ISSN 1755-5361

University of Essex

Department of Economics

Discussion Paper Series

No. 772 October 2015

Estimation of Spatial Autoregressions with Stochastic Weight Matrices

Abhimanyu Gupta

Estimation of spatial autoregressions with stochastic weight matrices

Abhimanyu Gupta*

Department of Economics, University of Essex, UK

October 9, 2015

Abstract

We examine a higher-order spatial autoregressive model with stochastic, but exogenous, spatial weight matrices. Allowing a general spatial linear process form for the disturbances that permits many common types of error specifications as well as potential 'long memory', we provide sufficient conditions for consistency and asymptotic normality of instrumental variables and ordinary least squares estimates. The implications of popular weight matrix normalizations and structures for our theoretical conditions are discussed. A set of Monte Carlo simulations examines the behaviour of the estimates in a variety of situations and suggests, like the theory, that spatial weights generated from distributions with 'smaller' moments yield better estimates. Our results are especially pertinent in situations where spatial weights are functions of stochastic economic variables.

JEL classifications: C21, C31, C36

Keywords: Spatial autoregression; stochastic spatial weights; spatial linear process; instrumental variables; ordinary least squares

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

1 Introduction

Spatial autoregressive (SAR) models, due to Cliff and Ord (1973), have recently become very popular in applied and theoretical research. These assume that, 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_{jn}, $j = 1, \ldots, p$, there exist unknown scalars $\lambda_{1n}, \lambda_{2n}, \ldots, \lambda_{pn}$ and an unknown $k \times 1$ vector β_n such that

$$y_n = \sum_{j=1}^{p} \lambda_{jn} W_{jn} y_n + X_n \beta_n + u_n, \tag{1.1}$$

where u_n is an $n \times 1$ vector of unobserved disturbances. The elements of W_{jn} measure distance between units, which may be geographic but in general can be (inverse) economic distances. W_{jn} are sometimes normalized in ways that make their elements dependent on n, e.g. row normalization, and X_n may contain spatial lags of basic explanatory variables. Both points imply triangular arrays and justify the n subscripting in (1.1), but the linear process type structure we permit for the disturbances also entails n subscripting on these. Subsequently we will drop n subscripts for brevity, but will occasionally remind the reader of the n-dependence of certain quantities.

While the majority of the literature on estimation and inference for SAR models, e.g. Kelejian and Prucha (1998, 1999, 2001), Lee (2002, 2003, 2004), Robinson (2010), Lee and Liu (2010), Su and Jin (2010), Lee and Yu (2013), Gupta and Robinson (2015), has assumed the W_j to be deterministic, examples abound that imply stochastically generated W_j. Most commonly a typical element of W_j is determined by economic variables that may themselves be stochastic. Conley and Ligon (2002) study cross-country spillovers in long-run growth rates using several distance measures. While one of them, geographic distance, is evidently fixed the other two measures, United Parcel Service shipping costs and airfare, are more difficult to justify to be fixed in repeated sampling. Both, at the very least, are subject to random shocks in the economic conditions of each pair of countries, among many other factors. Conley and Dupor (2003) take input-output relations as a measure of economic distance, and it is reasonable to imagine that these relations are stochastic and not fixed. In Yuzefovich (2003) spatial weight matrices are constructed using a variety of economic distances, e.g. trade between two countries and competition in borrowing from a common lender. These variables would generally be considered stochastic in econometric analyses that use such data. Another example is Baltagi, Fingleton, and Pirotte (2014), who construct a weight matrix using commuting frequencies between districts in the UK. Commuting frequencies between two districts depend heavily on macro and microeconomic factors that are stochastic, and therefore may be anticipated to be stochastic too. Souza (2015) considers a SAR model in which networks may form stochastically, captured by nonzero spatial weight matrix elements. Robinson (2008) briefly discusses a SAR with stochastic weights in the context of correlation testing.

In this paper we will justify instrumental variables (IV) and ordinary least squares (OLS) estimates for $(\lambda', \beta')'$ with stochastic but exogenous W_j. Asymptotic theory for IV estimates of SAR models was introduced first in Kelejian and Prucha (1998), and subsequently also studied by Lee (2003). IV is employed because the $W_j y$ are endogenous in general, but Lee (2002) demonstrated that OLS can deliver consistent and asymptotically normal estimates of SAR model parameters under certain circumstances, thus correcting a tendency to casually discard OLS as a suitable method for SAR estimation and inference. A more general treatment by Gupta and Robinson (2015) examined IV and OLS estimates for an increasing order version of (2.4) with $p, k \to \infty$ slowly with n, but with iid u elements.

Theory has been developed for estimation with endogenous W_j. Kelejian and Piras (2014) consider such a model and develop IV type estimates. Qu and Lee (2015) were critical of their restrictive assumptions, and instead use the near epoch dependence (NED) theory of Jenish and Prucha (2012) to establish consistency and asymptotic normality of estimates in a more general setting. However the intermediate case, with stochastic but exogenous W_j has received little theoretical attention. This case can cover situations of economic interest where spatial weights are generated by exogenous regressors, and can be examined in a very general framework that does not require NED process theory. For observations recorded at locations i and l the latter essentially requires the locations to be geographic (in the sense that they are in Euclidean space) due to a notion of dependence reducing as the distance between them increases. Thus it is not generally applicable to data that do not have a geographic interpretation. On the other hand, the SAR model has been considered to be particularly appealing because of its ability to handle data in general economic spaces, such as income space, where geographical interpretations may not be natural. If the locations indeed have a geographical interpretation, NED based theory provides powerful results and a greater ability to handle nonlinear models, cf. Xu and Lee (2015).

An additional innovation is that we allow for a general 'spatial linear process' structure in u, cf. Robinson and Thawornkaiwong (2012). They do not consider models with spatial lags in the dependent variables explicitly, nor do they provide theory for OLS estimates. In this sense we make a novel contribution to the literature also in the fixed W_j case that we formally cover.

The paper is organized as follows: Section 2 contains asymptotic theory for IV estimates, and Section 3 for ordinary least squares estimates. We discuss implications of common weight matrix normalizations and structures in Section 4, while Section 5 contains a small Monte Carlo simulation study. Section 6 concludes the paper. Two appendices contain theorem proofs and technical lemmas.

2 IV estimation

Let $Z = Z_n$ be a matrix of instruments with dimension $n \times p_1$, $p_1 \geq p$. Denoting $\theta = (\lambda', \beta')'$, define the IV estimate of θ as

$$\hat{\theta} = n^{-1}\bar{\bar{Q}}^{-1}\bar{\bar{K}}'\bar{J}^{-1}\left[Z,X\right]'y = \theta + \bar{\bar{Q}}^{-1}\bar{\bar{K}}'\bar{J}^{-1}q, \tag{2.1}$$

where $\bar{\bar{Q}} = \bar{\bar{Q}}_n = \bar{\bar{K}}'\bar{J}^{-1}\bar{\bar{K}}$ (dimension $p+k$) and $\bar{\bar{K}} = \bar{\bar{K}}_n = n^{-1}\left[Z,X\right]'\left[R,X\right]$ (dimension $(p_1+k)\times(p+k)$), with $R = [W_1y,\ldots,W_py]$, $\bar{J} = \bar{J}_n = n^{-1}\left[Z,X\right]'\left[Z,X\right]$ (dimension p_1+k), $q = n^{-1}\left[Z,X\right]'u$. Throughout the paper C denotes a generic positive constant, arbitrarily large but independent of n.

Assumption 1. *(1.1) holds with $u = u_n = (u_{1n},\ldots,u_{nn})'$, and*

$$u_{rn} = u_r = \sum_{l=1}^{\infty} c_{rl}\epsilon_l, \quad r = 1,\ldots,n, \quad n \geq 1, \tag{2.2}$$

where ϵ_l are scalar independent random variables with zero mean and unit variance, $c_{rl} = c_{rln}$, and satisfy

$$\sum_{l=1}^{\infty} c_{rl}^2 < C, \quad r = 1,\ldots,n, \quad n \geq 1. \tag{2.3}$$

Assumption 2. *The elements of W_j, $j = 1,\ldots,p$, are random variables that are uniformly $\mathcal{O}_p(1/h_n)$, as $n \to \infty$, with $h_n = h$ a bounded or divergent sequence that is bounded away from zero.*

Assumption 1 permits a wide variety of disturbance processes including SAR and spatial moving average (SMA), and implies that each u_i forms a triangular array. The square summability of linear process coefficients in (2.3) allows spatial 'long-memory'. Robinson and Thawornkaiwong (2012), who introduced this assumption, discuss it in detail. The time series literature commonly allows for martingale ϵ_l, but this is avoided in spatial settings as there may be no natural ordering available. Assumption 2 is an extension to stochastic weights of a commonly employed assumption that controls spatial weights, cf. Lee (2002, 2004), Gupta and Robinson (2015).

