Journal of Econometrics 186 (2015) 19–31

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Inference on higher-order spatial autoregressive models with increasingly many parameters

Abhimanyu Gupta [a], Peter M. Robinson [b,*]

[a] Department of Economics, University of Essex, UK

[b] Department of Economics, London School of Economics, UK

ARTICLE INFO

Article history:
Received 2 December 2013
Received in revised form
23 August 2014
Accepted 19 December 2014
Available online 7 January 2015

JEL classification:
C21
C31
C33
C36

Keywords:
Spatial autoregression
Increasingly many parameters
Central limit theorem
Rate of convergence
Spatial panel data

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 parameters to increase with the data, this has the advantage of accommodating some asymptotic regimes that are suggested by certain spatial settings, several of which are discussed. A small empirical example is also included, and a Monte Carlo study analyses various implications of the theory in finite samples.

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 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 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. This model, allowing $p > 1$, has been studied by, e.g., Blommestein (1983), Huang (1984), Huang and Ahn (1992), Anselin (2001), Lee and Liu (2010) and Badinger and Egger (2013). 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 zero, and they are frequently row-normalized such that each row sums to 1. If W_{in} has non-negative elements, this implies that its (j, l)th element can be interpreted as $w_{jl,in} = d_{jl,in} / \sum_{h=1}^{n} d_{jh,in}$, where $d_{jl,in}$ measures inverse distance between units j and l. Thus elements of the W_{in} are allowed to depend on n, so those of y_n form a triangular array. 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.

* Corresponding author. Tel.: +44 20 7955 7516; fax: +44 20 7955 6592.
E-mail addresses: a.gupta@essex.ac.uk (A. Gupta), p.m.robinson@lse.ac.uk (P.M. Robinson).

http://dx.doi.org/10.1016/j.jeconom.2014.12.008

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 OLS 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 OLS, but inefficient. On the other hand, it is computationally simpler than procedures which may have better statistical properties, such as generalized method of moments (Kelejian and Prucha, 1999; Lee and Liu, 2010), optimal IV (Lee, 2003), Gaussian pseudo maximum likelihood (Lee, 2004), and adaptive estimation (Robinson, 2010). Additionally, desirable asymptotic properties of OLS and IV 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. The model studied in this paper is defined explicitly in the next section.

Higher-order SAR models (such that $p > 1$ in (1.1)) raise serious identifiability problems, which become more acute if p is allowed to increase with n. Our assumptions for consistency and asymptotic normality imply identifiability, but the practitioner needs to be aware of what choices of the W_{in} can potentially afford it. It is obvious that no W_{in} can be a linear combination of the others, but this property is far from sufficient (see e.g. Anselin (2001)). One particular class of W_{in} will transparently avoid identifiability problems, in particular ones with 'single nonzero diagonal block structure'. To define these, denote by V_n an $n \times n$ block diagonal matrix with ith block V_{in}, a $m_i \times m_i$ matrix, where $\sum_{i=1}^{p} m_i = n$, and for $i = 1, \ldots, p$ obtain W_{in} from V_n by replacing each $V_{jn}, j \neq i$, by a matrix of zeros. Thus $V_n = \sum_{i=1}^{p} W_{in}$.

This structure can be thought of as extending a choice of W_n in (1.2) suggested by Case (1991, 1992), where each of p districts contains m farmers, so $n = mp$, and there is interdistrict independence, implying a block diagonal W_n, and also homogeneous within-district reactions, so

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

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. For (1.1) we might then take the W_{in} to have single nonzero diagonal block structure such that

$$m_i = m, \qquad V_{in} = B_m, \quad i = 1, \ldots, p, \tag{1.4}$$

allowing SAR coefficients to vary across districts. In asymptotic theory for (1.1) using (1.3), p is sometimes allowed to diverge with n, and if that happens with (1.4) the number of possibly distinct λ_i likewise increases. Of course (1.4) might then be extended to allow unequal m_i and V_{in} across i, and V_{in} with more complex structure. 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; Robinson, 1979, 2003; Portnoy, 1984, 1985), but not previously with SAR models. Pinkse et al. (2002) consider nonparametric series estimation of a model with spatial weights determined by an unknown function of economic distances.

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, followed by an empirical example in Section 6. We conduct a Monte Carlo study in Section 7, while Section 8 concludes. Proofs are in Appendices A and B

2. Model and basic assumptions

We rewrite (1.1) to stress the possible dependence of the parameter dimension, and the parameters themselves, on n and also introduce an endogenous regression component:

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

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

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

where $S_n = I_n - \sum_{i=1}^{p_n} \lambda_{in} W_{in}$, or equivalently $y_n = R_n \lambda_{(n)} + H_n \gamma_{(n)} + X_n \beta_{(n)} + U_n$ with $R_n = (W_{1n} y_n, \ldots, W_{p_n n} y_n)$. X_n will be taken to comprise exogenous elements, while s_n can also diverge and the elements of H_n are allowed to be correlated with u_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 OLS 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. He also provided a detailed discussion of this assumption. In IV estimation, any mixture of bounded and divergent h_{in} sequences may be employed. 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, $h_n = m - 1$.

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. For any $s \times q$ matrix $A = \left[a_{ij}\right]$ we define $\|A\|_R = \max_{i=1,\ldots,s} \sum_{j=1}^{q} \left|a_{ij}\right|$, the maximum absolute row-sum norm. The proof of the following can be found in the Appendix.

Proposition 2.1. *When for each $i = 1, \ldots, p_n$, $\|W_{in}\|_R \leq 1$ and each W_{in} has a single nonzero diagonal block structure, a sufficient condition for invertibility of S_n is that $|\lambda_{in}| < 1$, $i = 1, \ldots, p_n$.*

Assumption 4. $\left\|S_n^{-1}\right\|_R$, $\left\|S_n'^{-1}\right\|_R$, $\|W_{in}\|_R$ and $\left\|W_{in}'\right\|_R$ are bounded uniformly 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 conditional on the regressors 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.

3. Instrumental variables estimates

Because of endogeneity of the $W_{in}y_n$, $i = 1, \ldots, p_n$, IV estimation has been employed for estimation of SAR models, and we also introduce instruments for the endogenous component H_n. Let Z_n be an $n \times r_n$ matrix of instruments, with $r_n \geq p_n + s_n$ for all n. For (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 $X_n, W_nX_n, W_n^2X_n, W_n^3X_n, \ldots$, assuming the existence of a convergent power series for $(I_n - \lambda W_n)^{-1}$. The existence of such a series is guaranteed if any matrix norm of λW_n is less than 1 (see e.g. Kreyszig (2011), pg. 398). It was suggested that the part of the instrument matrix pertaining to the spatially lagged y_n be constructed from linearly independent subsets of the columns of $X_n, W_nX_n, W_n^2X_n, \ldots, W_n^jX_n$, for some $j \geq 1$. Our theory allows the number of instruments r_n to increase with n.

Write $\overline{\eta}(B)$ (respectively $\underline{\eta}(B)$) for the largest (smallest) eigenvalue of a square, symmetric matrix B. For a rectangular matrix A, define the spectral norm $\|A\| = \left\{\overline{\eta}\left(A'A\right)\right\}^{\frac{1}{2}}$. For the specification (2.1) with $\gamma_{(n)} = 0$,

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

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, (1.4) implies $\sum_{i=1}^{p_n}\lambda_{in}W_{in} = diag(\lambda_{1n}B_m, \lambda_{2n}B_m, \ldots, \lambda_{p_nn}B_m)$, so that $\left\|\sum_{i=1}^{p_n}\lambda_{in}W_{in}\right\| \leq \max_{i=1,\ldots,p_n}|\lambda_{in}|\,\|B_m\|$. 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 $\max_{i=1,\ldots,p_n}|\lambda_{in}| < 1$ suffices for (3.2). See also Proposition 2.1 for a similar result. If (3.1) is indeed valid, instruments may be constructed as subsets of the linearly independent columns of

$$X_n, W_{1n}X_n, W_{1n}^2X_n, \ldots, W_{2n}X_n, W_{2n}^2X_n, \ldots, W_{p_nn}X_n, W_{p_nn}^2X_n, \ldots \tag{3.3}$$

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, who will also need to find instruments for H_n.

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

$$\hat{\theta}_{(n)} = n^{-1}\bar{\bar{Q}}_n^{-1}\bar{\bar{K}}_n'\bar{J}_n^{-1}[Z_n, X_n]'y_n = \theta_{(n)} + \bar{\bar{Q}}_n^{-1}\bar{\bar{K}}_n'\bar{J}_n^{-1}q_n, \tag{3.4}$$

where $\bar{\bar{Q}}_n = \bar{\bar{K}}_n'\bar{J}_n^{-1}\bar{\bar{K}}_n$ (dimension $p_n + s_n + k_n$) and $\bar{\bar{K}}_n = n^{-1}[Z_n, X_n]'[R_n, H_n, X_n]$ (dimension $(r_n + k_n) \times (p_n + s_n + k_n)$), $\bar{J}_n = n^{-1}[Z_n, X_n]'[Z_n, X_n]$ (dimension $r_n + k_n$), $q_n = n^{-1}[Z_n, X_n]'U_n$. Assumption 3 implies that $y_n = S_n^{-1}H_n\gamma_{(n)} + S_n^{-1}X_n\beta_{(n)} + S_n^{-1}U_n$, so $R_n = A_n + B_n$ where $A_n = \left[G_{1n}\left(H_n\gamma_{(n)} + X_n\beta_{(n)}\right), \ldots, G_{p_nn}\left(H_n\gamma_{(n)} + X_n\beta_{(n)}\right)\right]$, $B_n = [G_{1n}U_n, \ldots, G_{p_nn}U_n]$ and $G_{in} = W_{in}S_n^{-1}$ for $i = 1, \ldots, p_n$. Also define $\bar{K}_n = n^{-1}[Z_n, X_n]'[A_n, H_n, X_n]$, $\bar{Q}_n = \bar{K}_n'\bar{J}_n^{-1}\bar{K}_n$, $\bar{L}_n = n^{-1}[A_n, H_n, X_n]'[A_n, H_n, X_n]$ (dimension $p_n + s_n + k_n$). Write, respectively, a'_{in}, p_{in} and $a_{ij,n}$ for the ith row, ith column and a typical element of $[Z_n, X_n]$, and z_{in} for the ith column of Z_n'. Let C denote a generic, arbitrarily large positive constant that is independent of n.

