

Large-sample inference on spatial dependence

P. M. Robinson

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP29/08

Large-Sample Inference on Spatial Dependence

P.M. ROBINSON

Department of Economics, London School of Economics,
Houghton Street, London WC2A 2AE, UK
E-mail: p.m.robinson@lse.ac.uk

September 1, 2008

Summary: We consider cross-sectional data that exhibit no spatial correlation, but are feared to be spatially dependent. We demonstrate that a spatial version of the stochastic volatility model of financial econometrics, entailing a form of spatial autoregression, can explain such behaviour. The parameters are estimated by pseudo Gaussian maximum likelihood based on log-transformed squares, and consistency and asymptotic normality are established. Asymptotically valid tests for spatial independence are developed.

Keywords: Spatial dependence, Parameter estimation, Asymptotic theory, Independence testing.

1 INTRODUCTION

The possibility of cross-sectional dependence haunts much analysis of econometric data. Rules of statistical inference based on cross-sectional or panel data frequently assume independence of observables or, more likely, of unobservable disturbances. These rules are typically invalidated if there is actually dependence. On the other hand, the modelling of cross-sectional dependence is hugely complicated by the usual lack of any natural ordering over the cross-section. This is in contrast to time series data, where dependence between variables at different times is frequently modelled as a function of their time difference, as is appropriate under stationarity. In the standard setting of equal spacing across time, elegant statistical procedures result, due to the ability to exploit the Toeplitz structure of the covariance matrix. When there is unequal spacing, matters are considerably complicated, but nevertheless there is still a natural ordering, and the ability to regard the observations as arising from sampling from, say, a continuous time process, and so it is still clear how one might proceed, for example under Gaussianity where it suffices to consider the mean and covariance structure.

The absence of any such natural ordering poses more of a dilemma. One might consider pairwise covariances or correlations, but without data replication these cannot be consistently estimated in the absence of suitable structure. When there is a spatial context, however, progress may be possible. Lattice data provide the simplest extension of time series, entailing equal spacing across two or more dimensions. Though there is a lack of a single obvious ordering, and difficulties due to end effects and in data simulation, and there are natural ways of extending stationary time series models, and the corresponding rules of statistical inference.

However lattice observations arise infrequently in econometrics. With spatio-temporal data, there may well still be regular spacing across time, but observations in geographical space are more likely to be at irregular intervals, in both dimensions, for example when these are identified with capital cities of countries. Matters are further complicated if observations are to be interpreted as aggregates across administrative regions of irregular shapes. And in many situations geographical distances may not be the most relevant measures. More generally, pairwise "economic distances" can be postulated, possibly varying with reversal of direction. Much of the methodology of spatial econometrics has pursued this setting, focussing on models of "spatial autoregressive" type, which depend on the availability of such measures of distance between each pair of observations.

In the spatial econometrics literature, dependence has been usually taken to be synonymous with correlation (an exception being Brett and Pinkse (1997)). On the other hand other areas of econometric research stress the distinction between these concepts. In particular, financial time series that exhibit lack of serial correlation frequently contain evidence of dependence, for example in serial correlation of second moments, and considerable activity has been devoted to modelling such phenomena. In the present paper we propose a model that combines features of a stochastic volatility model of financial econometrics with the spatial autoregressive model, deriving asymptotic statistical theory for estimates of its parameters, and present related tests for lack of dependence, justifying their asymptotic validity. The reference to stochastic volatility modelling is not necessarily intended to imply particular relevance to financial data, and the econometric and statistical literature on nonlinearity and testing for dependence covers other possible applications also. Generally in non-Gaussian settings, dependence and correlation have different meanings and there may be interest in nonlinear modelling and independence testing, our particular approach being specialised but parsimonious. One could also think of our model as applying not to raw data but uncorrelated but not necessarily independent innovations, possible spatial correlation and explanatory variables having been previously taken care of in a conventional fashion.

The following section describes the model and illustrates its ability to describe both dependence and lack of correlation. Section 3 describes the parameter estimates. Section 4 establishes their consistency. Section 5 establishes their asymptotic normality. Section 6 examines tests that might be used to test the hypothesis of spatial independence. Some concluding remarks are offered in Section 7. Proofs are left to an Appendix.