Assumption 3. *$P(S$ is non-singular$) = 1$, for all sufficiently large n.*

Assumption 3 ensures that a reduced form exists almost everywhere (a.e.) for y. Indeed, we can write (1.1) as

$$Sy = X\beta + u, \tag{2.4}$$

where $S = I_n - \sum_{j=1}^{p} \lambda_j W_j$, I_n denoting the s-dimensional identity matrix or, equivalently, $y = R\lambda + X\beta + u$. Assumption 3 implies that $y = S^{-1}X\beta + S^{-1}u$, a.e., so $R = A + B$ where $A = [G_1 X\beta, \ldots, G_p X\beta]$, $B = [G_1 u, \ldots, G_p u]$ and $G_j = W_j S^{-1}$ for $j = 1, \ldots, p$. Also define $\bar{K} = \bar{K}_n = n^{-1}[Z, X]'[A, X]$, $\bar{Q} = \bar{Q}_n = \bar{K}'\bar{J}^{-1}\bar{K}$ and introduce user chosen real numbers ζ_i, $i = 1, \ldots, 12$, such that $1 < \zeta_i < C$ for each i and $\zeta_j^{-1} + \zeta_{j+1}^{-1} = 1$ for odd j. The ζ_i will be used in Hölder inequalities in the proofs.

Assumption 4. *X, W_j and z_r are independent of ϵ_l, $r = 1, \ldots, n, j = 1, \ldots, p$, $l = 1, \ldots,$, where is z_r is the r-th column of Z'. Let $a_{rjn} = a_{rj}$ denote the (r, j)-th element of $[Z, X]$. Then*

$$\max_{1 \le r \le n,\ 1 \le j \le p_1 + k} \mathbb{E} \left| a_{rj} \right|^{2\zeta_1} < C, \tag{2.5}$$

and, as $n \to \infty$,

$$\bar{K} \xrightarrow{p} K,\ \bar{J} \xrightarrow{p} J, \tag{2.6}$$

where K, J are full-rank constant matrices, with J symmetric.

A consequence of Assumption 4 is that $\bar{Q} - Q = o_p(1)$, with $Q = K'J^{-1}K$. Condition (2.5) implies finite fourth moments for instruments and regressors. The requirement of the whole regressor matrix X being independent of the ϵ_l stems from the fact the instruments are typically constructed using linearly independent columns of $W_j^s X$, $j = 1, \ldots, p$, $s \ge 1$, cf. Kelejian and Prucha (1998). Evidently a given instrument vector then contains elements from different rows of X, as was noted by Gupta and Robinson (2015).

For a generic matrix F, define $\|F\|$ as the square root of the largest eigenvalue of FF' (the spectral norm), and $\|F\|_R$ as the largest absolute row-sum of F (the maximum row-sum norm). Denote

$$\chi_n = n^{-\frac{1}{2}} \max_{1\le j\le p} \left(\mathbb{E}\left\|W_j\right\|^{2\zeta_2\zeta_3}\right)^{\frac{1}{2\zeta_2\zeta_3}} \left(\mathbb{E}\left\|S^{-1}\right\|^{2\zeta_2\zeta_4}\right)^{\frac{1}{2\zeta_2\zeta_4}}. \tag{2.7}$$

Theorem 2.1. *Let Assumptions 1-4 hold and*

$$\chi_n \to 0, \text{ as } n \to \infty. \tag{2.8}$$

Then $\hat{\theta} - \theta = o_p(1)$.

The condition (2.8) limits the extent of spatial correlation. Note that (2.8) does not impose that $\|W_j\|$ or $\|S^{-1}\|$ have finite $2\zeta_2\zeta_3$-th or $2\zeta_2\zeta_4$-th moments, but allows these to grow with n. In this sense it is not as strong as may be imagined at first glance. We will look at a specific example with potentially unbounded moments in Section 4. There is an implication of being able to 'trade-off' the magnitude of moments of $\|W_j\|$ and $\|S^{-1}\|$ in χ_n and a_{ij} by choices of ζ_i, $i = 1, 2, 3, 4$. Some of the existing literature on SAR models with fixed weights imposes restrictions on $\|W_j\|_R$ or $\|S^{-1}\|_R$, but these are evidently stronger than those based on the spectral norm. Indeed, taking $\zeta_i = 2$, $i = 1, 2, 3, 4$, for simplicity, the inequality $\|F\|^2 \le \|F\|_R \|F'\|_R$ immediately implies $\chi_n = \mathscr{O}(\chi_{n,R})$ with

$$\chi_{n,R} = n^{-\frac{1}{2}} \max_{1\le j\le p} \left(\mathbb{E}\left\|W_j\right\|_R^8 \mathbb{E}\left\|W_j'\right\|_R^8 \mathbb{E}\left\|S^{-1}\right\|_R^8 \mathbb{E}\left\|S'^{-1}\right\|_R^8\right)^{\frac{1}{16}}.$$

Let $\mathbb{1}(\cdot)$ denote indicator function.

Assumption 5. $\sup_{l\ge1} \mathbb{E}\left(\epsilon_l^2 \mathbb{1}\left(|\epsilon_l| > \delta\right)\right) \to 0$, *as* $\delta \to \infty$.

Assumption 6. *With* a_r' *denoting the* r*-th row of* $[Z, X]$ *and* Φ *a positive definite* *(p.d.) constant matrix,*

$$n^{-1}\sum_{r,s=1}^{n}\sum_{l=1}^{\infty}c_{rl}c_{sl}a_r a_s' \xrightarrow{p} \Phi, \text{ as } n\to\infty, \tag{2.9}$$

$$n^{-1}\sup_{l\geq 1}\left\|\sum_{r=1}^{n}a_r c_{rl}\right\|^2 \xrightarrow{p} 0, \text{ as } n\to\infty. \tag{2.10}$$

Assumption 5 avoids identity of distribution for the ϵ_l, (2.9) simply asserts convergence of the covariance matrix of $n^{-\frac{1}{2}}[Z,X]'U$ while (2.10) is the form of the Lindeberg condition required for the central limit theorem.

Theorem 2.2. *Let Assumptions 1-6 and* (2.8) *hold. Then*

$$n^{\frac{1}{2}}\left(\hat{\theta}-\theta\right)\xrightarrow{d}\mathcal{N}\left(0, Q^{-1}K'J^{-1}\Phi J^{-1}KQ^{-1}\right), \text{ as } n\to\infty.$$

3 OLS estimation

Define the OLS estimate

$$\tilde{\theta} = n^{-1}\bar{\bar{L}}^{-1}\left[R,X\right]'y = \theta + \bar{\bar{L}}^{-1}w, \tag{3.1}$$

where $\bar{\bar{L}} = \bar{\bar{L}}_n = n^{-1}\left[R,X\right]'\left[R,X\right]$ (dimension $p+k$), $w = w_n = n^{-1}\left[R,X\right]'u$. Also define $\bar{L} = \bar{L}_n = n^{-1}[A,X]'[A,X]$. Assumption 2 needs to be strengthened to the following sufficient condition:

Assumption 7. *The ζ_i are chosen such that $\zeta_5\zeta_7 = 2\zeta_{11}$ and*

$$\max_{1\leq j\leq p}\mathbb{E}\left(\left\{\max_{1\leq r,s\leq n}|w_{rs,j}|\right\}^{2\zeta_{11}}\right) = \mathcal{O}\left(h^{-2\zeta_{11}}\right),$$

where $w_{rs,j}$ is the (r,s)-th element of W_j, $j=1,\ldots,p$.

This assumption implies $\max_{1\leq r,s\leq n,\,1\leq j\leq p}|w_{rs,j}| = \mathcal{O}_p\left(h^{-1}\right)$. Various bounds depending on the distribution of $w_{rs,j}$ exist in the extreme value literature for the expectation, but Assumption 7 ensures also that the familiar case with fixed $w_{rs,j} = \mathscr{O}(h^{-1})$ is formally covered. The restriction $\zeta_5\zeta_7 = 2\zeta_{11}$ is satisfied in the case where the Cauchy Schwarz inequality is used in place of the Hölder inequality, implying that $\zeta_i = 2$ for all i.

Assumption 8. *X, W_j are independent of ϵ_l, $l \geq 1$, $j = 1, \ldots, p$. Let $t_{rjn} = t_{rj}$ denote the (r, j)-th element of $[A, X]$. Then*

$$\max_{1 \leq r \leq n,\, 1 \leq j \leq p+k} \mathbb{E} \left| t_{rj} \right|^{2\zeta_1} < C, \tag{3.2}$$

and, as $n \to \infty$,

$$\bar{L} \xrightarrow{p} L, \tag{3.3}$$

where L is a constant, symmetric and non-singular matrix.

