin{aligned}\max_{1\le i\le n} f_{in}^2 &= \max_{1\le i\le n}\frac{(\alpha'\Psi_n L_n^{-1} t_{in})^2}{\sigma^2\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1} t_{in})^2} \\ &\le \frac{\left\|L_n^{-1}\right\|^2\left\|\Psi_n'\alpha\right\|^2\max_{1\le i\le n}\|t_{in}\|^2}{\sigma^2\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1} t_{in})^2}. \qquad (A.28)\end{aligned}$$

For the denominator of (A.28), note that

$$\begin{aligned}\sum_{i=1}^{n}(\alpha'\Psi_n L_n^{-1} t_{in})^2 &= \alpha'\Psi_n L_n^{-1}\sum_{i=1}^{n} t_{in}t_{in}' L_n^{-1}\Psi_n'\alpha \\ &\ge n\left\|\Psi_n'\alpha\right\|^2(\overline{\eta}(L_n))^{-1} \ge nc\left\|\Psi_n'\alpha\right\|^2,\end{aligned}$$

using Assumption 9. Thus (A.28) is $\mathscr{O}(c_n/n)$ by Assumptions 5, 9 and Lemma B.4. This is negligible by (4.5) and therefore the Lindeberg condition is satisfied.

The asymptotic covariance matrix exists, and is positive definite, by Assumptions 9 and 10. The proof of the consistency of the covariance matrix estimate is omitted. □

B Technical lemmas

Lemma B.1. *Let Assumptions 3 and 4 hold. Then $\|G_{in}\|_R$ and $\|G_{in}'\|_R$ are uniformly bounded for all $i = 1,\ldots,p_n$ and $n \ge 1$.*

Proof. For any $i = 1, \ldots, p_n$,

$$\|G_{in}\|_R = \left\| S_n^{-1} W_{in} \right\|_R \leq \left\| S_n^{-1} \right\|_R \|W_{in}\|_R \leq C$$

where the last inequality follows from Assumption 4. The claim for G'_{in} follows similarly. □

Lemma B.2. *Let Assumptions 2, 3 and 4 hold. Then, for all $i = 1, \ldots, p_n$, the elements of G_{in} are uniformly $\mathscr{O}\left(h_n^{-1}\right)$ as $n \to \infty$.*

Proof. Denote by $w'_{j,in}$ the j-th row of W_{in}. Then the (j,k)-th element of G_{in} is given by $w'_{j,in} S_n^{-1} e_{k,n}$, where $e_{k,n}$ is the n-dimensional vector with unity in the k-th position and zeros elsewhere. Then

$$\left| w'_{j,in} S_n^{-1} e_{k,n} \right| = \left\| w'_{j,in} S_n^{-1} e_{k,n} \right\|_R \leq \|w_{j,in}\|_R \left\| S_n^{-1} \right\|_R \|e_{k,n}\|_R = \mathscr{O}\left(\frac{1}{h_n}\right).$$

where the last inequality follows from Assumptions 2 and 4. □

Lemma B.3. *Let Assumptions 2, 3 and 4 hold. Then, for all $i = 1, \ldots, p_n$, the elements of a product consisting of any finite number of the G_{in} or their transposes are uniformly $\mathscr{O}\left(h_n^{-1}\right)$ as $n \to \infty$. In particular $G'_{in}G_{jn}$ and $G'_{in}G_{jn}$ have elements that are $\mathscr{O}\left(h_n^{-1}\right)$ uniformly in $i, j = 1, \ldots, p_n$ as $n \to \infty$.*

Proof. Similar to proof of Lemma B.2. □

Lemma B.4. *Let Assumptions 3-5 hold. Then the elements of A_n are uniformly $\mathscr{O}(k_n)$.*

Proof. Let $g'_{i,jn}$ be the $i-th$ row of G_{jn}. Then a typical $(i,j)-th$ element of A_n is $g'_{i,jn} X_n \beta$. Now $\left| g'_{i,jn} X_n \beta \right| \leq \left\| g'_{i,jn} \right\|_R \|X_n \beta\|_R = \mathscr{O}(k_n)$ since $\|G_{jn}\|_R$ is uniformly bounded by Lemma (B.1) and by Assumption 5. □

References

Arbia, G., 2006. Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence. Springer-Verlag, Berlin.

Berk, K. N., 1974. Consistent autoregressive spectral estimates. The Annals of Statistics 2, 489–502.

Case, A. C., 1991. Spatial patterns in household demand. Econometrica 59, 953–965.

Case, A. C., 1992. Neighborhood influence and technological change. Regional Science and Urban Economics 22, 491–508.

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

Davis, P. J., 1979. Circulant Matrices. Wiley Interscience, New York.

Gupta, A., 2013. Essays on Spatial Autoregressive Models with Increasingly Many Parameters. Ph. D. thesis, London School of Economics.

Horn, R. A., Johnson, C. R., 1985. Matrix Analysis. Cambridge University Press, New York.

Huber, P. J., 1973. Robust regression: asymptotics, conjectures and Monte Carlo. The Annals of Statistics 1, 799–821.

Kelejian, H. H., Prucha, I. R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. The Journal of Real Estate Finance and Economics 17, 99–121.

Kelejian, H. H., Prucha, I. R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.

Kelejian, H. H., Prucha, I. R., 2002. 2SLS and OLS in a spatial autoregressive model with equal spatial weights. Regional Science and Urban Economics 32, 691–707.

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

Kelejian, H. H., Prucha, I. R., Yuzefovich, Y. A., 2006. Estimation problems in models with spatial weighting matrices which have blocks of equal elements. Journal of Regional Science 46, 507–515.

Kolympiris, C., Kalaitzandonakes, N., Miller, D., 2011. Spatial collocation and venture capital in the US biotechnology industry. Research Policy 40, 1188–1199.

Kreyszig, E., 2011. Introductory Functional Analysis with Applications. Wiley India, New Delhi.

Lee, L. F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252–277.

Lee, L. F., 2003. Best spatial two-stage least squares estimators for a spatial autoregressive model with autoregressive disturbances. Econometric Reviews 22, 307–335.

Lee, L. F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

Lee, L. F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory 26, 187–230.

Portnoy, S., 1984. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. The Annals of Statistics 12, 1298–1309.

Portnoy, S., 1985. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large; II. Normal approximation. The Annals of Statistics 13, 1403–1417.

Robinson, P. M., 1979. Distributed lag approximation to linear time-invariant systems. The Annals of Statistics 7, 507–515.

Robinson, P. M., 2003. Denis Sargan: some perspectives. Econometric Theory 19, 481–494.

Robinson, P. M., 2010. Efficient estimation of the semiparametric spatial autoregressive model. Journal of Econometrics 157, 6–17.

Yuzefovich, Y. A., 2003. Two Essays on Spatial Econometrics. Ph. D. thesis, University of Maryland.