2 A MODEL FOR SPATIAL DEPENDENCE

Introduce first sequences $\eta_i, \varepsilon_i, i = 1, 2, \dots$, of zero-mean independent and identically distributed (iid) random variables, having finite variances σ_η^2 and σ_ε^2 , respectively, and such that η_i and ε_j are independent for all i, j . Next define

$$\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)', \quad (2.1)$$

and an $n \times n$ weight matrix W , having zero diagonal elements, and for a scalar ρ define

$$S(\rho) = I - \rho W, \quad (2.2)$$

I being the $n \times n$ identity matrix. For some $\rho_0 \in (-1, 1)$, put $S_0 = S(\rho_0)$, and define the $n \times 1$ vector

$$\zeta = (\zeta_1, \dots, \zeta_n)' \quad (2.3)$$

by the spatial autoregressive model (see Cliff and Ord (1973))

$$S_0 \zeta = \varepsilon. \quad (2.4)$$

Though the ζ_i are unobservable, as are the η_i , we observe

$$x_i = \eta_i e^{\alpha_0 + \zeta_i}, \quad i = 1, \dots, n, \quad (2.5)$$

where α_0 is a scalar constant. This is a model analogous to the stochastic volatility model of financial econometrics of Taylor (1986), and extensively developed and applied since. We will develop asymptotic theory of parameter estimates as $n \rightarrow \infty$, in which case all non-diagonal elements of W (as well as its dimension), can vary as n increases, especially as some normalization restriction is typically placed on W (see Assumption A2 below). In this case the ζ_i , and correspondingly the x_i , form triangular arrays. However, as is common, for notational convenience we suppress reference to this.

When $\rho_0 = 0$, the x_i are clearly independent. For $\rho_0 \neq 0$ spatial independence is lost, though there is still no spatial correlation, as we now demonstrate. For the purpose of the immediately following argument assume also that the ε_i have a moment generating function; thus so also do the ζ_i . We have

$$E(x_i) = 0, \quad (2.6)$$

$$E(x_i^2) = \sigma_\eta^2 e^{2\alpha_0} E(e^{2\zeta_i}) < \infty, \quad (2.7)$$

and, for $i \neq j$,

$$E(x_i x_j) = 0, \quad (2.8)$$

so that the x_i are uncorrelated. Thus they exhibit no spatial correlation. However, consider now

$$x_i^2 = \eta_i^2 e^{2\alpha_0 + 2\zeta_i}. \quad (2.9)$$

Then, for $i \neq j$, using also (2.7),

$$Cov(x_i^2, x_j^2) = \sigma_\eta^4 e^{4\alpha_0} \left\{ E\left(e^{2(\zeta_i + \zeta_j)}\right) - E(e^{2\zeta_i}) E(e^{2\zeta_j}) \right\}. \quad (2.10)$$

For $\rho_0 \neq 0$, the ζ_i are not independent, so the expression in braces is non-zero, and thus the x_i^2 do exhibit spatial correlation. Of course, other nonlinear functions of x_i , such as $|x_i|^\lambda$ for any $\lambda > 0$, will also do so, but for simplicity we focus on squares.

Parametric functional expressions are available on making distributional assumptions on the ε_i . For example, suppose they are Gaussian. From (2.4), we can write

$$\zeta_i = t'_i \varepsilon, \quad (2.11)$$

where t'_i is the i -th row of

$$T_0 = S_0^{-1}. \quad (2.12)$$

Thus

$$E(e^{2\zeta_i}) = e^{2\|t_i\|^2 \sigma_\varepsilon^2} \quad (2.13)$$

and for $i \neq j$

$$E(e^{2(\zeta_i + \zeta_j)}) = E(e^{2(t_i + t_j)' \varepsilon}) = e^{2\|t_i + t_j\|^2 \sigma_\varepsilon^2} \quad (2.14)$$

where for any real matrix A , $\|A\|$ denotes the square root of the largest eigenvalue of $A'A$. Thus

$$Cov(x_i^2, x_j^2) = \sigma_\eta^4 e^{4\alpha_0 + 2\|t_i\|^2 + 2\|t_j\|^2} (e^{4t'_i t_j \sigma_\varepsilon^2} - 1), \quad i \neq j. \quad (2.15)$$

Note that when $\rho_0 = 0$ the elements of t_i are all zero except for the i -th, so indeed (2.15) then reduces to zero, but otherwise it is generally non-zero.