Define

$$\pi_n = h^{-\frac{1}{2}} \left(\max_{1 \leq j \leq p} \mathbb{E} \left\| W_j' \right\|_R^{\zeta_6\zeta_9} \right)^{\frac{1}{2\zeta_6\zeta_9}} \left(\mathbb{E} \left\| S'^{-1} \right\|_R^{\zeta_5\zeta_8} \right)^{\frac{1}{2\zeta_5\zeta_8}} \left(\mathbb{E} \left\| S'^{-1} \right\|_R^{\zeta_6\zeta_{10}} \right)^{\frac{1}{2\zeta_6\zeta_{10}}}. \tag{3.4}$$

Theorem 3.1. *Let Assumptions 1-3, 7, 8 hold and*

$$h^{-1} + \chi_n + \pi_n \longrightarrow 0 \text{ as } n \to \infty. \tag{3.5}$$

Then $\tilde{\theta} - \theta \xrightarrow{p} 0$, as $n \to \infty$.

For consistency of OLS estimates $h \to \infty$ is necessary even with deterministic W_j (cf. Lee (2002), Gupta and Robinson (2015)), and (3.5) strengthens the restrictions on spatial correlation relative to h.

Assumption 9. $\sup_{l \geq 1} \mathbb{E} \left(\epsilon_l^4 \mathbb{1} \left(|\epsilon_l| > \delta \right) \right) \to 0$, *as $\delta \to \infty$.*

This assumption seems hard to relax for a CLT. Indeed, even for (1.1) with $p = 1$, W_j fixed and no linear process structure, Lee (2002) required $\mathbb{E}\left|u_r\right|^{4+\eta} < C$, for some $\eta > 0$. Gupta and Robinson (2015) relaxed this slightly to $\mathbb{E}u_r^4 < C$, with increasing p, k but restricted themselves to iid u_r. Here we avoid identity of distribution of ϵ_l, and u_r, but require the uniform integrability of the ϵ_l^4 that Assumption 9 entails.

Assumption 10. *With t_r' denoting the r-th row of $[A, X]$ and Ψ a p.d. constant matrix,*

$$n^{-1} \sum_{r,s=1}^{n} \sum_{l=1}^{\infty} c_{rl} c_{sl} t_r t_s' \xrightarrow{p} \Psi > 0, \text{ as } n \to \infty, \tag{3.6}$$

$$n^{-1} \sup_{l \geq 1} \left\| \sum_{r=1}^{n} t_r c_{rl} \right\|^2 \xrightarrow{p} 0, \text{ as } n \to \infty. \tag{3.7}$$

Theorem 3.2. *Let Assumptions 1-3, 7-9 and (3.5) hold and*

$$h^{-1} n^{\frac{1}{2}} \left(\mathbb{E} \left\| S'^{-1} \right\|_R^{2\zeta_{12}} \right)^{\frac{1}{2\zeta_{12}}} \longrightarrow 0, \text{ as } n \to \infty. \tag{3.8}$$

Then

$$n^{\frac{1}{2}} \left(\tilde{\theta} - \theta \right) \xrightarrow{d} \mathcal{N} \left(0, L^{-1} \Psi L^{-1} \right), \text{ as } n \to \infty.$$

The proof requires some care to ensure that (2.3) does not need strengthening. Lee (2002) established that asymptotic normality of OLS relies not just on divergence of h, but sufficiently fast divergence, viz. $n^{\frac{1}{2}} = o(h)$. Condition (3.8) indicates the additional requirement that arises when the W_j are stochastic.

In the Cauchy Schwarz case with all $\zeta_i = 2$, we obtain

$$\chi_n = n^{-\frac{1}{2}} \left(\max_{1 \leq j \leq p} \mathbb{E} \left\| W_j \right\|^8 \mathbb{E} \left\| S^{-1} \right\|^8 \right)^{\frac{1}{8}}, \tag{3.9}$$

$$\pi_n = h^{-\frac{1}{2}} \left(\max_{1 \leq j \leq p} \mathbb{E} \left\| W_j' \right\|_R^4 \right)^{\frac{1}{8}} \left(\mathbb{E} \left\| S'^{-1} \right\|_R^4 \right)^{\frac{1}{4}}, \tag{3.10}$$

(3.8) becomes $h^{-1}n^{\frac{1}{2}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^4\right)^{\frac{1}{4}} \to 0$, as $n \to \infty$ and (2.5), (3.2) require finite fourth moments for the a_{rj} and t_{rj} respectively.

4 Normalizations of weight matrices

In this section we discuss the effect of various normalizations of the W_j on (2.8) and (3.5). For simplicity of exposition we will focus on the Cauchy Schwarz ($\zeta_i = 2$, all i) case, given in (3.9) and (3.10). Due to the $n^{-\frac{1}{2}}$ factor it is not necessary that the elements of W_j and S^{-1} have finite eighth moments for (2.8) to hold, but is it not sufficient either. Similarly due to the $h^{-\frac{1}{2}}$ factor and $h \to \infty$ finite fourth moments for elements of W_j' or S'^{-1} are neither necessary nor sufficient for (3.5) to hold. Both (2.8) and (3.5) can be compared to conditions imposed in Gupta and Robinson (2015) for deterministic W_j elements, where $\max_{1\leq j\leq p}\|W_j\| + \|S^{-1}\| \leq C$ was assumed, for which a necessary condition was boundedness of the elements of W_j, S^{-1}. Thus (2.8) and (3.5) may be viewed as controlling the spatial correlation asymptotically, and in particular controlling the magnitudes of the moments of the W_j and S^{-1} without them necessarily existing.

Various sufficient conditions may be found for (2.8) and (3.5) to hold. For example, suppose that

$$W_j = \left\|W_j^*\right\|^{-1} W_j^*, \tag{4.1}$$

for some matrices W_j^*. Then $\|W_j\|^s \leq 1$, for any $s > 0$, while

$$\left\|S^{-1}\right\| \leq \sum_{l=0}^{\infty}\left(\sum_{j=1}^{p}\|\lambda_j W_j\|\right)^l \leq C, \tag{4.2}$$

if

$$\sum_{j=1}^{p}|\lambda_j| < 1, \tag{4.3}$$

the power series existing a.e. under Assumption 3. The W_j can have a special 'single nonzero diagonal block' structure in some applications. Here there are $m_j \times m_j$ matrices V_j, $j = 1, \ldots, p$ and $\sum_{j=1}^{p} m_j = n$, such that each W_j has V_j as its j-th diagonal block and zeroes elsewhere. In this case Gupta and Robinson (2015) prove that

$$\max_{1 \le j \le p} |\lambda_j| < 1 \tag{4.4}$$

can replace the more general condition given in (4.3). Thus under (4.1), condition (2.8) is satisfied always if (4.3) holds for general W_j and (4.4) holds for W_j with 'single nonzero diagonal block' structure. Another popular normalization is row-normalization, where each row of W_j sums to 1. If the elements of W_j are non-negative, this implies that $\|W_j\|_R = 1$, and negligibility of $\chi_{n,R}$ follows if (4.3) holds in the general case and (4.4) holds in 'single nonzero diagonal block' case, as illustrated for the spectral norm. More generally, if $\|W_j\|_M \le C$, where $\|\cdot\|_M$ denotes a generic matrix norm, then $\mathbb{E}\|W_j\|_M^s \le C$, any $s > 0$. A sufficient condition for $\mathbb{E}\|S^{-1}\|_M \le C$ is that $\left\|\sum_{j=1}^{p} \lambda_j W_j\right\|_M$ has a moment generating function. Some normalization of the W_j is necessary to identify λ_j, and if these result in any of the favourable conditions listed above then the difference from the deterministic W_j case is lessened.

On the other hand, all types of normalizations are not economically justified. For instance Bell and Bockstael (2000) point out that row normalization is not justified in certain models with real estate data while Lee and Yu (2014) discuss problems in estimation that occur with row normalized weight matrices and present some simulation evidence using non normalized matrices. Thus the moment conditions implied by (2.8) and (3.5) will be different under various normalizations of W_j. The choice of normalization ultimately lies with the practitioner, but it seems like (4.1) is the most attractive option. Unlike row normalization it doesn't change the content of the spatial weight matrices because it preserves relative distances, and performs the task of stabilizing moments. Kelejian and Prucha (2010) provide an excellent discussion of normalizations and their implications, particularly for parameter spaces.