Assumption 5. X_n and z_{in} are independent of u_i, $i = 1, \ldots, n$, and $\max_{1\leq i\leq n,\, 1\leq j\leq r_n+k_n}\mathbb{E}a_{ij,n}^2 < C$.

The elements of the instrument matrix corresponding to spatial lags typically contain elements from different rows of X_n by construction, hence the independence requirement on all of X_n. The assumption implies $\mathbb{E}\|a_{in}\|^2 = \mathcal{O}(r_n + k_n)$ and $\mathbb{E}\|p_{in}\|^2 = \mathcal{O}(n)$, uniformly in i.

Assumption 6. As $n \to \infty$, $\overline{\eta}(\bar{J}_n) = \mathcal{O}_p(1)$ and $\left\{\underline{\eta}(\bar{K}_n'\bar{K}_n)\right\}^{-1} = \mathcal{O}_p(1)$.

Assumption 7. As $n \to \infty$, $\left\{\underline{\eta}(\bar{J}_n)\right\}^{-1} = \mathcal{O}_p(1)$ and $\overline{\eta}(\bar{K}_n'\bar{K}_n) = \mathcal{O}_p(1)$.

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. (2.1) has an intercept), for a row-normalized W_{in} with equal off-diagonal elements (such as (1.4)) $W_{in}y_n$ is asymptotically collinear with the intercept. In this case the second part of Assumption 6 fails, and in fact so does Assumption 12, introduced later. This problem is discussed further in Kelejian and Prucha (2002). However this does not preclude inclusion of an intercept, as the identification conditions may in fact hold if weight matrices are not row-normalized. A necessary condition for both Assumption 12 and $\left\{\underline{\eta}(\bar{K}_n'\bar{K}_n)\right\}^{-1} = \mathcal{O}_p(1)$ 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.

Lemma 3.1. *Under Assumptions 6 and 7 respectively, as $n \to \infty$,*

(i) $\left\{\underline{\eta}\left(\bar{Q}_n\right)\right\}^{-1} = \mathcal{O}_p(1)$.
(ii) $\overline{\eta}\left(\bar{Q}_n\right) = \mathcal{O}_p(1)$.

Theorem 3.1. *Let Assumptions 1–6 hold and*

$$\frac{1}{p_n} + \frac{1}{r_n} + \frac{1}{s_n} + \frac{1}{k_n} + \frac{p_n(r_n + k_n)}{n} \to 0 \quad as\ n \to \infty. \tag{3.5}$$

Then

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

Condition (3.5) details the restrictions on the growth of the numbers of instruments and regressors, and implies a restriction on the growth of the parameter space because $p_n + s_n \leq r_n$. The theorem also holds when r_n (therefore p_n and s_n) and/or k_n are fixed, with (3.5) being altered to exclude reference to the fixed quantity. Slightly weakened conditions yield the same result for the just identified case $p_n + s_n = r_n$, where $\bar{\bar{K}}_n$ and $\bar{K}_n$ are square matrices so that $\hat{\theta}_{(n)} = n^{-1}\bar{\bar{K}}_n^{-1}[Z_n, X_n]'y_n$ and $\bar{Q}_n^{-1} = \bar{K}_n^{-1}\bar{J}_n\bar{K}_n'^{-1}$.

Corollary 3.2. *Let $p_n + s_n = r_n$, Assumptions 1–5 hold, and let*

$$\left\{\underline{\eta}(\bar{K}_n'\bar{K}_n)\right\}^{-1} = \mathcal{O}_p(1), \tag{3.6}$$

and

$$\frac{1}{p_n} + \frac{1}{s_n} + \frac{1}{k_n} + \frac{p_n(p_n + s_n + k_n)}{n} \to 0 \quad as\ n \to \infty. \tag{3.7}$$

Then

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

A natural estimate of σ^2 is

$$\hat{\sigma}_{(n)}^2 = n^{-1}\left\|y_n - [R_n, X_n]\hat{\theta}_{(n)}\right\|^2. \tag{3.8}$$

Assumption 8. As $n \to \infty$, $\overline{\eta}(\bar{L}_n) = \mathscr{O}_p(1)$.

Theorem 3.2. *Let Assumptions 1–6 and 8 hold, and*

$$\frac{1}{p_n} + \frac{1}{r_n} + \frac{1}{s_n} + \frac{1}{k_n} + \frac{(p_n + s_n + k_n)(r_n + k_n)}{n} \to 0 \quad as \quad n \to \infty. \tag{3.9}$$

Then

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

A similar theorem holds in the just identified case $p_n + s_n = r_n$ but we omit the statement for brevity. Here the requirement that h_n be bounded away from zero is crucial (see (A.7)), 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.

Assumption 9. $n^{-1} \max_{1\le i \le n} \|a_{in}\|^2 \xrightarrow{p} 0$ as $n \to \infty$.

This is a standard kind of requirement for a Lindeberg condition to hold. In particular if Z_n and X_n have uniformly bounded nonstochastic elements, it is satisfied if $r_n + k_n = o(n)$.

Assumption 10. There exist constant matrices K_n and J_n (J_n symmetric) satisfying $\underline{\lim}_{n\to\infty} \underline{\eta}\left(K_n' K_n\right) > 0$, $\overline{\lim}_{n\to\infty} \overline{\eta}\left(K_n' K_n\right) < \infty$, $\underline{\lim}_{n\to\infty} \underline{\eta}\left(J_n\right) > 0$ and $\overline{\lim}_{n\to\infty} \overline{\eta}\left(J_n\right) < \infty$, such that $(r_n + k_n)^{\frac{1}{2}} \left(\left\|\bar{K}_n - K_n\right\|, \left\|\bar{J}_n - J_n\right\|\right) = o_p(1)$, as $n \to \infty$.

We define $Q_n = K_n' J_n^{-1} K_n$, and note that $\overline{\lim}_{n\to\infty} \overline{\eta}\left(Q_n\right) < \infty$ and $\underline{\lim}_{n\to\infty} \underline{\eta}\left(Q_n\right) > 0$ can be shown using Assumption 10, as in Lemma 3.1. The presence of X_n in both Z_n and A_n makes it difficult to establish elementwise laws of large numbers (or stochastic orders) for the matrices inside norms, but we illustrate the restriction imposed by the rate above using the bound derived in (A.2) in Appendix A as an example of a rate for the norm of a matrix of increasing dimension in the context of this paper. Indeed, $(r_n + k_n)^{\frac{1}{2}} \left\|\bar{\bar{K}}_n - \bar{K}_n\right\| = \mathscr{O}_p\left(p_n^{\frac{1}{2}} (r_n + k_n)/n^{\frac{1}{2}}\right)$, which would be negligible under (3.10) below. Another example is found in the familiar $X_n' X_n/n$ block. If the rows x_{in}' of X_n are iid with elements having finite fourth moments, then $(r_n + k_n)^{\frac{1}{2}} \left\|X_n' X_n/n - \mathbb{E} x_{in} x_{in}'\right\| = \mathscr{O}_p\left((r_n + k_n)^{\frac{1}{2}} k_n^{\frac{1}{2}}/n^{\frac{1}{2}}\right)$, which is again negligible under (3.10). Denote by Ψ_n a generic $q \times (p_n + k_n)$ matrix of constants with full row-rank.

Theorem 3.3. *Let Assumptions 1–10 hold and*

$$\frac{1}{p_n} + \frac{1}{r_n} + \frac{1}{s_n} + \frac{1}{k_n} + \frac{p_n\left(r_n^2 + k_n^2\right)}{n} + \frac{(k_n + s_n)(r_n + k_n)}{n} \to 0 \quad as\ n \to \infty. \tag{3.10}$$

Then

$$\frac{n^{\frac{1}{2}}}{(p_n + s_n + k_n)^{\frac{1}{2}}} \Psi_n \left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \xrightarrow{d} N\left(0, \Xi_1\right), \quad as\ n \to \infty,$$

where $\Xi_1 = \sigma^2 \lim_{n\to\infty} (p_n + s_n + k_n)^{-1} \Psi_n Q_n^{-1} \Psi_n'$, *which may be consistently estimated by* $\hat{\sigma}^2_{(n)} (p_n + s_n + k_n)^{-1} \Psi_n \bar{\bar{Q}}_n^{-1} \Psi_n'$.

Corollary 3.3. *Let $p_n + s_n = r_n$, Assumptions 1–5, 7–10, (3.6) hold and*

$$\frac{1}{p_n} + \frac{1}{s_n} + \frac{1}{k_n} + \frac{p_n^3}{n} + \frac{p_n\left(s_n^2 + k_n^2\right)}{n} + \frac{s_n k_n}{n} \to 0 \quad as\ n \to \infty. \tag{3.11}$$

Then

$$\frac{n^{\frac{1}{2}}}{(p_n + s_n + k_n)^{\frac{1}{2}}} \Psi_n \left(\hat{\theta}_{(n)} - \theta_{(n)}\right) \xrightarrow{d} N\left(0, \Xi_2\right) \quad as\ n \to \infty,$$

where $\Xi_2 = \sigma^2 \lim_{n\to\infty} (p_n + s_n + k_n)^{-1} \Psi_n K_n^{-1} J_n K_n'^{-1} \Psi_n'$, *which may be consistently estimated by* $\hat{\sigma}^2_{(n)} (p_n + s_n + k_n)^{-1} \Psi_n \bar{\bar{K}}_n^{-1} \bar{J}_n \bar{\bar{K}}_n'^{-1} \Psi_n'$.

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 and s_n are 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 $n^{\frac{1}{2}}/(p_n + s_n + k_n)^{\frac{1}{2}}$-norming is needed to ensure a finite asymptotic covariance matrix, and implies slower than $n^{\frac{1}{2}}$-consistency due to the growing parameter space dimension, while conditions (3.10) and (3.11) restrict the growth. 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 normalization as these rows have increasing dimension. The norming can change if the rows of Ψ_n contain many zero elements, indeed the number of nonzero 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. If we are interested in the first l elements of $\theta_{(n)}$, we may take Ψ_n to be a $1 \times (p_n + k_n)$ non-null row vector with all elements after the lth entry equal to zero. We then recover a $n^{\frac{1}{2}}$-consistency result which indicates that the behaviour of simple t-statistics does not change from the fixed-dimension model (1.1) to (2.1).