3 PSEUDO-MAXIMUM LIKELIHOOD ESTIMATION

Though W is chosen by the practitioner, the parameters α_0 , ρ_0 , σ_η^2 and σ_ε^2 are generally unknown. Given further distributional assumptions they can be estimated by maximum likelihood, but this is a computationally onerous procedure, and asymptotic statistical properties are difficult to derive. Instead, we consider a Gaussian pseudo-likelihood procedure based on logs. Denote

$$y_i = \log x_i^2, \quad \beta_0 = E \log \eta_i^2, \quad \nu_i = \log \eta_i^2 - \beta_0. \quad (3.1)$$

We deduce from (2.9)

$$y_i = 2\alpha_0 + \beta_0 + \nu_i + 2\zeta_i. \quad (3.2)$$

Define also

$$\mu_0 = 2\alpha_0 + \beta_0, \quad (3.3)$$

$$\xi_i = \nu_i + 2\zeta_i, \quad (3.4)$$

and write

$$y_i = \mu_0 + \xi_i, \quad (3.5)$$

or in vector form

$$y = \mu_0 \ell + \xi, \quad (3.6)$$

where

$$\xi = (\xi_1, \dots, \xi_n)', \quad y = (y_1, \dots, y_n)' \quad (3.7)$$

and ℓ is the $n \times 1$ vector of 1's. We could rewrite (3.6) as

$$S(\rho_0)y = \mu_0 S(\rho_0)\ell + S(\rho_0)\nu + 2\varepsilon, \quad (3.8)$$

where $\nu = (\nu_1, \dots, \nu_n)'$, so that the y_i follow a kind of constrained spatial autoregressive moving average. Note that if the row sums of W are normalized the intercept term in (3.8) becomes $\mu_0(1 - \rho_0)\ell$.

The y_i have mean μ_0 , and y has covariance matrix

$$\sigma_0^2 I + \sigma_\varepsilon^2 T_0 T_0', \quad (3.9)$$

where $\sigma_0^2 = \text{Var}(\nu_i)$. Though ξ in general is non-Gaussian, we will apply Gaussian estimation procedures to (3.6). This means that parameters must be identifiable from first and second moments of the y_i . We can only identify μ_0 from $E(y_i)$. Also, (3.9) reduces to $(\sigma_0^2 + \sigma_\varepsilon^2)I$ when $\rho_0 = 0$, whence we cannot identify both σ_0^2 and σ_ε^2 . Though our work is motivated by the possible presence of spatial dependence, there is interest also testing for spatial independence, i.e. $\rho_0 = 0$, so we restrict to a parsimonious version of the model, in which we constrain $\sigma_\varepsilon^2 = 1$. Since T_0 depends only on the parameter ρ_0 , we are left with three unknown parameters, summarized in the vector

$$\theta_0 = (\mu_0, \sigma_0^2, \rho_0)'. \quad (3.10)$$

Let μ, σ^2, ρ be any admissible values, of $\mu_0, \sigma_0^2, \rho_0$, and define

$$\theta = (\mu, \sigma^2, \rho)', \quad (3.11)$$

$$T(\rho) = S(\rho)^{-1}, \quad (3.12)$$

$$\Omega(\sigma^2, \rho) = \sigma^2 I + T(\rho)T(\rho)'. \quad (3.13)$$

The Gaussian pseudo-maximum likelihood estimate (PMLE) of θ_0 is defined as

$$\hat{\theta} = \arg \min_{\theta \in \Theta} Q(\theta), \quad (3.14)$$

where

$$Q(\theta) = \frac{1}{n} \log \det \{ \Omega(\sigma^2, \rho) \} + \frac{1}{n} (y - \mu \ell)' \Omega(\sigma^2, \rho)^{-1} (y - \mu \ell), \quad (3.15)$$

and Θ is a compact subset of $\mathbb{R} \times (0, \infty) \times (-1, 1)$, in particular

$$\Theta = \Theta_\mu \times \Theta_{\sigma^2} \times \Theta_\rho, \quad (3.16)$$

where

$$\Theta_\mu = [c_1, c_2], \quad \Theta_{\sigma^2} = [c_3, c_4], \quad \Theta_\rho = [c_5, c_6], \quad (3.17)$$

where $-\infty < c_1 < c_2 < \infty$, $0 < c_3 < c_4 < \infty$, $-1 < c_5 < c_6 < 1$.