For another sufficient condition we focus on $p = 1$ (writing $W_1 = W$, $\lambda_1 = \lambda$), with $W = I_r \otimes B_m$, B_m a symmetric $m \times m$ matrix, so $n = rm$. This is the type of block diagonal weight matrix used by Case (1991, 1992), and sometimes referred to as a Balanced Group Interaction (BGI) setting, cf. Hillier and Martellosio (2013). It implies inter group independence for clustered data. Note that it does *not* have 'single nonzero diagonal block' structure. We take $r, m \to \infty$, which is a combination of 'increasing domain' and 'infill' asymptotics. Suppose that the elements of B_m are such that $\mathbb{E}\left\|B_m\right\|^8 = \mathscr{O}\left(m^{\xi}\right)$ and $\mathbb{E}\left\|\left(I_m - \lambda B_m\right)^{-1}\right\|^8 = \mathscr{O}\left(m^{\psi}\right)$, $\zeta, \psi \geq 0$. Then $\mathbb{E}\left\|W\right\|^8 = \mathscr{O}\left(m^{\xi}\right)$ and $\mathbb{E}\left\|S^{-1}\right\|^8 = \mathscr{O}(m^{\psi})$ due to their block diagonality with equal blocks, implying $\chi_n = \mathscr{O}\left(r^{-\frac{1}{2}} m^{\frac{\xi+\psi}{8} - \frac{1}{2}}\right) = o(1)$ always if $\xi + \psi \leq 4$ and if $m = o\left(r^{\frac{4}{\xi+\psi-4}}\right)$ when $\xi + \psi > 4$. This condition allows unbounded moments for spatial weight matrices.

5 Monte Carlo

Finite sample performance of IV and OLS estimates with fixed W_i has been examined before, cf. Gupta and Robinson (2015), and our aim in this section is rather different from previous literature. We seek information on how estimates behave as spatial weight moments change. Our design takes $p = 2$, $k = 2$ with $\lambda_1 = 0.2$, $\lambda_2 = 0.3$, $\beta_1 = 1$, $\beta_2 = 0.7$, X generated from $U(0,1)$ and u_i iid standard normal. The dependent variable is generated as $y = \lambda_1 W_1 y + \lambda_2 W_2 y + X\beta + u$. Our setup has W_j with the 'single non-zero diagonal block' structure discussed earlier. In particular we define $W_1 = diag\left[V_{1(m\times m)}, 0_{(m\times m)}\right], W_2 = diag\left[0_{(m\times m)}, V_{2(m\times m)}\right]$, so $n = 2m$, and generate V_j as iid replications using t_v and χ^2_v distributions with $v = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 100$, and $m = 48, 96, 144$. Instruments were selected

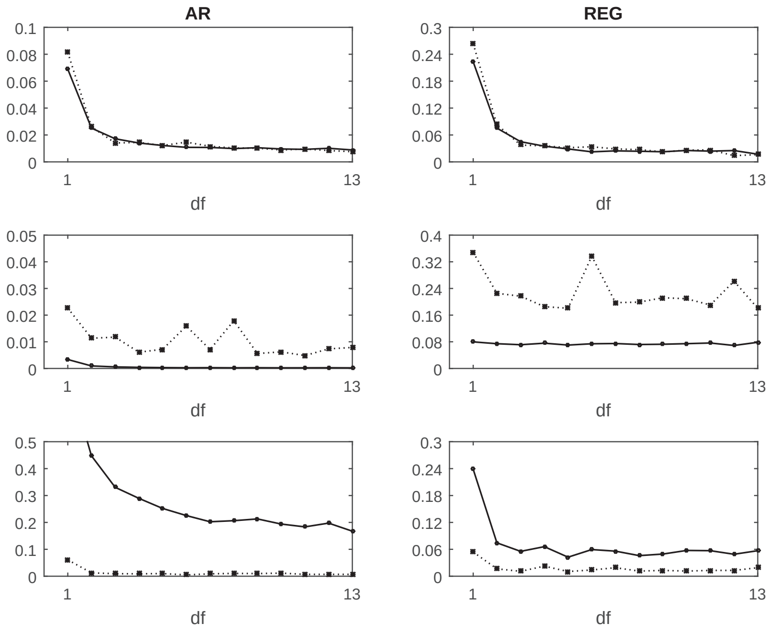

Figure 5.1: Monte Carlo bias (top row), variance (middle row) and size (bottom row) for IV estimates (dotted line) and OLS estimates (solid line) with $m = 48$, W_1 and W_2 generated from t_v, $v = 1, 2, \ldots, 12, 100$

to be $Z = [X, W_1X, W_2X]$.

Figures 5.1-5.6 display Monte Carlo bias (top row), variance (middle row) and size (bottom row) for IV (dotted line) and OLS (solid line) estimates of θ. The left columns display the results for λ_i and right columns for β_i, with average statistics reported in both cases. The horizontal axes in each figure correspond to values of v. Because of variations in the range over which the statistics change, it is unavoidable to employ different scales for the vertical axes across sub-figures and the reader should take care to ensure that they note the scale.

If $Z \sim t_v$ then $\mathbb{E}Z^l = v^{\frac{l}{2}} \prod_{i=1}^{\frac{l}{2}} (v - 2i)^{-1} (2i - 1)$, for even l such that $0 < l < v$ and $\mathbb{E}Z^l = 0$, for odd l such that $0 < l < v$. The expression for even l is evidently decreasing in v for given l, implying that Z has 'smaller' moments for larger v. On the other hand, for $Z \sim \chi^2_v$ we have $\mathbb{E}Z^l = 2^l \Gamma(v/2)^{-1} \Gamma(l + v/2)$, where $\Gamma(\cdot)$ is the Gamma function. This is evidently increasing in v, implying that Z has 'larger' moments as v increases. Our conditions suggest that IV and OLS estimates have better properties

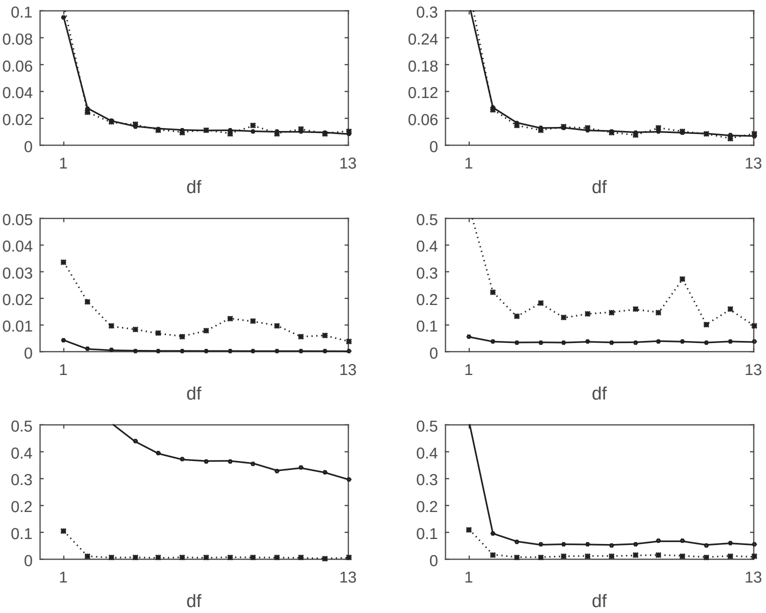

Figure 5.2: Monte Carlo bias (top row), variance (middle row) and size (bottom row) for IV estimates (dotted line) and OLS estimates (solid line) with $m = 96$, W_1 and W_2 generated from t_v, $v = 1, 2, \ldots, 12, 100$

for 'smaller' moments. Figures 5.1-5.6 seem to corroborate this. For spatial weights generated by t_v, the bias reduces dramatically as v increases for small values of v and then reduces more modestly as v increases for large values of v. The property holds for both autoregressive coefficients λ_i and regression coefficients β_i. The converse is true for weights generated using χ^2_v (Figures 5.4-5.6), where the bias reduces as v decreases, confirming our theoretical results. Variances follow a similar pattern, but the decrease is not monotonic for IV estimates using either t_v or χ^2_v, while for OLS estimates it is rather modest, with low variances throughout. The properties hold for all values of m. The spikes in variances of $\hat{\beta}_i$ in Figures 5.1 and 5.2 seem to be a result of randomness and appear less marked in the largest sample size of Figure 5.3.