Corollary 3.4. (i) *Let Assumptions 1–10 and (3.10) hold. Then*

$$n^{\frac{1}{2}} \left(\hat{\theta}_{(n)} - \theta_{(n)}\right)_l \xrightarrow{d} N\left(0, \sigma^2 \Omega_1^l\right), \quad as\ n \to \infty,$$

where $\left(\hat{\theta}_{(n)} - \theta_{(n)}\right)_l$ *denotes the first l elements of* $\hat{\theta}_{(n)} - \theta_{(n)}$ *while the limiting top-left $l \times l$ block of Q_n^{-1} is denoted Ω_1^l.*

(ii) *Let $p_n + s_n = r_n$, Assumptions 1–5, 7–10, (3.6) and (3.11) hold. Then*

$$n^{\frac{1}{2}} \left(\hat{\theta}_{(n)} - \theta_{(n)}\right)_l \xrightarrow{d} N\left(0, \sigma^2 \Omega_2^l\right), \quad as\ n \to \infty,$$

where Ω_2^l denotes the limiting top-left $l \times l$ block of $K_n^{-1} J_n K_n'^{-1}$. 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 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, in this setting the condition $p_n^3/n \to 0$ (keeping k_n fixed for simplicity) translates to $p^2/m \to 0$. 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 7 illustrates the behaviour of estimates in this setting.

4. Least squares estimates

The model we consider in this section is (2.1) without the endogenous component H_n. Redefine $\theta_{(n)} = \left(\lambda'_{(n)}, \beta'_{(n)}\right)'$, and define its OLS estimate

$$\tilde{\theta}_{(n)} = n^{-1}\bar{\bar{L}}_n^{-1}\left[R_n, X_n\right]' y_n = \theta_{(n)} + \bar{\bar{L}}_n^{-1} w_n, \quad (4.1)$$

where $\bar{\bar{L}}_n = n^{-1}\left[R_n, X_n\right]'\left[R_n, X_n\right]$, $w_n = n^{-1}\left[R_n, X_n\right]' U_n$. Note that in the context of this section $\bar{L}_n = n^{-1}\left[A_n, X_n\right]'\left[A_n, X_n\right]$. Write t'_{in}, m_{in} and $t_{ij,n}$ for the ith row, ith column and typical element of $[A_n, X_n]$ respectively.

Assumption 11. X_n is independent of u_i, $i = 1, \ldots, n$, $\max_{1\le i\le n,\ p_n<j\le p_n+k_n} \mathbb{E}t_{ij,n}^2 < \infty$ and $\max_{1\le i\le n,\ 1\le j\le p_n} \mathbb{E}t_{ij,n}^2 < Ck_n^2$.

The independence of the entire regressor matrix X_n from u_i is needed due to each element of A_n containing elements from different rows of X_n. The assumption implies that $\mathbb{E}\left\|t_{in}\right\|^2 = \mathcal{O}\left(k_n + p_n k_n^2\right)$ and $\mathbb{E}\left\|m_{in}\right\|^2 = \mathcal{O}\left(nk_n^2\right)$, uniformly in i. The second part imposes finite second moments on elements of X_n. A sufficient condition for the last part is $\mathbb{E}\left\|X_n\right\|_R^2 < Ck_n^2$. This part restricts the expectation of the squared elements of A_n, motivated by the fact that the elements of A_n are $\mathcal{O}(k_n)$ if X_n is taken to consist of nonstochastic, uniformly bounded constants.

Assumption 12. As $n \to \infty$, $\left\{\underline{\eta}(\bar{L}_n)\right\}^{-1} = \mathcal{O}_p(1)$.

Theorem 4.1. *Let Assumptions 1–4, 8, 11, 12 hold, and*

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

Then

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

Lee (2002) demonstrated consistency of OLS 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. To obtain a central limit theorem, we additionally assume

Assumption 13. $\mathbb{E}u_i^4 = \mu_4 \le C$ for $i = 1, \ldots, n$.

This is slightly weaker than Lee's (2002) condition $\mathbb{E}\left|u_i\right|^{4+\epsilon} \le C$ for some $\epsilon > 0$, for both consistency and asymptotic normality. We now introduce the OLS residual-based estimate of σ^2, defined as

$$\tilde{\sigma}_{(n)}^2 = n^{-1}\left\|y_n - \left[R_n, X_n\right]\tilde{\theta}_{(n)}\right\|^2. \quad (4.3)$$

Theorem 4.2. *Let Assumptions 1–4, 8, 11–13 hold, and*

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

Then

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

Assumption 14. $n^{-1}\max_{1\le i\le n}\left\|t_{in}\right\|^2 \xrightarrow{p} 0$ as $n \to \infty$.

This condition can be satisfied for uniformly bounded and constant X_n elements if $p_n k_n^2 = o(n)$, as discussed for Assumption 9.

Assumption 15. There exists a constant, symmetric matrix L_n with $\underline{\lim}_{n\to\infty}\underline{\eta}\left(L_n\right) > 0$ and $\overline{\lim}_{n\to\infty}\overline{\eta}\left(L_n\right) < \infty$, and such that $\max\left\{p_n^{\frac{1}{2}}k_n, n^{\frac{1}{2}}p_n^{\frac{1}{2}}h_n^{-1}\right\}\left\|\bar{L}_n - L_n\right\| = o_p(1)$, as $n \to \infty$.

A similar comment to that after Assumption 10 pertaining to the rate condition holds. Here (A.19) in Appendix A can be used as a guide rate, and it is straightforward to see that the resulting orders would be negligible under (4.5).

Theorem 4.3. *Let Assumptions 1–4, 8, 11–15 hold, and*

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

Then

$$\frac{n^{\frac{1}{2}}}{\left(p_n + k_n\right)^{\frac{1}{2}}}\Psi_n\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right) \xrightarrow{d} N\left(0, \Xi_3\right) \quad \text{as } n \to \infty,$$

where $\Xi_3 = \sigma^2\lim_{n\to\infty}\left(p_n + k_n\right)^{-1}\Psi_n L_n^{-1}\Psi'_n$, *which may be estimated consistently using* $\tilde{\sigma}_{(n)}^2\left(p_n + k_n\right)^{-1}\Psi_n\bar{\bar{L}}_n^{-1}\Psi'_n$.

Corollary 4.1. *Let Assumptions 1–4, 8, 11–15 and (4.5) hold. Then*

$$n^{\frac{1}{2}}\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)_l \xrightarrow{d} N\left(0, \sigma^2\Omega_3^l\right), \quad \text{as } n \to \infty,$$

where $\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right)_l$ *denotes the first* l *elements of* $\hat{\theta}_{(n)} - \theta_{(n)}$ *while the limiting top-left* $l \times l$ *block of* L_n^{-1} *is denoted* Ω_3^l, *and the asymptotic covariance matrix is estimated as in Theorem 4.3.*

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 $n = NT$. Let $y_{t,N}$ be the $N \times 1$ vector of observations on the dependent variable for the tth panel, where t may correspond to a time period or a more general spatial unit like a school, village or district. Also let X_{tN} and F_N be $N \times k_1$ and $N \times k_2$ matrices of regressors respectively. X_{tN} contains panel-varying regressors while F_N does not. Consider the model

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

where U_{tN} is the $N \times 1$ vector of disturbances for each panel, formed of iid components. The α_t, $t = 1, \ldots, T$, are scalar fixed effect parameters, the λ_i, $i = 1, \ldots, p$, are scalar and β is a $k_1 \times 1$ panel-invariant parameter vector, whereas ζ_t is a $k_2 \times 1$ parameter vector that can vary with t. For this reason, F_N may be thought of as controlling for 'quasi' fixed-effects. Denote $y_n = \left(y'_{1n}, \ldots, y'_{Tn}\right)'$, $X_n = \left(X'_{1n}, \ldots, X'_{Tn}\right)'$, $U_n = \left(U'_{1n}, \ldots, U'_{Tn}\right)'$, $\alpha = \left(\alpha_1, \ldots, \alpha_T\right)'$ and $\zeta = \left(\zeta_1, \ldots, \zeta_T\right)'$. We can then stack (5.1) to obtain

$$y_n = \left(I_T \otimes l_N\right)\alpha + X_n\beta + \left(I_T \otimes F_N\right)\zeta + \sum_{i=1}^{p}\lambda_i\left(I_T \otimes W_{iN}\right)y_n + U_n. \quad (5.2)$$

This model extends that of Kelejian et al. (2006), and was employed by Yuzefovich (2003). As $n \to \infty$, $N \to \infty$ and/or $T \to \infty$, and the dimension of ζ can diverge. Also, since the $I_T \otimes W_{iN}$ are block diagonal it would be natural to fear that spatial autoregressive parameters differ across panels, or at least among subsets of the panels. For example, allowing a separate spatial parameter for each panel when $p = 1$ implies the model

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

where $W_N^{(i)}$ has a single nonzero diagonal block structure with $V_{in} = W_N$, say, $i = 1, \ldots, T$. (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. In this context condition $p_n^3 k_n^4 = o(n)$ (a sufficient condition for asymptotic normality of OLS in Section 4) translates to $T^6/N \to 0$ as $N, T \to \infty$.

5.2. Models with circulant weight matrices

Define W_{in}^* as the symmetric circulant matrix with first row elements given by

$$w_{1j,in}^* = \begin{cases} 0 & \text{if } j = 1 \text{ or } j = i+2, \ldots, n-i; \\ 1 & \text{if } j = 2, \ldots, i+1 \text{ or } j = n-i+1, \ldots, n. \end{cases} \quad (5.4)$$

Thus the weight matrix W_{in}^* encapsulates a binary neighbourhood criterion for i neighbours on either 'side' of a unit (see Das et al. (2003)). For instance, with $n = 4$

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

Now define

$$W_{in} = \left\| W_{in}^* \right\|^{-1} W_{in}^*, \quad (5.5)$$

where $\left\| W_{in}^* \right\| = \overline{\eta}\left(W_{in}^*\right) = 2i$ because W_{in}^* is a symmetric, circulant matrix (see e.g. Davis (1979) p. 73), implying that (5.5) is row-normalization. Then W_{in} is also a symmetric circulant matrix with first row elements given by $w_{1j,in}^*/2i$.