4 CONSISTENCY OF ESTIMATES

We introduce the following assumptions.

Assumption A1 *The ε_i, η_i are iid with zero means, ε_i is independent of η_j , for all i, j , and $E(\varepsilon_i^4) < \infty, E(\nu_i^4) < \infty$.*

The identity of distribution aspect, and indeed the independence, can be somewhat relaxed, but we opt for simplicity.

Assumption A2 *For all $n, \|W\| \leq 1$.*

The spatial autoregression literature imposes various conditions on W . One is that the row sums of W are normalized to 1, which implies that 1 is an eigenvalue of WW' (and of W if W is symmetric), so that Assumption A2 requires that there be no other eigenvalue that is larger in absolute value. When all the elements of W are non-negative, 1 is then also the maximum row sum norm of W (see Horn and Johnson, 1988 p.295). In a sense, Assumption A2 is costless because some normalization is necessary in order to identify ρ_0 , and indeed $\|W\| = 1$ not only achieves this but is natural from a stability perspective because then

$$\|T(\rho)\| \leq \sum_{j=1}^{\infty} |\rho|^j \|W\| \leq \sum_{j=0}^{\infty} |\rho|^j = (1 - |\rho|)^{-1}, \quad (4.1)$$

which is finite for all $\rho \in (-1, 1)$. It is possible to impose more general conditions on W , such as ones on $T(\rho)$ that are uniform in ρ (see e.g. Lee (2004)), but we prefer in this respect to separate requirements on W from other aspects.

Define $H = \Omega^{-\frac{1}{2}}(\sigma^2, \rho) \Omega(\sigma_0^2, \rho_0) \Omega^{-\frac{1}{2}}(\sigma^2, \rho)$ (where we employ the positive definite square root), and then

$$\begin{aligned} r(\sigma^2, \rho) &= \frac{1}{n} \text{tr} \left\{ \Omega(\sigma^2, \rho)^{-1} \Omega(\sigma_0^2, \rho) \right\} - \frac{1}{n} \log \det \left\{ \Omega(\sigma^2, \rho)^{-1} \Omega(\sigma_0^2, \rho) \right\} - 1 \\ &= \frac{1}{n} \text{tr} \{H\} - \frac{1}{n} \log \det \{H\} - 1 \\ &= \frac{1}{n} \sum_{j=1}^n (\lambda_j - \log \lambda_j - 1), \end{aligned} \quad (4.2)$$

where the λ_j are eigenvalues of H .

Assumption A3 *For any $\delta > 0$,*

$$\liminf_{n \rightarrow \infty} \inf_{\{\|\sigma^2 - \sigma_0, \rho - \rho_0\| > \delta\} \cap \{\Theta_{\sigma^2} \times \Theta_{\rho}\}} r(\sigma^2, \rho) > 0. \quad (4.3)$$

Because H is positive definite the $\lambda_j, j = 1, \dots, n$, are positive, and for all j , the j -th summand in (4.2) is non-negative, and positive when $\lambda_j \neq 1$. Of course

$\lambda_j = 1$ for all n only when $\Omega(\sigma^2, \rho) = \Omega(\sigma_0^2, \rho_0)$, so that Assumption A3 is an identifiability condition. It seems difficult in general to reduce it to something more comprehensible (see also the identifiability assumption employed by Lee (2004) in his asymptotic theory for the Gaussian PMLE of spatial autoregression).

Assumption A4 $\theta_0 \in \Theta$.

Theorem 1 *Let Assumptions A1-A4 hold. Then*

$$\hat{\theta} \rightarrow_p \theta_0, \quad \text{as } n \rightarrow \infty. \quad (4.4)$$

5 ASYMPTOTIC NORMALITY OF ESTIMATES

Using the consistency just established, and additional conditions, we proceed to establish asymptotic normality of the Gaussian PMLE.