Empirical sizes are to be compared to the nominal size of 5%. With t_v weights (Figures 5.1-5.3) IV estimates undersize both for λ_i and β_i. OLS estimates oversize for λ_i and do not even touch 20% but converge to the nominal size for β_i. The story is much the same for χ^2_v weights, except the approach to 5% for OLS estimates of

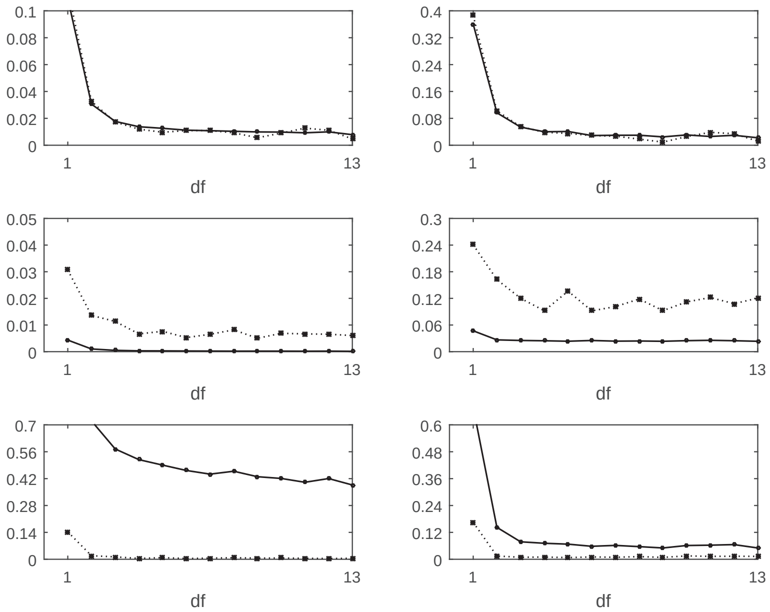

Figure 5.3: Monte Carlo bias (top row), variance (middle row) and size (bottom row) for IV estimates (dotted line) and OLS estimates (solid line) with $m = 144$, W_1 and W_2 generated from t_v, $v = 1, 2, \ldots, 12, 100$

regression coefficients is rather slow, while IV doesn't undersize as much as in the t_v^2 case, cf. Figures 5.4-5.6. The undersizing with IV estimates persists with χ_v^2 weights for the smallest v, while there is oversizing for large v. For moderate v the sizes are acceptable. This behaviour doesn't seem to be different across values of m.

Experiments were also conducted with exactly the same design as above but W_1 and W_2 subsequently row normalized. The stabilization effect of this row normalization is such that all statistics are acceptable, no matter which distribution generates the weights and the value of v. We opt not to include figures here because fundamentally these do not differ much from results already seen in the literature and our focus is rather different. We recall from Section 4 that row normalization changes the content of the spatial weight matrix and has been criticized, but division by spectral norm is less controversial and has the same stabilising effect in simulations carried out but not reported.

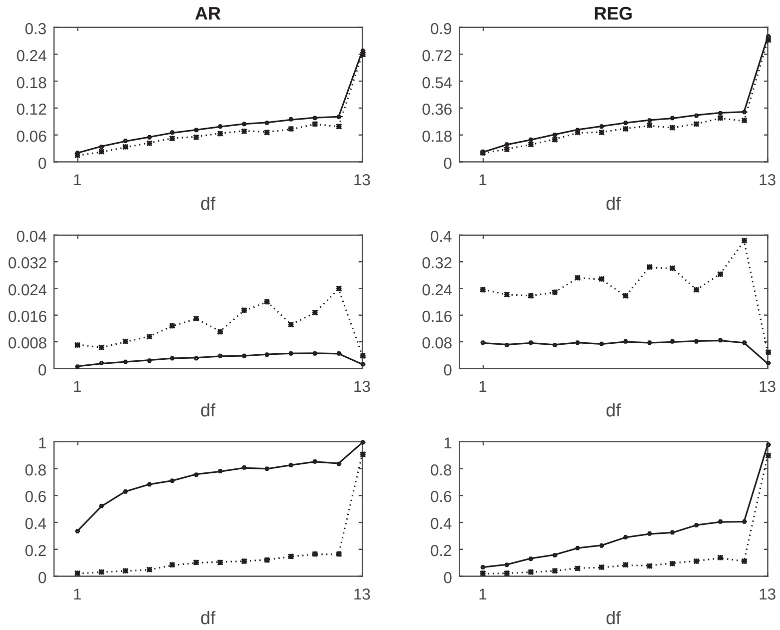

Figure 5.4: Monte Carlo bias (top row), variance (middle row) and size (bottom row) for IV estimates (dotted line) and OLS estimates (solid line) with $m = 48$, W_1 and W_2 generated from χ^2_v, $v = 1, 2, \ldots, 12, 100$

6 Conclusion and extensions

We examined IV and OLS estimates for the parameters of a SAR models with stochastic weight matrices and spatia linear process dependence in disturbances, finding that estimates perform better when spatial weights have smaller moments. We also discussed the implications of popular weight matrix normalizations on our conditions. In the dependent disturbances setup that we consider, heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimation is an important problem. With deterministic W_j this has been considered in the literature, cf. Kelejian and Prucha (2007, 2010) and Robinson and Thawornkaiwong (2012) for HAC estimation with SAR/SMA disturbances and disturbances satisfying Assumption 1 respectively. These approaches are straightforward to extend to the case with stochastic W_j and using them in practice requires no change in earlier techniques. It is also reasonable to anticipate the same lesson about spatial weight moments when constructing these

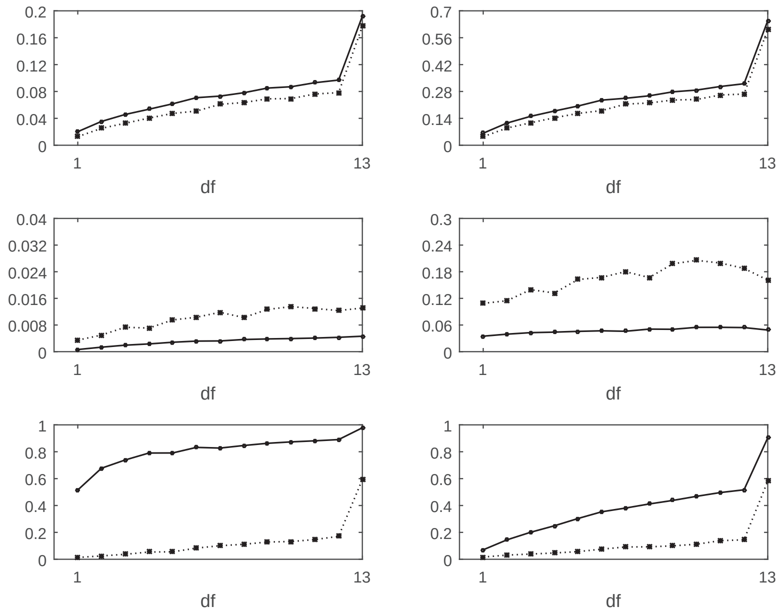

Figure 5.5: Monte Carlo bias (top row), variance (middle row) and size (bottom row) for IV estimates (dotted line) and OLS estimates (solid line) with $m = 96$, W_1 and W_2 generated from χ^2_v, $v = 1, 2, \ldots, 12, 100$

robust covariance matrix estimates.

Acknowledgements

I thank Peter Robinson for several comments that led to a much improved paper. I am also grateful to Nicolas Debarsy, Javier Hidalgo, Ingmar Prucha and Renata Rabovic for useful suggestions and discussions.

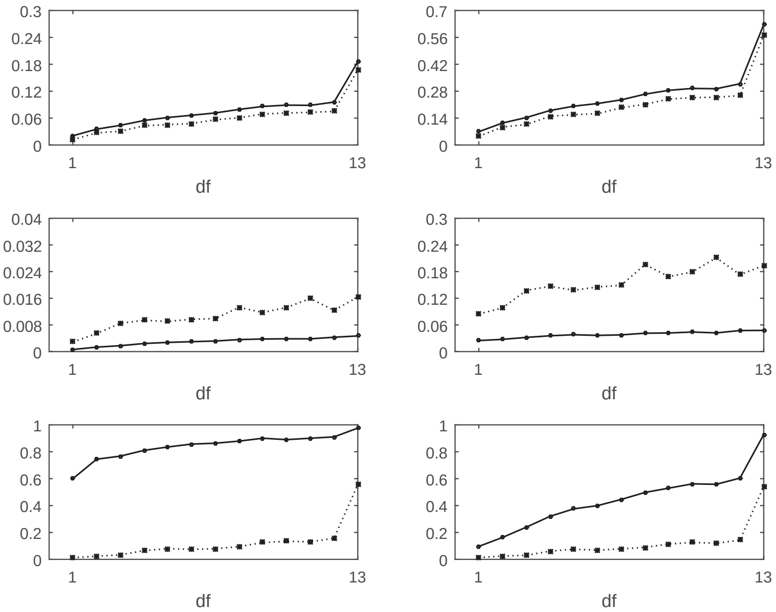

Figure 5.6: Monte Carlo bias (top row), variance (middle row) and size (bottom row) for IV estimates (dotted line) and OLS estimates (solid line) with $m = 144$, W_1 and W_2 generated from χ^2_v, $v = 1, 2, \ldots, 12, 100$

Appendices

A Proofs of theorems

For any matrices $\bar{F}$ and $\bar{\bar{F}}$ of equal dimension, we will write $\bar{\bar{\Delta}}^F = \bar{F} - \bar{\bar{F}}$.