The maximum number of neighbours is determined by n. We must have $p_n < \lceil n/2 \rceil$ where $\lceil a \rceil$ denotes the smallest integer greater than or equal to a. The W_{in} are easily seen to be linearly independent by showing linear independence of the first rows. Indeed take any finite set $\{W_{in} : i = 1, \ldots, s\}$, 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. It immediately follows that $\xi_s = 0$, implying $\sum_{i=1}^{s-1} \xi_i w_{1,in}' = 0$ so $\xi_{s-1} = 0$. We conclude that $\xi_i = 0$ for each $i = 1, \ldots, s$.

Note that OLS cannot be used to estimate (2.1) with these W_{in}, because as the number of neighbours increases with n this is reflected in additional weight matrices. It is possible to obtain consistent and asymptotically normal IV estimates, however.

6. Empirical example

In this section we provide a small empirical illustration of how a richer model may provide more insight into economic problems. Kolympiris et al. (2011) (henceforth KKM) 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. Denoting by d_{lk} the distance in miles between the lth and kth DBFs, they estimated

$$y_n = \sum_{i=1}^{3} \lambda_i W_{in}^b y_n + X_n\beta + U_n, \quad (6.1)$$

where W_{in}^b is the row-normalized version of a weight matrix having off-diagonal (l, k)th element equal to 1 if $10(i-1) < d_{lk} \le 10i$, $i = 1, 2, 3$, and if $d_{lk} = 0$ for $i = 1$. Thus each weight matrix is based on one of 3 sequential 10-mile rings from the origin DBF. Because the number of neighbours may be taken to increase with sample size, OLS was used. y_n is taken to comprise the natural logs of the amount of VC funding (million \$) received by each of $n = 816$ DBFs. The authors include several explanatory variables in X_n, described below. The number of VCFs within 10 miles of the origin DBF (*VC*010), between 10.01 and 20 miles (*VC*1020), between 20.01 and 30 miles (*VC*2030), similarly the number of DBFs within three 10-mile rings of the origin DBF (*DBF*010, *DBF*1020, *DBF*2030) capture the effects of being in areas of high VCF or DBF concentration. There are also firm-specific factors for the VCFs. These include a variable that measures the natural logarithm of the average (weighted by funding share) distance from each DBF to the VCFs that fund it (*VCDist*), the average total amount invested in all firms by each VCF that funded the origin DBF (*VCSize*), the average age of each VCF (as of 2007) that funded the origin DBF (*VCAge*), the average number of other DBFs a funding VCF of the origin DBF co-invested in through syndication (*VCSynd*), a dummy variable (*Foreign*) that equals 1 if the origin DBF received funds from non-US VCFs (and equals 0 otherwise). Variables controlling for DBF-specific factors include the age of a DBF (*Age*, *Age*^{2}) to capture concavity in the sense that a mature firm will rely less on VCF, yet also be more attractive to VCFs. The dummy variable *SBIR* equals one if the DBF received Small Business Innovation Research (SBIR) grants (and equals 0 otherwise), while another dummy variable (*Tax*) equals one if the DBF's state had an R&D tax credit from 1990 to 2007. Also included are an index of the cost of doing business in the DBF's state (*Business*), the distance from a DBF to the closest university (*University*) and the average total number of non-biotech establishments (*Establishments*) in the DBF's zip code from 1990 to 2007. The final two variables recognize that additional factors can affect the cost of doing business in ways that influence the VC funding levels of a given DBF. The first is defined as $Cross = W_{NS} \times D$, where W_{NS} is a $n \times n$ non-standardized spatial weight matrix that identifies the firms located within a 1-mile radius from the origin firm, and D is the $n \times 1$ vector of the weighted distance for each firm from its financing VCFs. *Cross* increases in both the number of firms within 1-mile and in distance from funding VCFs, so it captures the potential increase of VC funds realized by firms closely surrounded by a large number of firms funded by distant VCFs. The second (*LocalVC*) measures the number of DBFs located less than 10 miles from the origin DBF which were funded by VCF(s) located in the same radius that also funded the origin DBF, accounting for potential cost efficiencies realized by VCFs when they invest in proximate firms. An intercept is also included. We omit a more detailed discussion of the variables to conserve space, but this is available in KKM.

KKM discover that only the spatial lag corresponding to the first 10 mile ring is significant. Of the regressors in X_n, only the following were found to be significant at the 5% level: *Age*, *Age*^{2}, *SBIR*, *VCFSynd*, *Foreign*, *LocalVC*, *VCDist* and *DBF*010, apart from the intercept. We extend the analysis by reducing the radius of the rings used in defining neighbours to 1 mile, in keeping with our suggestion that with a fairly large sample of 816 observations a richer model can be estimated. The specification is

$$y_n = \sum_{i=1}^{10} \lambda_i W_{in}^b y_n + X_n\beta + U_n, \quad (6.2)$$

Table 6.1
OLS estimates of λ_i in (6.2) and (6.3).

Parameter	Specification	
	(6.2)	(6.3)
λ_1	0.0964* (2.7428)	0.1051* (4.6698)
λ_2	0.0823* (2.1982)	
λ_3	−0.0058 (0.1482)	
λ_4	0.0594 (1.5344)	
λ_5	0.0046 (0.1299)	
λ_6	0.0128 (0.3753)	
λ_7	0.0261 (0.8168)	
λ_8	0.0265 (0.7951)	
λ_9	0.0560 (1.5007)	
λ_{10}	0.0408 (1.1477)	

In parentheses are absolute t-ratios (based on heteroskedasticity-consistent standard errors).
* 5% significance.

where now W_{in}^b is the row-normalized version of a weight matrix having off-diagonal (l, k)th element equal to 1 if $i - 1 < d_{lk} \leq i$, $i = 1 \dots, 10$, and if $d_{lk} = 0$ for $i = 1$.

The first columns on the right of Tables 6.1 and 6.2 report the autoregressive and regression coefficient estimates, respectively, from (6.2). Only the first two spatial lags are significant at the 5% level, implying that spatial dependence is restricted to a small geographical radius. A test of $\lambda_1 = \lambda_2$ returned a t-statistic of 0.2609, not ruling out equal spatial dependence from each ring. The regression coefficient estimates broadly correspond in sign and magnitude to those in KKM, with three exceptions. Allowing for 'thinner' rings increases the significance of *DBF*010, *DBF*2030 and *VC*010. The first was found to be significant at the 10% level by KKM but is now significant at even 0.1%, while the latter two were insignificant but are now significant at the 5% level. The impact on VC funding of having competing DBFs in the geographical area may operate in two opposite directions. On the one hand areas of high DBF concentration may attract more VC funding for all firms, while on the other the competition for VC funds may have a negative impact. It is evident that allowing for a richer model has enabled us to pick up the non-negligible latter impact. Similarly a concentration of VCFs in close vicinity has a positive impact on DBF funding. In light of the analysis above we also estimate

$$y_n = \lambda_1 \left(W_{1n}^b + W_{2n}^b\right) y_n + X_n\beta + U_n. \tag{6.3}$$

The results are reported in the second columns on the right of Tables 6.1 and 6.2. There are no major changes in the conclusions qualitatively. *DBF*010 remains significant but *DBF*2030 no longer does, while *VC*1020 also becomes significant. The intercept loses significance marginally.

7. Monte Carlo experiments

Two sets of Monte Carlo experiments were carried out with three aims. First, to assess the extent to which more data improve precision of estimates despite increasing number of parameters, and to understand the rate at which this may happen. 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.

Table 6.2
OLS estimates of (6.2) and (6.3): regression coefficients.

Variables	Specification	
	(6.2)	(6.3)
Intercept	−1.4542* (2.1906)	−1.1678 (1.9235)
Age (years)	0.2499* (5.1138)	0.2520* (5.1236)
Age^2 (years)	−0.0120* (4.3680)	−0.0120* (4.3631)
SBIR (binary)	0.2368* (2.3253)	0.2120* (2.0637)
Tax (binary)	−0.1503 (0.6215)	−0.0674 (0.2861)
VCDist (miles)	−0.0020 (0.6485)	−0.0028 (0.9147)
Business	0.0034 (0.5681)	0.0003 (0.0721)
Establishments	0.0174 (0.2938)	0.0315 (0.5469)
VCSize	0.0000 (0.9062)	0.0000 (0.9063)
VCAge	0.0049 (0.7880)	0.0037 (0.6048)
VCSynd level	0.0336* (5.4671)	0.0350* (5.7465)
Foreign (binary)	1.1754* (12.6853)	1.1819* (12.8610)
VCDist	0.2172* (5.8581)	0.2214* (6.0478)
Cross	0.0012 (1.2103)	0.0000 (0.0618)
LocalVC	0.0229* (3.3887)	0.0226* (3.3145)
DBF010	−0.0057* (2.9790)	−0.0032* (2.0458)
DBF1020	−0.0022 (0.8710)	−0.0016 (0.6400)
DBF2030	−0.0061* (2.0223)	−0.0052 (1.8100)
VCF010	0.0049* (2.1070)	0.0071* (3.2732)
VCF1020	0.0052 (1.5818)	0.0065* (2.0636)
VCF2030	0.0048 (1.5528)	0.0043 (1.4227)

In parentheses are absolute t-ratios (based on heteroskedasticity-consistent standard errors).
* 5% significance.

In the first set of experiments we chose weight matrices as (1.3) and (1.4). The number of regressors was kept fixed at $k_n = 2$ for simplicity, and we experimented with $p = 2, 6, 18$ and $m = 50, 150, 450$. Note that in this setting we have $p_n = p$ and $n = pm$ so that $p_n^2/n = p/m$. The elements of X_n were generated from a uniform distribution on $(0, 1)$, and kept fixed over replications. The u_i were generated as iid draws from a standard normal ($\sigma^2 = 1$) distribution, and instruments were constructed as in (3.3) using only first-order spatial lags of the regressors ($j = 1$). The vector y_n was generated using (1.1) and (1.4) (without H_n) in each of the 1000 replications. We chose $\beta_1 = 1$, $\beta_2 = 0.5$ and $\lambda_1 = 0.7$, $\lambda_2 = 0.8$, $\lambda_3 = 0.5$, $\lambda_4 = 0.8$, $\lambda_5 = 0.3$, $\lambda_6 = 0.6$, $\lambda_7 = 0.7$, $\lambda_8 = 0.8$, $\lambda_9 = 0.5$, $\lambda_{10} = 0.8$, $\lambda_{11} = 0.3$, $\lambda_{12} = 0.6$, $\lambda_{13} = 0.7$, $\lambda_{14} = 0.8$, $\lambda_{15} = 0.5$, $\lambda_{16} = 0.8$, $\lambda_{17} = 0.3$, $\lambda_{18} = 0.6$. For our analysis we employ Monte Carlo average mean-squared error (AMSE) defined as $\text{AMSE}\left(\hat{\theta}_{(n)}\right) = s^{-1}\sum_{i=1}^{s} \text{MSE}\left(\hat{\theta}_{in}\right)$, with $\text{MSE}\left(\hat{\theta}_{in}\right)$ denoting the Monte Carlo MSE for the IV estimate of the ith element $\hat{\theta}_{in}$ in the $s \times 1$ vector $\hat{\theta}_{(n)}$. $\text{AMSE}\left(\tilde{\theta}_{(n)}\right)$ is defined analogously.

Define the relative AMSE (RAMSE) as $\text{RAMSE}(Q_1, Q_2) = \text{AMSE}(Q_1)/\text{AMSE}(Q_2)$, for two generic estimators Q_1 and Q_2. Table 7.1 presents AMSE and RAMSE of the IV and OLS estimates for various combinations of p and m. The AMSE for both reduces

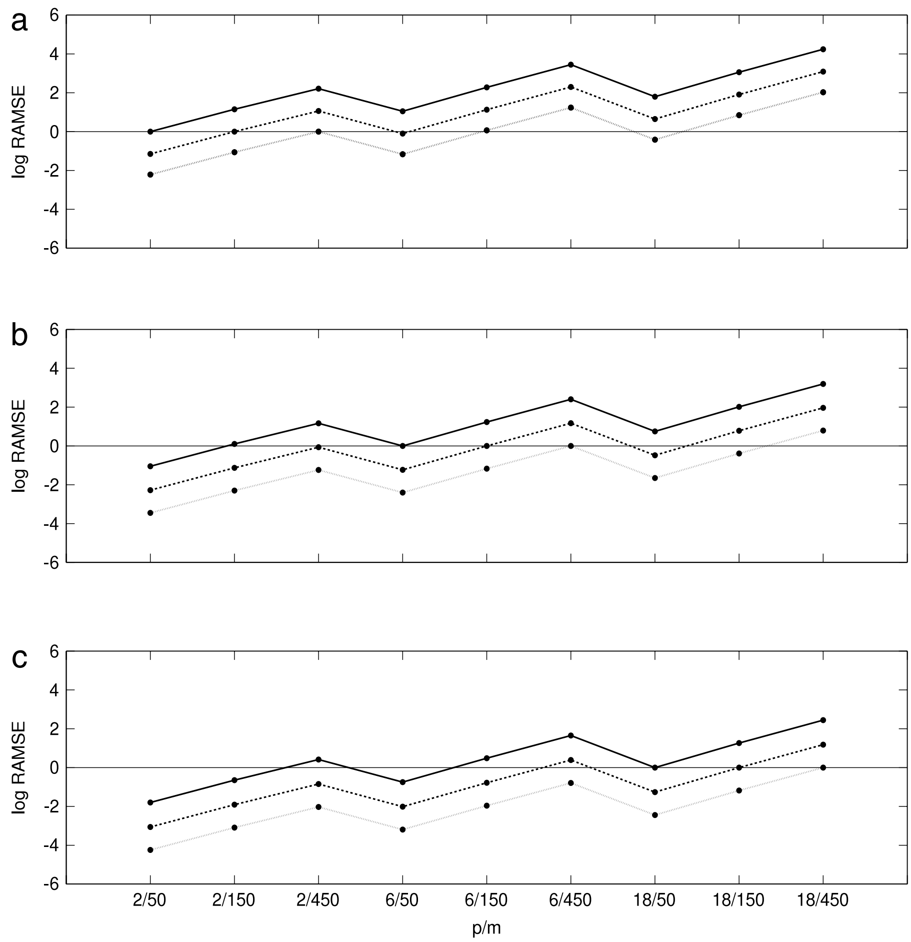

Fig. 7.1. Monte Carlo log RAMSE for IV estimates $\hat{\theta}_{(n)}$.

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 indicates that even though OLS is asymptotically efficient IV outperforms it in many cases. Additional simulations (not reported here) indicate that OLS 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 $h_n = m - 1$, supporting the theory that $h_n \to \infty$ is required for consistency of OLS with a fast rate of divergence required for asymptotic normality, while the properties of IV do not depend on the divergence of h_n.

Figs. 7.1(a)–(c) and 7.2(a)–(c) plot the log RAMSE (LRAMSE) for $\hat{\theta}_{(n)}$ and $\tilde{\theta}_{(n)}$ for all possible ratios p/m. Fig. 7.1(a) illustrates LRAMSE 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 LRAMSE. Each solid dot on the figure marks out the LRAMSE 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 LRAMSE for $m = 50$, the heavy dotted line for $m = 150$ and the light dotted line for $m = 450$. Fig. 7.1(b), (c) repeats the analysis for $p = 6$ and $p = 18$ respectively, again the solid lines joining the LRAMSE for $m = 50$, the heavy dotted lines for $m = 150$ and the light dotted lines for $m = 450$. Fig. 7.2(a)–(c) does exactly the same for the OLS estimate $\tilde{\theta}_{(n)}$. By definition the LRAMSE of an estimate with itself corresponds to a value of zero. Negative LRAMSE 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.

Table 7.1
Monte Carlo AMSE and RAMSE of IV and OLS estimates using (1.4).

p	m	50	150	450
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
	RAMSE$\left(\hat{\theta}_{(n)}, \tilde{\theta}_{(n)}\right)$	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
	RAMSE$\left(\hat{\theta}_{(n)}, \tilde{\theta}_{(n)}\right)$	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
	RAMSE$\left(\hat{\theta}_{(n)}, \tilde{\theta}_{(n)}\right)$	0.6250	0.8479	0.9218

We now analyse Fig. 7.1(a) in detail. The solid line increases through LRAMSE comparisons for estimates with $p = 2$ but higher values of m. While the value of LRAMSE 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 LRAMSE then increases in the $p = 6$ range, before dropping and then increasing again in the $p = 18$ range. We make two comments about these results, reflective of the discussion of Table 7.1. First, estimates become more precise for given p with increasing m. Second, many more

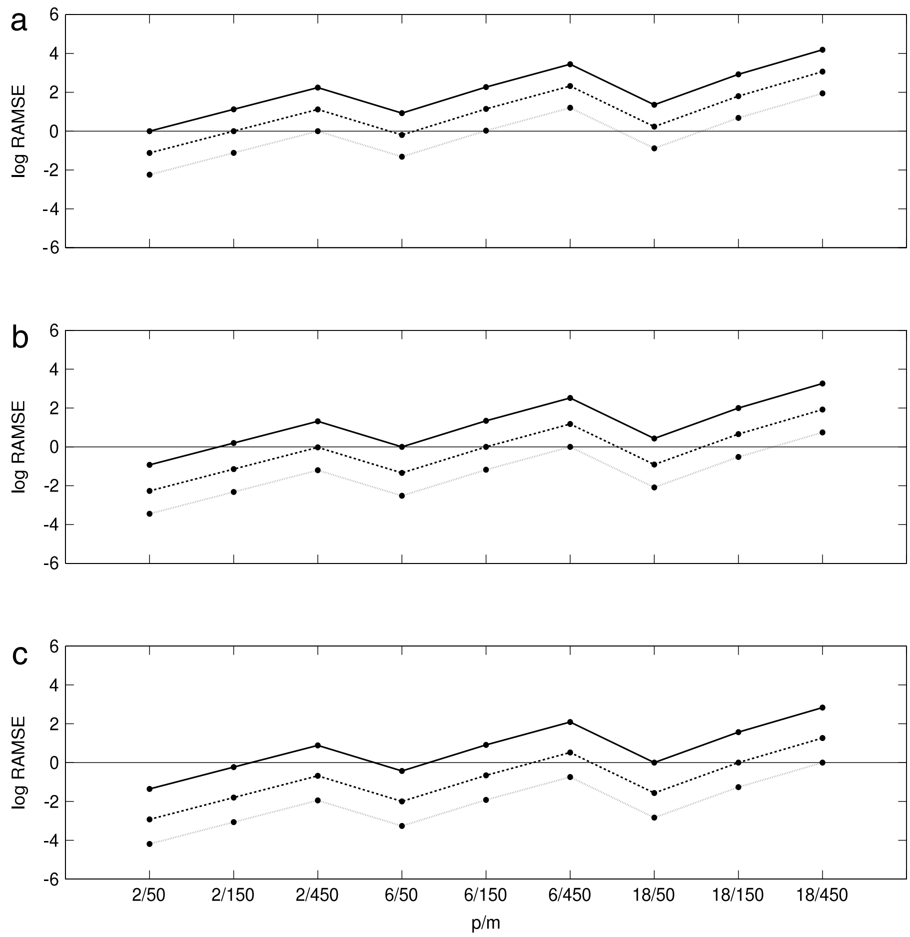

Fig. 7.2. Monte Carlo log RAMSE for OLS estimates $\tilde{\theta}_{(n)}$.

Table 7.2
Monte Carlo AMSE of IV estimates using (5.5).

p	n	108	432	864
2	AMSE $\left(\hat{\theta}_{(n)}\right)$	0.3981	0.0810	0.0350
4	AMSE $\left(\hat{\theta}_{(n)}\right)$	2.4391	0.5684	0.2355

parameters can be estimated much more accurately with increasing data. The first point suggests that in practice, when a data set 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 in SAR models, as discussed earlier. The heavy dotted ($m = 150$) and light dotted ($m = 450$) lines behave in the same way, the difference being that the initial values of LRAMSE are negative. This is because, for instance, the first point on the heavy dotted line corresponds to the LRAMSE 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. Analysis of the other figures indicates the same pattern in the rate of change of the LRAMSE.