Proof of Theorem 2.1. By (2.1) $\hat{\theta} - \theta = \bar{Q}^{-1}\bar{\bar{\Delta}}^Q \left(\hat{\theta} - \theta\right) - \bar{Q}^{-1}\bar{\bar{\Delta}}^{K\prime}\bar{J}^{-1}q + \bar{Q}^{-1}\bar{K}'\bar{J}^{-1}q$, so

$$\left(I_{p+k} - \bar{Q}^{-1}\bar{\bar{\Delta}}^Q\right)\left(\hat{\theta} - \theta\right) = -\bar{Q}^{-1}\bar{\bar{\Delta}}^{K\prime}\bar{J}^{-1}q + \bar{Q}^{-1}\bar{K}'\bar{J}^{-1}q. \tag{A.1}$$

We first show $\bar{\bar{\Delta}}^K = o_p(1)$. Write e_r for the $n \times 1$ vector with unity in the r-th position and zeroes elsewhere and b_i for the i-th column of $[Z, X]$. By the law of iterated expectations, the expectation of the square of a typical (i, j)-th element, $i, j = 1, \ldots, p_1 + k$, of $\bar{\bar{\Delta}}^K = n^{-1}[Z, X]'[B, 0]$ is

$$n^{-2}\mathbb{E}\left(b_i' G_j u u' G_j' b_i\right) = n^{-2}\mathbb{E}\left(\sum_{r,s=1}^{n} b_i' G_j e_r \mathbb{E}\left(u_{rn}u_{sn}\right) e_s' G_j' b_i\right). \tag{A.2}$$

Now $\mathbb{E}\left(u_{rn}u_{sn}\right) = \sum_{k,l=1}^{\infty} c_{rk}c_{sl}\mathbb{E}\left(\epsilon_k\epsilon_l\right) = \sum_{k=1}^{\infty} c_{rk}c_{sk} \leq \left(\sum_{k=1}^{\infty} c_{rk}^2\right)^{\frac{1}{2}}\left(\sum_{k=1}^{\infty} c_{sk}^2\right)^{\frac{1}{2}} \leq C$, by Assumption 5 and Cauchy-Schwarz inequality, so (A.2) is bounded by Cn^{-2} times

$$\mathbb{E}\left(b_i' G_j \sum_{r,s=1}^{n} e_r e_s' G_j' b_i\right) = \mathbb{E}\left(b_i' G_j \sum_{r=1}^{n} e_r e_r' G_j' b_i\right) = \mathbb{E}\left(b_i' G_j G_j' b_i\right). \tag{A.3}$$

The term inside the expectation on the far right is bounded by $\left\|b_i\right\|^2\left\|G_j\right\|^2$ so, by the Hölder inequality (for expectations), (A.3) is bounded by

$$\left(\mathbb{E}\left\|b_i\right\|^{2\zeta_1}\right)^{\frac{1}{\zeta_1}}\left(\mathbb{E}\left\|G_j\right\|^{2\zeta_2}\right)^{\frac{1}{\zeta_2}}. \tag{A.4}$$

The expectation inside parentheses in the first factor in (A.4) equals $\mathbb{E}\left(\sum_{r=1}^{n} a_{ri}^2\right)^{\zeta_1}$, which, by the Hölder inequality (for sums of real numbers) is bounded by

$$n^{\zeta_1\left(1-\frac{1}{\zeta_1}\right)}\sum_{r=1}^{n}\mathbb{E}\left|a_{ri}\right|^{2\zeta_1}=\mathscr{O}\left(n^{\zeta_1}\right), \tag{A.5}$$

by (2.5). The second factor in (A.4) is bounded by

$$\left\{\mathbb{E}\left(\|W_j\|^{2\zeta_2}\left\|S^{-1}\right\|^{2\zeta_2}\right)\right\}^{\frac{1}{\zeta_2}}\leq\left\{\left(\mathbb{E}\|W_j\|^{2\zeta_2\zeta_3}\right)^{\frac{1}{\zeta_3}}\left(\mathbb{E}\left\|S^{-1}\right\|^{2\zeta_2\zeta_4}\right)^{\frac{1}{\zeta_4}}\right\}^{\frac{1}{\zeta_2}}=\mathscr{O}\left(n\chi_n^2\right), \tag{A.6}$$

once again using Hölder's inequality. Combining (A.5) and (A.6) yields

$$\left(\mathbb{E}\|b_i\|^{2\zeta_1}\right)^{\frac{1}{\zeta_1}}\left(\mathbb{E}\|G_j\|^{2\zeta_2}\right)^{\frac{1}{\zeta_2}}=\mathscr{O}\left(n^2\chi_n^2\right), \tag{A.7}$$

whence Markov's inequality implies that

$$\left\|\bar{\bar{\Delta}}^K\right\|=\mathscr{O}_p\left(\chi_n\right)=o_p(1), \tag{A.8}$$

by (2.8). By (2.6) and (A.8),

$$\left\|\bar{\bar{\Delta}}^Q\right\|\leq\left\|\bar{\bar{\Delta}}^K\right\|\left\|\bar{J}^{-1}\right\|\left(\left\|\bar{\bar{\Delta}}^K\right\|+2\left\|\bar{K}\right\|\right)=\mathscr{O}_p\left(\left\|\bar{\bar{\Delta}}^K\right\|\right)=o_p(1). \tag{A.9}$$

Finally, the expectation of the square of a typical element of q is $n^{-2}\mathbb{E}\left(b_i'uu'b_i\right)=n^{-2}\mathbb{E}\left(\sum_{r,s=1}^{n}b_i'e_r\mathbb{E}\left(u_{rn}u_{sn}\right)e_s'b_i\right)\leq Cn^{-2}\mathbb{E}\|b_i\|^2\leq Cn^{-1}$, so

$$q=\mathscr{O}_p\left(n^{-\frac{1}{2}}\right). \tag{A.10}$$

Using (A.8), (A.9), (A.10) and (2.6) in (A.1) we obtain the desired result. □

Proof of Theorem 2.2. In view of (A.1), (A.8), (A.9) and (A.10) it suffices to show $n^{\frac{1}{2}}q\overset{d}{\longrightarrow}\mathcal{N}(0,\Phi)$. The proof now follows Robinson and Thawornkaiwong (2012) (henceforth RT), who modified one of Robinson and Hidalgo (1997). Write

$$d = d_n = n^{-\frac{1}{2}} \sum_{r=1}^{n} a_r u_r = n^{-\frac{1}{2}} \sum_{l=1}^{\infty} f_l \epsilon_l,$$

where $f_l = f_{ln} = \sum_{r=1}^{n} a_r c_{rl}$. By Lemma A1 of RT, there exists a sequence $N = N_n$, increasing in n without bound, such that $d - d_N = o_p(1)$, where $d_N = n^{-\frac{1}{2}} \sum_{l=1}^{N} f_l \epsilon_l$. Writing $E = E_n = n^{-1} \sum_{l=1}^{N} f_l f_l'$, again Lemma A1 of RT implies that $E \xrightarrow{p} \Phi$, by Assumption 6. Let $\alpha \in \mathbb{R}^{p+k}$ such that $\|\alpha\|^2 = 1$ and $c_N = \alpha' E^{-\frac{1}{2}} d_N$, $v_l = v_{ln} = n^{-\frac{1}{2}} \alpha' E^{-\frac{1}{2}} f_l$. Then $c_N = \sum_{l=1}^{N} v_l \epsilon_l$, and Assumption 6 implies that $\{f_l \epsilon_l, 1 \le l \le N\}$ is a martingale difference sequence for each $N \ge 1$. We show $c_N \xrightarrow{d} \mathcal{N}(0,1)$, conditional on X, z_r and W_j, $j = 1, \ldots, p$, which follows by Theorem 2 of Scott (1973) if, conditional on X, z_r and W_j, $j = 1, \ldots, p$, as $n \to \infty$,

$$\mathbb{E}\left(\sum_{l=1}^{N} v_l^2 \epsilon_l^2 | \epsilon_j, j < l\right) \xrightarrow{p} 1, \tag{A.11}$$

and for all $\xi > 0$,

$$\mathbb{E}\left(\sum_{l=1}^{N} v_l^2 \mathbb{E}\left(\epsilon_l^2 \mathbb{1}\left(|v_l \epsilon_l| > \xi\right) | z_r, X, W_1, \ldots, W_p\right)\right) \to 0. \tag{A.12}$$