The second set of experiments employed the weight matrices (5.5). We considered $p = 2, 4$, $n = 108, 432, 864$ and chose $\lambda_1 = 0.2$, $\lambda_2 = 0.1$, $\lambda_3 = 0.2$, $\lambda_4 = 0.1$ to satisfy (3.2). The remainder of the design was kept identical to that considered in the first set of experiments. In this setting only IV estimates were considered (cf. Section 5), with instruments given again by first-order spatial lags of X_n. Table 7.2 presents the results. The AMSE is seen to decline with increasing n but remain rather high for $p = 4$ even when $n = 864$. The choice of lag order should be economically motivated but seems to depend heavily on n, with rather large n needed to justify even moderate choices of p.

Table 7.3 displays average Monte Carlo sizes of tests based on IV and OLS and Theorems 3.3 and 4.3, for various combinations of p and n, and using both specifications for weight matrices. These

Table 7.3
Monte Carlo average sizes.

p	m	Using (1.4)			p	n	Using (5.5)		
		50	150	450			108	432	864
2	$\hat{\theta}_{(n)}$	0.0590	0.0473	0.0520	2	$\hat{\theta}_{(n)}$	0.0305	0.0403	0.0498
	$\tilde{\theta}_{(n)}$	0.0612	0.0443	0.0460					
6	$\hat{\theta}_{(n)}$	0.0464	0.0484	0.0452	4	$\hat{\theta}_{(n)}$	0.0172	0.0428	0.0535
	$\tilde{\theta}_{(n)}$	0.0623	0.0530	0.0507					
18	$\hat{\theta}_{(n)}$	0.0488	0.0493	0.0511					
	$\tilde{\theta}_{(n)}$	0.0771	0.0560	0.0542					

sizes should be compared with the nominal 5%. The table indicates that the sizes approach the nominal value as n increases, though not always monotonically. When using (5.5) there tends to be mostly undersizing, quite severe for smaller n owing to the issues discussed above. On the other hand, when using (1.4) OLS usually oversizes while IV usually undersizes, but neither does so greatly. On the whole, the asymptotic distributions seem to provide a reasonable approximation.

8. Conclusion

Asymptotic theory has been presented for a general class of spatial autoregressive models with increasingly many parameters, and with at least one non-intercept regressor. The theory is thus not applicable to the pure SAR model or the SAR model with only an intercept regressor, though analogous theory for this can be developed using parameter estimates such as the Gaussian pseudo maximum likelihood estimate (PMLE). The latter (studied by Lee (2004) in the $p = 1$ case) provides consistent and efficient estimates under weaker conditions on h_n than OLS and also, unlike OLS and IV, can estimate models with no explanatory variables, or only an intercept. It may be worth investigating data-dependent choices of p, though in the SAR context a higher order does not necessarily reflect greater distance.

Acknowledgements

The authors thank two referees and an associate editor for helpful comments, and are also grateful to Christos Kolympiris for sharing the data for the empirical example. The second author's research was supported by ESRC Grant ES/J007242/1.

Appendix A. Proofs of results in Sections 2–4

Write $a_n = p_n + s_n + k_n$, $b_n = r_n + k_n$, $c_n = p_n k_n^2 + k_n$ and $\tau_n = n^{\frac{1}{2}}/a_n^{\frac{1}{2}}$. For any matrices F_n, $\bar{F}_n$ and $\bar{\bar{F}}_n$ of equal dimension, we will write $\bar{\bar{\Delta}}_n^F = \bar{F}_n - \bar{\bar{F}}_n$, $\bar{\Delta}_n^F = F_n - \bar{F}_n$.

Proof of Proposition 2.1. By definition, $S_n = I_n - \mathrm{diag}\left[\lambda_{1n}V_{1n}, \ldots, \lambda_{p_n n}V_{p_n}\right]$, with unit diagonal elements. With $v_{lk,in}$ denoting a typical element of V_{in}, due to row-normalization we have $\sum_{s\neq r, m\neq l} |\lambda_{in}|^2 \left|v_{rs,in}\right| \left|v_{lm,in}\right| < 1$ if $\lambda_{in}^2 < 1$, whence the claim follows from Horn and Johnson (1985), p. 381, Corollary 6.4.11 (b). □

Proof of Lemma 3.1. (i) For any $a_n \times 1$ vector x_n satisfying $\|x_n\| = 1$, $x_n'\bar{K}_n'\bar{J}_n^{-1}\bar{K}_n x_n \geq \underline{\eta}\left(\bar{K}_n'\bar{K}_n\right)/\overline{\eta}\left(\bar{J}_n\right)$, whence the lemma follows by Assumption 6. The proof of (ii) is similar. □

Proof of Theorem 3.1. Write

$$\hat{\theta}_{(n)} - \theta_{(n)} = \bar{Q}_n^{-1}\bar{\bar{\Delta}}_n^{Q}\left(\hat{\theta}_{(n)} - \theta_{(n)}\right) - \bar{Q}_n^{-1}\bar{\bar{\Delta}}_n^{K'}\bar{J}_n^{-1}q_n + \bar{Q}_n^{-1}\bar{K}_n'\bar{J}_n^{-1}q_n. \quad (A.1)$$

By elementary norm inequalities $\left\|\bar{\bar{\Delta}}_n^{Q}\right\| \leq \left\|\bar{\bar{\Delta}}_n^{K}\right\| \left\|\bar{J}_n^{-1}\right\| \left(\left\|\bar{\bar{\Delta}}_n^{K}\right\| + 2\left\|\bar{K}_n\right\|\right)$, where $\mathbb{E}\left\|\bar{\bar{\Delta}}_n^{K}\right\|^2$ is bounded by $\sigma^2\sum_{i=1}^{b_n}\mathbb{E}\left\|p_{in}\right\|^2\sum_{j=1}^{p_n}\left\|G_{jn}\right\|^2/n^2 = \mathcal{O}\left(p_n b_n/n\right)$, by Assumption 5 and Lemma B.1. We conclude that

$$\left\|\bar{\bar{\Delta}}_n^{K}\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}\right), \quad (A.2)$$

by Markov's inequality. Then

$$\left\|\bar{\bar{\Delta}}_n^{Q}\right\| = \mathcal{O}_p\left(\max\left\{n^{-1}p_n b_n, n^{-\frac{1}{2}}p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}\right\}\right) = \mathcal{O}_p\left(n^{-\frac{1}{2}}p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}\right), \quad (A.3)$$

by Assumption 7. Likewise $\mathbb{E}\left\|q_n\right\|^2 = \sigma^2\sum_{i=1}^{n}\mathbb{E}\left\|a_{in}\right\|^2/n^2 = \mathcal{O}\left(b_n/n\right)$, by Assumption 5. By Markov's inequality

$$\left\|q_n\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}b_n^{\frac{1}{2}}\right). \quad (A.4)$$

From (A.1),

$$\left(1 - \left\|\bar{Q}_n^{-1}\right\|\left\|\bar{\bar{\Delta}}_n^{Q}\right\|\right)\left\|\hat{\theta}_{(n)} - \theta_{(n)}\right\| \leq \left\|\bar{Q}_n^{-1}\right\|\left\|\bar{J}_n^{-1}\right\|\left\|q_n\right\|\left(\left\|\bar{\bar{\Delta}}_n^{K}\right\| + \left\|\bar{K}_n\right\|\right). \quad (A.5)$$

By (A.3) the first factor on the LHS converges in probability to one by (3.5) and Lemma 3.1(i), the first two factors on the RHS of (A.5) are $\mathcal{O}_p(1)$ by Assumption 7. The third and fourth factors have orders given in (A.2) and (A.4) respectively, implying that

$$\left\|\hat{\theta}_{(n)} - \theta_{(n)}\right\| = \mathcal{O}_p\left(\max\left\{n^{-1}p_n^{\frac{1}{2}}b_n, n^{-\frac{1}{2}}b_n^{\frac{1}{2}}\right\}\right) = \mathcal{O}_p\left(n^{-\frac{1}{2}}b_n^{\frac{1}{2}}\right). \quad (A.6)$$

This is negligible by (3.5). The proof of Corollary 3.2 is similar. □

Proof of Theorem 3.2. Write

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

From Assumption 1, $n^{-1}U_n'U_n = \sigma^2 + o_p(1)$. Also by (A.6) and (A.17) the modulus of the second term is bounded by $\left\|\hat{\theta}_{(n)} - \theta_{(n)}\right\| \left\|w_n\right\| = \mathcal{O}_p\left(\max\left\{b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n, p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}\right\}\right)$ while the third term has modulus bounded by

$$\left\|\hat{\theta}_{(n)} - \theta_{(n)}\right\|^2\left(\left\|\bar{L}_n\right\| + \left\|\bar{\bar{\Delta}}_n^{L}\right\|\right) = \mathcal{O}_p\left(b_n\max\left\{n^{-1}, n^{-\frac{3}{2}}p_n^{\frac{1}{2}}k_n a_n^{\frac{1}{2}}, \left(nh_n\right)^{-1}p_n\right\}\right)$$

using (A.6), (A.19) and Assumption 8. Because $p_n^{\frac{1}{2}}k_n a_n^{\frac{1}{2}}b_n/n^{\frac{3}{2}}$ and $p_n b_n/nh_n$ are dominated by $b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n$ and $p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}/n^{\frac{1}{2}}h_n^{\frac{1}{2}}$ respectively under (3.9), we have

$$\hat{\sigma}_{(n)}^2 - \sigma^2 = \mathcal{O}_p\left(\max\left\{n^{-1}b_n^{\frac{1}{2}}c_n^{\frac{1}{2}}, \left(n^{\frac{1}{2}}h_n^{\frac{1}{2}}\right)^{-1}p_n^{\frac{1}{2}}b_n^{\frac{1}{2}}, n^{-1}b_n\right\}\right), \quad (A.7)$$

which is negligible by (3.9) and because h_n is bounded away from zero. □

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

$$\tau_n\alpha'\Psi_n\left(\hat{\theta}_{(n)}-\theta_{(n)}\right)$$
$$=\tau_n\alpha'\Psi_n\bar{Q}_n^{-1}\bar{\bar{\Delta}}_n^{Q}\left(\hat{\theta}_{(n)}-\theta_{(n)}\right)-\tau_n\alpha'\Psi_n\left(\bar{Q}_n^{-1}\bar{\bar{\Delta}}_n^{K'}\bar{J}_n^{-1}\right.$$
$$-\bar{Q}_n^{-1}\bar{\Delta}_n^{Q}Q_n^{-1}\bar{K}_n'\bar{J}_n^{-1}+Q_n^{-1}\bar{\Delta}_n^{K'}\bar{J}_n^{-1}$$
$$\left.-\ Q_n^{-1}K_n'\bar{J}_n^{-1}\bar{\Delta}_n^{J}J_n^{-1}\right)q_n+\tau_n\alpha'\Psi_nQ_n^{-1}K_n'J_n^{-1}q_n. \quad (A.8)$$

We first show that the first term on the RHS of (A.8) is $o_p(1)$. It has modulus bounded by $\tau_n\left\|\alpha\right\|\left\|\Psi_n\right\|\left\|\hat{\theta}_{(n)}-\theta_{(n)}\right\|\left\|Q_n^{-1}\right\|\left\|\bar{\bar{\Delta}}_n^{Q}\right\|=\mathcal{O}_p\left(p_n^{\frac{1}{2}}b_n/n^{\frac{1}{2}}\right)$, by (A.3), (A.6) and Assumption 7, 10. This is negligible by (3.10) because $b_n^2\leq 2\left(r_n^2+k_n^2\right)$. Similarly (A.4), Assumption 6, 7, 10 and Lemma 3.1 imply that the second term on the RHS of (A.8) is $\mathcal{O}_p\left(b_n^{\frac{1}{2}}\max\left\{\left\|\bar{\bar{\Delta}}_n^{K}\right\|,\left\|\bar{\Delta}_n^{Q}\right\|,\left\|\bar{\Delta}_n^{K}\right\|,\left\|\bar{\Delta}_n^{J}\right\|\right\}\right)$, which is negligible by (A.2), Assumption 10 and (3.10), noting that $\left\|\bar{\Delta}_n^{Q}\right\|\leq\left\|\bar{\Delta}_n^{K}\right\|\left\|J_n^{-1}\right\|\left(\left\|\bar{\Delta}_n^{K}\right\|+2\left\|K_n\right\|\right)$.

Because $\tau_n\alpha'\Psi_nQ_n^{-1}K_n'J_n^{-1}q_n=n^{-\frac{1}{2}}a_n^{-\frac{1}{2}}\sum_{i=1}^{n}\alpha'\Psi_nQ_n^{-1}K_n'J_n^{-1}a_{in}u_i$, we will consider $n\alpha'\Psi_nQ_n^{-1}K_n'J_n^{-1}q_n/v_n=\sum_{i=1}^{n}c_{in}u_i$, where $c_{in}=\alpha'\Psi_nQ_n^{-1}K_n'J_n^{-1}a_{in}/v_n$ and $v_n=\sigma\left(\sum_{i=1}^{n}(\alpha'\Psi_nQ_n^{-1}K_n'J_n^{-1}a_{in})^2\right)^{\frac{1}{2}}$. $\{c_{in}u_i, i=1,\ldots,n, n\geq 1\}$ forms a martingale difference sequence, by Assumption 5. It suffices to show $\sum_{i=1}^{n}c_{in}u_i\overset{d}{\to}N(0,1)$, conditional on z_{in}, X_n, which follows by Theorem 2 of Scott (1973) if conditional on z_{in}, X_n, as $n\to\infty$,

$$\mathbb{E}\left(\sum_{i=1}^{n}c_{in}^2u_i^2\mid u_j, j<i\right)\overset{p}{\to}1 \quad (A.9)$$

and for all $\xi>0$

$$\mathbb{E}\left(\sum_{i=1}^{n}c_{in}^2\mathbb{E}\left(u_i^2\mathbf{1}\left(|c_{in}u_i|>\xi\right)\mid z_{in},X_n\right)\right)\to 0, \quad (A.10)$$

with $\mathbf{1}(\cdot)$ denoting indicator function. The LHS of (A.9) equals 1, while the LHS of (A.10) is bounded by

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

Since the u_i^2 are uniformly integrable, it suffices to show that $\max_{1\leq i\leq n}c_{in}^2\overset{p}{\to}0$ as $n\to\infty$, as the last factor equals $1/\sigma^2$. Now, $\max_{1\leq i\leq n}c_{in}^2\leq\left\|Q_n^{-1}K_n'J_n^{-1}\right\|^2\left\|\Psi_n'\alpha\right\|^2\max_{1\leq i\leq n}\left\|a_{in}\right\|^2/v_n^2\leq\max_{1\leq i\leq n}\left\|a_{in}\right\|^2\left\|Q_n^{-1}K_n'J_n^{-1}\right\|^2\sigma^{-2}\overline{\eta}\left(Q_n\right)/n$ for large n, because $\sum_{i=1}^{n}a_{in}a_{in}'=nJ_n$. This is negligible as $n\to\infty$ by Assumptions 6, 7 and 9. The asymptotic covariance matrix exists, and is positive definite, by Lemma 3.1. The proof of consistency of the covariance matrix estimate is omitted, while the proof of Corollary 3.3 is similar, existence and positive definiteness of the asymptotic covariance matrix guaranteed by Assumptions 6 and 7. □

Proof of Theorem 4.1. Note that $s_n=0$ for the proofs of the results in Section 4. Write

$$\tilde{\theta}_{(n)}-\theta_{(n)}=\bar{L}_n^{-1}\bar{\bar{\Delta}}_n^{L}\left(\tilde{\theta}_{(n)}-\theta_{(n)}\right)+\bar{L}_n^{-1}w_n. \quad (A.11)$$

It is clear that

$$\left\|w_n\right\|\leq\left\|n^{-1}\left[A_n,X_n\right]'U_n\right\|+\left\|n^{-1}\left[B_n,0\right]'U_n\right\|. \quad (A.12)$$

Now

$$\mathbb{E}\left\|n^{-1}\left[A_n,X_n\right]'U_n\right\|^2=\mathcal{O}\left(n^{-1}c_n\right), \quad (A.13)$$

as in the proof Theorem 3.1, but using Assumption 11.

Write $g_{rs,in}$ for the (r,s)th element of G_{in}. Though we do not assume that the u_{in} have finite fourth moment in the statement of Theorem 4.1, it is convenient to present in the current proof also some calculations that use Assumption 13 and will be useful in the proof of Theorem 4.3. In this case, the square of the second term on the RHS of (A.12) has expectation $\sum_{i=1}^{p_n}\mathbb{E}\left(U_n'G_{in}'U_n/n\right)^2=\mu_4\sum_{i=1}^{p_n}\sum_{j=1}^{n}\left(g_{jj,in}/n\right)^2+\left(\sigma^4/n^2\right)\sum_{i=1}^{p_n}\sum_{j,k=1}^{n}\left(g_{jj,in}g_{kk,in}+g_{jk,in}g_{kj,in}+g_{jk,in}^2\right)=\mathcal{O}(p_n/h_n^2\max\{1/n,1\})$, by Lemma B.2. Hence

$$\left\|n^{-1}\left[B_n,0\right]'U_n\right\|=\mathcal{O}_p\left(h_n^{-1}p_n^{\frac{1}{2}}\right), \quad (A.14)$$

so that

$$\left\|w_n\right\|=\mathcal{O}_p\left\{\max\left(n^{-\frac{1}{2}}c_n^{\frac{1}{2}},h_n^{-1}p_n^{\frac{1}{2}}\right)\right\}. \quad (A.15)$$

However, with only second moments for the u_i we have

$$\left\|n^{-1}\left[B_n,0\right]'U_n\right\|\leq n^{-1}\left\|\left[B_n,0\right]\right\|\left\|U_n\right\|=\mathcal{O}_p\left(h_n^{-\frac{1}{2}}p_n^{\frac{1}{2}}\right), \quad (A.16)$$

by calculations used for bounding the first term on the RHS of (A.18), and so

$$\left\|w_n\right\|=\mathcal{O}_p\left\{\max\left(n^{-\frac{1}{2}}c_n^{\frac{1}{2}},h_n^{-\frac{1}{2}}p_n^{\frac{1}{2}}\right)\right\}. \quad (A.17)$$

Also $\bar{\bar{\Delta}}_n^{L}=n^{-1}[B_n,0]'[B_n,0]+n^{-1}[A_n,X_n]'[B_n,0]+n^{-1}[B_n,0]'[A_n,X_n]$ so we have

$$\left\|\bar{\bar{\Delta}}_n^{L}\right\|\leq n^{-1}\left\|\left[B_n,0\right]\right\|^2+2n^{-1}\left\|\left[A_n,X_n\right]'\left[B_n,0\right]\right\|. \quad (A.18)$$

The expectation of the first term on the RHS is bounded by $\sigma^2\sum_{j=1}^{p_n}tr\left(G_{jn}G_{jn}'\right)/n\leq Cp_n/h_n$, by Lemmas B.2 and B.3. Assumption 11 and Lemma B.1 imply that the square of the second term has expectation bounded by $\sigma^2\sum_{i=1}^{a_n}\mathbb{E}\left\|m_{in}\right\|^2\sum_{j=1}^{p_n}\left\|G_{jn}\right\|^2/n^2=\mathcal{O}\left(p_nk_n^2a_n/n\right)$, so that

$$\left\|\bar{\bar{\Delta}}_n^{L}\right\|=\mathcal{O}_p\left(\max\left\{h_n^{-1}p_n,n^{-\frac{1}{2}}p_n^{\frac{1}{2}}k_na_n^{\frac{1}{2}}\right\}\right). \quad (A.19)$$