The LHS of (A.11) equals 1, while the LHS of (A.12) is bounded by

$$\max_{1 \le l \le N} \mathbb{E}\left\{\epsilon_l^2 \mathbb{1}\left(\epsilon_l^2 > \frac{\xi^2}{\max\limits_{1 \le l \le N} v_l^2}\right)\right\} \mathbb{E}\left(\sum_{l=1}^{N} v_l^2\right).$$

By Assumption 5, it suffices to show that $\max_{1 \le l \le N} v_l^2 = o_p(1)$ as $n \to \infty$, as the rightmost factor equals 1. Now, $\max_{1 \le l \le N} v_l^2 \le n^{-1} \left\|E^{-\frac{1}{2}}\right\|^2 \sup_{l \ge 1} \left\|\sum_{r=1}^{n} a_r c_{rl}\right\|^2 = o_p(1)$ by Assumptions 4, 5 and 6. $\square$

Proof of Theorem 3.1. By (3.1) $\tilde{\theta} - \theta = \bar{L}^{-1}\bar{\bar{\Delta}}^L \left(\tilde{\theta} - \theta\right) + \bar{L}^{-1}w$, so

$$\left(I_{p+k} - \bar{L}^{-1}\bar{\bar{\Delta}}^L\right)\left(\tilde{\theta} - \theta\right) = \bar{L}^{-1}w. \tag{A.13}$$

Note that $\|w\| \leq \|n^{-1}[A, X]'u\| + \|n^{-1}[B, 0]'u\|$, where the first term on the RHS is readily shown to be negligible, as we deduced (A.10), but using (3.2) in Assumption 8 instead of (2.5) in Assumption 4 because here Z is replaced by A. Next $n^{-1}[B, 0]'u = o_p(1)$ by (3.5) and Lemma B.3, so that $w = o_p(1)$. It remains to prove that $\bar{\bar{\Delta}}^L = o_p(1)$, for which first note that

$$\left\|\bar{\bar{\Delta}}^L\right\| \leq n^{-1}\|B\|^2 + 2n^{-1}\left\|[A, X]'[B, 0]\right\|. \tag{A.14}$$

The first term on the RHS is $\mathscr{O}_p(\pi_n)$, by the proof of Lemma B.3, and is negligible by (3.5). The second term on the RHS is bounded exactly like $\left\|\bar{\bar{\Delta}}^K\right\| = n^{-1}[Z, X]'[B, 0]$ in the proof of Theorem 2.1, but again using (3.2) in Assumption 8 instead of (2.5) in Assumption 4 because here Z is replaced by A, see also the proof of Theorem 4.1 in Gupta and Robinson (2015). Therefore

$$\left\|\bar{\bar{\Delta}}^L\right\| = \mathscr{O}_p(\chi_n) = o_p(1), \tag{A.15}$$

and the theorem is proved.

□

Proof of Theorem 3.2. We claim that it is sufficient to prove

$$n^{-\frac{1}{2}}[A, X]'u \xrightarrow{d} \mathcal{N}(0, \Psi), \tag{A.16}$$

for which, in view of (A.13), (A.15) and $w = n^{-1}[A, X]'u + n^{-1}[B, 0]'u$ it is enough to show that

$$n^{-\frac{1}{2}}[B,0]'u = o_p(1). \tag{A.17}$$

To show (A.17), we can exploit Assumption 9 to obtain a sharper bound for $[B,0]'u$ than the one used in the proof of Theorem 3.1, where Assumption 5 sufficed. Indeed, by Lemma B.4, $n^{-\frac{1}{2}}[B,0]'u = \mathscr{O}_p\left(n^{\frac{1}{2}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{2\zeta_{12}}} h^{-1}\right) = o_p(1)$, by (3.8). The proof of (A.16) follows exactly as the proof of asymptotic normality of $n^{\frac{1}{2}}q$ in Theorem 2.2, and we omit the details, noting only that here A replaces Z, and Assumption 10 replaces Assumption 6. □

B Lemmas

Lemma B.1. *Under the conditions of Theorem 3.1, the expectation of an absolute typical element of $G_j'G_j$ is $\mathscr{O}\left(\pi_n^2\right)$, uniformly in j.*

Proof. For $r,s = 1,\ldots,n$, a typical absolute element of $G_j'G_j$ is $\left|g_{r,j}'G_je_s\right| = \left|e_s'G_j'g_{r,j}\right|$, where $g_{r,j}'$ is the r-th row of G_j'. Using Hölder's inequality as before, this has expectation bounded by

$$\left(\mathbb{E}\left\|g_{r,j}\right\|_R^{\zeta_5}\right)^{\frac{1}{\zeta_5}}\left(\mathbb{E}\left\|G_j'\right\|_R^{\zeta_6}\right)^{\frac{1}{\zeta_6}} \leq \left(\mathbb{E}\left\|g_{r,j}\right\|_R^{\zeta_5}\right)^{\frac{1}{\zeta_5}}\left(\mathbb{E}\left(\left\|W_j'\right\|_R^{\zeta_6}\left\|S'^{-1}\right\|_R^{\zeta_6}\right)\right)^{\frac{1}{\zeta_6}}. \tag{B.1}$$

Consider the first factor on the RHS of (B.1). $g_{r,j}$ has elements $w_{s,j}'S^{-1}e_r = e_r'S'^{-1}w_{s,j}$, where $w_{s,j}'$ is the s-th row of W_j, $s = 1,\ldots,n$, so this factor is

$$\begin{aligned}
\left(\mathbb{E}\left(\max_{1\leq s\leq n}\left|w_{s,j}'S^{-1}e_r\right|^{\zeta_5}\right)\right)^{\frac{1}{\zeta_5}} &\leq \left(\mathbb{E}\left(\max_{1\leq s\leq n}\left\|w_{s,j}\right\|_R^{\zeta_5}\left\|S'^{-1}\right\|_R^{\zeta_5}\right)\right)^{\frac{1}{\zeta_5}} \\
&\leq \left(\mathbb{E}\left(\max_{1\leq s\leq n}\left\|w_{s,j}\right\|_R^{\zeta_5\zeta_7}\right)\right)^{\frac{1}{\zeta_5\zeta_7}}\left(\mathbb{E}\left(\left\|S'^{-1}\right\|_R^{\zeta_5\zeta_8}\right)\right)^{\frac{1}{\zeta_5\zeta_8}} \\
&= \left(\mathbb{E}\left(\max_{1\leq r,s\leq n}\left|w_{rs,j}\right|^{\zeta_5\zeta_7}\right)\right)^{\frac{1}{\zeta_5\zeta_7}}\left(\mathbb{E}\left(\left\|S'^{-1}\right\|_R^{\zeta_5\zeta_8}\right)\right)^{\frac{1}{\zeta_5\zeta_8}}
\end{aligned}$$

$$= \mathscr{O}\left(h^{-1}\left(\mathbb{E}\left(\left\|S'^{-1}\right\|_R^{\zeta_5\zeta_8}\right)\right)^{\frac{1}{\zeta_5\zeta_8}}\right), \tag{B.2}$$

by the Hölder inequality and Assumption 7. The second factor on the RHS of (B.1) is bounded by

$$\left(\max_{1\le j\le p}\mathbb{E}\left\|W_j'\right\|_R^{\zeta_6\zeta_9}\right)^{\frac{1}{\zeta_6\zeta_9}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{\zeta_6\zeta_{10}}\right)^{\frac{1}{\zeta_6\zeta_{10}}}, \tag{B.3}$$

by another application of Hölder's inequality, whence the claim follows from (B.1), (B.2), (B.3) and the definition of π_n. □

Lemma B.2. *Under the conditions of Theorem 3.2, the expectation of the absolute product of two typical elements of* G_j *is* $\mathscr{O}\left(\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{\zeta_{12}}}h^{-2}\right)$, *uniformly in* j.