From (A.11),

$$\left(1-\left\|\bar{L}_n^{-1}\right\|\left\|\bar{\bar{\Delta}}_n^{L}\right\|\right)\left\|\tilde{\theta}_{(n)}-\theta_{(n)}\right\|\leq\left\|\bar{L}_n^{-1}\right\|\left\|w_n\right\|. \quad (A.20)$$

By (A.19) the first factor on the LHS above converges in probability to one by (4.2) and Assumption 12. The first factor on the RHS of (A.20) is $\mathcal{O}_p(1)$ by Assumption 12, so

$$\left\|\tilde{\theta}_{(n)}-\theta_{(n)}\right\|=\mathcal{O}_p\left\{\max\left(n^{-\frac{1}{2}}c_n^{\frac{1}{2}},h_n^{-1}p_n^{\frac{1}{2}}\right)\right\}, \quad (A.21)$$

by (A.15) under Assumptions 1 and 13 but

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

by (A.17) under Assumption 1 only. These are both negligible by (4.2). □

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

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

From (A.21) and (A.15) the second term has modulus bounded by

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

while the modulus of the third term is bounded by

$$\left\|\tilde{\theta}_{(n)} - \theta_{(n)}\right\|^2 \left(\left\|\bar{L}_n\right\| + \left\|\bar{\bar{\Delta}}_n^L\right\|\right) = \mathcal{O}_p\left(\max\left\{n^{-1}c_n, h_n^{-2}p_n, n^{-\frac{1}{2}}h_n^{-1}p_n^{\frac{1}{2}}c_n^{\frac{1}{2}}\right\}\right),$$

using (A.19), (A.21) and Assumption 8. We conclude that

$$\tilde{\sigma}^2_{(n)} - \sigma^2 = \mathcal{O}_p\left(\max\left\{n^{-1}c_n, h_n^{-2}p_n, n^{-\frac{1}{2}}h_n^{-1}p_n^{\frac{1}{2}}c_n^{\frac{1}{2}}\right\}\right). \quad (A.23)$$

This is negligible by (4.4). □

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

$$\tau_n\alpha'\Psi_n\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right) = \tau_n\alpha'\Psi_n\bar{L}_n^{-1}\left(\bar{\bar{\Delta}}_n^L\left(\tilde{\theta}_{(n)} - \theta_{(n)}\right) + \bar{\Delta}_n^L L_n^{-1}w_n\right) + \tau_n\alpha'\Psi_n L_n^{-1}w_n. \quad (A.24)$$

We first show that first term on the RHS of (A.24) is negligible in probability. It has modulus bounded by $C\tau_n\|\Psi_n\|\left\|\bar{L}_n^{-1}\right\| \left(\left\|\bar{\bar{\Delta}}_n^L\right\|\left\|\tilde{\theta}_{(n)} - \theta_{(n)}\right\| + \left\|\bar{\Delta}_n^L\right\|\left\|L_n^{-1}\right\|\|w_n\|\right) = \mathcal{O}_p(\max\{c_n^{\frac{1}{2}}, n^{\frac{1}{2}}p_n^{\frac{1}{2}}/h_n\}\max\{p_n/h_n, p_n^{\frac{1}{2}}k_na_n^{\frac{1}{2}}/n^{\frac{1}{2}}, \|\bar{\Delta}_n^L\|\})$, by (A.15), (A.19), (A.21) and Assumption 12. This is $\mathcal{O}_p(\max\{p_n^{\frac{1}{2}}k_na_n^{\frac{1}{2}}c_n^{\frac{1}{2}}/n^{\frac{1}{2}}, p_nc_n^{\frac{1}{2}}/h_n, p_nk_n a_n^{\frac{1}{2}}/h_n, n^{\frac{1}{2}}p_n^{\frac{3}{2}}/h_n^2\})$, (using Assumption 15) which is negligible by (4.5) because

$$p_nk_n^2a_nc_n \le C\left(p_n^3k_n^4 + p_n^2k_n^5\right), \qquad p_n^2c_n \le Cp_n^3k_n^2,$$
$$np_n^3 = n^2p_n^2n^{-1}p_n$$
$$p_n^2k_n^2a_n = p_n^3k_n^2 + p_n^2k_n^3 = np_n\left(n^{-1}p_n^2k_n^2 + n^{-1}p_nk_n^3\right).$$

The second term on the RHS of (A.24) is

$$\tau_n\alpha'\Psi_nL_n^{-1}w_n = \tau_n\alpha'\Psi_nL_n^{-1}n^{-1}[A_n, X_n]'U_n + \tau_n\alpha'\Psi_nL_n^{-1}n^{-1}[B_n, 0]'U_n. \quad (A.25)$$

The modulus of the second term on the RHS of (A.25) is bounded by τ_n times

$$\|\alpha\|\,\|\Psi_n\|\,\left\|\bar{L}_n^{-1}\right\|\,\left\|n^{-1}[B_n, 0]'U_n\right\|. \quad (A.26)$$

The second factor on the RHS above is $\mathcal{O}\left(a_n^{\frac{1}{2}}\right)$, the third is $\mathcal{O}_p(1)$ by Assumption 12, and the fourth is $\mathcal{O}_p\left(p_n^{\frac{1}{2}}/h_n\right)$ by (A.14). Therefore (A.26) is $\mathcal{O}_p\left(p_n^{\frac{1}{2}}a_n^{\frac{1}{2}}/h_n\right)$ and so the second term on the RHS of (A.25) is $\mathcal{O}_p\left(n^{\frac{1}{2}}p_n^{\frac{1}{2}}/h_n\right)$. Under (4.5) this is $o_p(1)$, so we need to compute only the asymptotic distribution of the first term on the RHS in (A.25), which we write as $\sum_{i=1}^n f_{in}u_i$ with $f_{in} = \alpha'\Psi_nL_n^{-1}t_{in}/d_n$ and $d_n = \sigma\left\{\sum_{i=1}^n(\alpha'\Psi_nL_n^{-1}t_{in})^2\right\}^{\frac{1}{2}}$. As in the proof of Theorem 3.3, $\max_{1\le i\le n} f_{in}^2 \le n^{-1}\max_{1\le i\le n}\|t_{in}\|^2 \left\|L_n^{-1}\right\|^2\sigma^{-2}\overline{\eta}(L_n)$, for large enough n. This is negligible as $n \to \infty$ by Assumptions 8, 12 and 14. The asymptotic covariance matrix exists, and is positive definite, by Assumptions 8 and 12. The proof of consistency of the covariance matrix estimate is omitted. □

Appendix B. Technical lemmas

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

Proof. For any $i = 1, \ldots, p_n$, $\|G_{in}\|_R = \left\|S_n^{-1}W_{in}\right\|_R \le \left\|S_n^{-1}\right\|_R \|W_{in}\|_R \le C$ where the last inequality follows from Assumption 4. The claim for G_{in}' follows similarly. □

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

Proof. Denote by $w_{j,in}'$ the jth 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 $n \times 1$ vector with unity in the kth position and zeros elsewhere. Then $\left|w_{j,in}'S_n^{-1}e_{k,n}\right| \le \left\|w_{j,in}\right\|_R \left\|S_n^{-1}\right\|_R \left\|e_{k,n}\right\|_R = \mathcal{O}\left(h_n^{-1}\right)$. □

Lemma B.3. *Let Assumptions 2–4 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 $\mathcal{O}\left(h_n^{-1}\right)$ as $n \to \infty$.*

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

References

Anselin, L., 2001. Rao's score test in spatial econometrics. J. Statist. Plann. Inference 97, 113–139.
Arbia, G., 2006. Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence. Springer-Verlag, Berlin.
Badinger, H., Egger, P., 2013. Estimation and testing of higher-order spatial autoregressive panel data error component models. J. Geograph. Syst. 15, 453–489.
Berk, K.N., 1974. Consistent autoregressive spectral estimates. Ann. Statist. 2, 489–502.
Blommestein, H.J., 1983. Specification and estimation of spatial econometric models: A discussion of alternative strategies for spatial economic modelling. Region. Sci. Urban Econom. 13, 251–270.
Case, A.C., 1991. Spatial patterns in household demand. Econometrica 59, 953–965.
Case, A.C., 1992. Neighborhood influence and technological change. Region. Sci. Urban Econ. 22, 491–508.
Cliff, A.D., Ord, J.K., 1973. Spatial Autocorrelation. Pion, London.
Das, D., Kelejian, H.H., Prucha, I.R., 2003. Finite sample properties of estimators of spatial autoregressive models with autoregressive disturbances. Papers Region. Sci. 82, 1–121.
Davis, P.J., 1979. Circulant Matrices. Wiley Interscience, New York.
Horn, R.A., Johnson, C.R., 1985. Matrix Analysis. Cambridge University Press, New York.
Huang, J.S., 1984. The autoregressive moving average model for spatial analysis. Austral. J. Statist. 26, 169–178.
Huang, D., Anh, V.V., 1992. Estimation of spatial ARMA models. Austral. J. Statist. 34, 513–530.
Huber, P.J., 1973. Robust regression: asymptotics, conjectures and Monte Carlo. Ann. Statist. 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. J. Real Estate Fin. Econom. 17, 99–121.
Kelejian, H.H., Prucha, I.R., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. Internat. Econom. Rev. 40, 509–533.
Kelejian, H.H., Prucha, I.R., 2002. 2SLS and OLS in a spatial autoregressive model with equal spatial weights. Region. Sci. Urban Econ. 32, 691–707.
Kelejian, H.H., Prucha, I.R., Yuzefovich, Y.A., 2006. Estimation problems in models with spatial weighting matrices which have blocks of equal elements. J. Region. Sci. 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.
Pinkse, J., Slade, M.E., Brett, C., 2002. Spatial price competition: A semiparametric approach. Econometrica 70, 1111–1153.
Portnoy, S., 1984. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. Consistency. Ann. Statist. 12, 1298–1309.
Portnoy, S., 1985. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large; II. Normal approximation. Ann. Statist. 13, 1403–1417.
Robinson, P.M., 1979. Distributed lag approximation to linear time-invariant systems. Ann. Statist. 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. J. Econometrics 157, 6–17.
Scott, D.J., 1973. Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. Adv. Appl. Probab. 5, 119–137.
Yuzefovich, Y.A., 2003. Two Essays on Spatial Econometrics (Ph. D. thesis), University of Maryland.