Proof. For $p,q,r,s = 1,\ldots,n$, the expectation of the absolute product of typical elements of G_j is

$$\mathbb{E}\left|w_{r,j}'S^{-1}e_s w_{p,j}'S^{-1}e_q\right| \le \mathbb{E}\left(\max_{1\le r\le n}\left\|w_{r,j}\right\|_R^2\left\|S'^{-1}\right\|_R^2\right), \tag{B.4}$$

which is bounded by $\left(\mathbb{E}\left(\max_{1\le r\le n}\left\|w_{r,j}\right\|_R^{2\zeta_{11}}\right)\right)^{\frac{1}{\zeta_{11}}}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{\zeta_{12}}}$, whence the result follows by Assumption 7 because $\left\|w_{r,j}\right\|_R = \max_{1\le s\le n}\left|w_{rs,j}\right|$. □

Lemma B.3. *Under the conditions of Theorem 3.1,*

$$n^{-1}[B,0]'u = \mathscr{O}_p\left(\pi_n\right).$$

Proof. First note that $\mathbb{E}\left\|u\right\|^2 = \sum_{r=1}^{n}\sum_{l=1}^{\infty}c_{rl}^2 = \mathscr{O}(n)$, by (2.3), so $\|n^{-1}[B,0]'u\| = \mathscr{O}_p\left(n^{-\frac{1}{2}}\left\|B\right\|\right)$ by Markov's inequality. Next

$$\mathbb{E}\left\|B\right\|^2 \le \mathbb{E}\left(trB'B\right) = \mathscr{O}\left(\max_{1\le j\le p}\mathbb{E}\left(u'G_j'G_ju\right)\right),$$

the RHS being

$$\mathscr{O}_p\left(\pi_n^2\sum_{r,s=1}^{n}\sum_{j,l=1}^{\infty}c_{rj}c_{sl}\mathbb{E}\left(\epsilon_j\epsilon_l\right)\right) = \mathscr{O}_p\left(\pi_n^2\sum_{r,s=1}^{n}\sum_{l=1}^{\infty}\left(c_{rl}^2+c_{sl}^2\right)\right)$$
$$= \mathscr{O}_p\left(\pi_n^2\sum_{r=1}^{n}\sum_{l=1}^{\infty}c_{rl}^2\right) = \mathscr{O}_p\left(n\pi_n^2\right),$$

by Lemma B.1, the inequality $|ab| \leq \left(a^2+b^2\right)/2$ for real numbers a, b and Assumption 1. The claim follows by Markov's inequality, □

Lemma B.4. *Under the conditions of Theorem 3.2,*

$$n^{-1}[B,0]'u = \mathscr{O}_p\left(\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{2\zeta_{12}}}h^{-1}\right).$$

Proof. Write $g_{rs,j}$ for a typical element of G_j, $r,s=1,\ldots,n$. It is sufficient to evaluate $\mathbb{E}\left(n^{-1}u'G_ju\right)^2 = n^{-2}\sum_{r,s,t,v=1}^{n}\mathbb{E}\left(u_ru_su_tu_v\right)\mathbb{E}\left(g_{rs,j}g_{tv,j}\right)$, with $j=1,\ldots,p$, and then use Markov's inequality. By Assumption 1 and Lemma B.2 the RHS is

$$n^{-2}\sum_{r,s,t,v=1}^{n}\sum_{j,k,l,m=1}^{\infty}c_{rj}c_{sk}c_{tl}c_{vm}\mathbb{E}\left(\epsilon_j\epsilon_k\epsilon_l\epsilon_m\right)\mathbb{E}\left(g_{rs,j}g_{tv,j}\right)$$
$$= \mathscr{O}\left(n^{-2}\left(\mathbb{E}\left\|S'^{-1}\right\|_R^{2\zeta_{12}}\right)^{\frac{1}{\zeta_{12}}}h^{-2}\left[\sum_{r,s,t,v=1}^{n}\sum_{j=1}^{\infty}\mathbb{E}\left(\epsilon_j^4\right)c_{rj}c_{sj}c_{tj}c_{vj}+\right.\right.$$
$$\left.\left.+\sum_{r,s,t,v=1}^{n}\sum_{j,k=1}^{\infty}\left(c_{rj}c_{sj}c_{tk}c_{vk}+c_{rj}c_{sk}c_{tj}c_{vk}+c_{rj}c_{sk}c_{tk}c_{vj}\right)\right]\right). \tag{B.5}$$

By Assumption 9 and the ℓ_p norm inequality, the first sum inside square brackets in (B.5) is bounded in absolute value by a constant times

$$\sum_{r,s,t,v=1}^{n}\sum_{j=1}^{\infty}\left|c_{rj}c_{sj}c_{tj}c_{vj}\right| \leq C\sum_{r,s,t,v=1}^{n}\sum_{j=1}^{\infty}\left(c_{rj}^2c_{sj}^2+c_{tj}^2c_{vj}^2\right)$$
$$\leq C\sum_{r,s,t,v=1}^{n}\sum_{j=1}^{\infty}\left(c_{rj}^4+c_{sj}^4+c_{tj}^4+c_{vj}^4\right)$$

$$\leq \; C \sum_{r=1}^{n} \sum_{j=1}^{\infty} c_{rj}^4 \leq C \sum_{r=1}^{n} \left(\sum_{j=1}^{\infty} c_{rj}^2 \right)^2 \leq Cn.$$

Now consider the first product inside parentheses in the second sum inside square brackets in (B.5). By similar techniques this is bounded in absolute value by

$$\sum_{r,s,t,v=1}^{n} \sum_{j,k=1}^{\infty} |c_{rj}c_{sj}| \, |c_{tk}c_{vk}| = \left(\sum_{r,s=1}^{n} \sum_{j=1}^{\infty} |c_{rj}c_{sj}| \right)^2 \leq \; C \left(\sum_{r,s=1}^{n} \sum_{j=1}^{\infty} \left(c_{rj}^2 + c_{sj}^2 \right) \right)^2$$

$$\leq \; C \left(\sum_{r=1}^{n} \left(\sum_{j=1}^{\infty} c_{rj}^2 \right) \right)^2 \leq Cn^2.$$

The remaining two products inside parentheses in the second sum inside square brackets in (B.5) are similarly shown to be $\mathscr{O}\left(n^2\right)$. We have established that the term inside square brackets in (B.5) is $\mathscr{O}\left(n^2\right)$, whence the claim follows. □

References

Baltagi, B. H., B. Fingleton, and A. Pirotte (2014). Spatial lag models with nested random effects: An instrumental variable procedure with an application to English house prices. *Journal of Urban Economics 80*, 76–86.

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

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. and J. K. Ord (1973). *Spatial Autocorrelation.* London: Pion.

Conley, T. G. and B. Dupor (2003). A spatial analysis of sectoral complementarity. *Journal of Political Economy 111*, 311–352.

Conley, T. G. and E. Ligon (2002). Economic distance and cross-country spillovers. *Journal of Economic Growth 7*, 157–187.

Gupta, A. and P. M. Robinson (2015). Inference on higher-order spatial autoregressive models with increasingly many parameters. *Journal of Econometrics 186*, 19–31.

Hillier, G. and F. Martellosio (2013). Properties of the maximum likelihood estimator in spatial autoregressive models. Mimeo.

Jenish, N. and I. R. Prucha (2012). On spatial processes and asymptotic inference under near-epoch dependence. *Journal of Econometrics 170*, 178 – 190.

Kelejian, H. H. and G. Piras (2014). Estimation of spatial models with endogenous weighting matrices, and an application to a demand model for cigarettes. *Regional Science and Urban Economics 46*, 140–149.

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

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

Kelejian, H. H. and I. R. Prucha (2001). On the asymptotic distribution of the Moran *I* test statistic with applications. *Journal of Econometrics 104*, 219–257.

Kelejian, H. H. and I. R. Prucha (2007). HAC estimation in a spatial framework. *Journal of Econometrics 140*, 131–154.

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

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. and X. Liu (2010). Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. *Econometric Theory 26*, 187–230.

Lee, L. F. and J. Yu (2013). Near unit root in the spatial autoregressive model. *Spatial Economic Analysis, Special Issue: Contributions to Spatial Econometrics: A Festschrift in Honour of Harry H. Kelejian 8*, 314–351.

Lee, L. F. and J. Yu (2014). Efficient GMM estimation of spatial dynamic panel data models with fixed effects. *Journal of Econometrics 180*, 174–197.

Qu, X. and L. F. Lee (2015). Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics 184*, 209–232.

Robinson, P. M. (2008). Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics 147*, 5–16.

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

Robinson, P. M. and F. J. Hidalgo (1997). Time series regression with long-range dependence. *The Annals of Statistics 25*, 77–104.

Robinson, P. M. and S. Thawornkaiwong (2012). Statistical inference on regression with spatial dependence. *Journal of Econometrics 167*, 521–542.

Scott, D. J. (1973). Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Advances in Applied Probability 5*, 119–137.

Souza, P. C. L. (2015). Estimating network effects without network data. Mimeo.

Su, L. and S. Jin (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics 157*, 18–33.

Xu, X. and L. F. Lee (2015). A spatial autoregressive model with a nonlinear transformation of the dependent variable. *Journal of Econometrics 186*, 1–18.

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