Define the 3×3 symmetric matrices A and B , as follows. Write $\Omega_0 = \Omega(\sigma_0^2, \rho_0)$. The (i, j) -th element of A is a_{ij} , where $a_{12} = a_{13} = 0$ and

$$a_{11} = 2 \lim_{n \rightarrow \infty} \frac{1}{n} \ell' \Omega_0^{-1} \ell, \quad (5.1)$$

$$a_{22} = \lim_{n \rightarrow \infty} \frac{1}{n} \text{tr} \{ \Omega_0^{-2} \}, \quad (5.2)$$

$$a_{23} = 8 \lim_{n \rightarrow \infty} \frac{1}{n} \text{tr} \{ \Omega_0^{-2} T_0 W T_0 T_0' \}, \quad (5.3)$$

$$a_{33} = 32 \lim_{n \rightarrow \infty} \text{tr} \{ \Omega_0^{-1} T_0 W T_0 T_0' \Omega_0^{-1} T_0 (W T_0 + T_0' W') T_0' \}, \quad (5.4)$$

and the (i, j) -th element of B is b_{ij} , where $b_{11} = 2\alpha_{11}$ and

$$b_{12} = \lim_{n \rightarrow \infty} \frac{1}{n} E [\ell' \Omega_0^{-1} \xi \text{tr} \{ (\xi \xi' - \Omega_0) \Omega_0^{-1} \}], \quad (5.5)$$

$$b_{13} = 16 \lim_{n \rightarrow \infty} \frac{1}{n} E [\ell' \Omega_0^{-1} \xi \text{tr} \{ (\xi \xi' - \Omega_0) \Omega_0^{-1} T_0 W T_0 T_0' \Omega_0^{-1} \}], \quad (5.6)$$

$$b_{22} = E \lim_{n \rightarrow \infty} \frac{1}{n} E [\text{tr}^2 \{ (\xi \xi' - \Omega_0) \Omega_0^{-1} \}], \quad (5.7)$$

$$b_{23} = - \lim_{n \rightarrow \infty} \frac{8}{n} E [\text{tr} \{ (\xi \xi' - \Omega_0) \Omega_0^{-1} \} \text{tr} \{ (\xi \xi' - \Omega_0) \Omega_0^{-1} T_0 W T_0 T_0' \}], \quad (5.8)$$

$$b_{33} = 64 \lim_{n \rightarrow \infty} \frac{1}{n} E [\text{tr}^2 \{ (\xi \xi' - \Omega_0) \Omega_0^{-1} T_0 W T_0 T_0' \Omega_0^{-1} \}], \quad (5.9)$$

where we assume:

Assumption A5 *The matrices A and B exist, and are finite and non-singular.*

We also impose standard additional conditions for a central limit theorem.

Assumption A6 θ_0 is an interior point of Θ_0 .

Assumption A7 For some $\delta > 0$

$$E \left(|\varepsilon_i|^{4+\delta} + |\nu_i|^{4+\delta} \right) < \infty. \quad (5.10)$$

Theorem 2 Let Assumptions A1-A3 and A5-A7 hold. Then as $n \rightarrow \infty$,

$$n^{\frac{1}{2}} \left(\hat{\theta} - \theta_0 \right) \rightarrow_d \mathcal{N} \left(0, A^{-1} B A^{-1} \right).$$

6 TESTING FOR SPATIAL INDEPENDENCE

Theorem 2 can be applied to set confidence regions for θ_0 or its individual elements, but it is also a basis for testing hypotheses. One of leading interest is

$$H_0 : \rho_0 = 0, \quad (6.1)$$

which in the setting of our model is equivalent to independence of the x_i .

We present first a result which is largely, but not strictly, a corollary of Theorem 2. In this connection, we introduce

Assumption A8

$$\text{tr} \{W(W + W')\} \rightarrow \infty, \quad \text{as } n \rightarrow \infty. \quad (6.2)$$

Notice that Assumption A3 would require, under H_0 , that $\text{tr} \{W(W + W')\}$ increase at rate n . We could indeed have relaxed conditions for Theorem 2 to permit a slower rate, which would have been reflected in the convergence rate of $\hat{\rho}$ (see also Lee, 2004). Assumption A2 implies that $\text{tr} \{W(W + W')\} = O(n)$, and thus that no faster rate would be possible.

For notational convenience define also

$$\sigma_\xi^2 = \sigma_0^2 + 4, \quad (6.3)$$

which is the variance of ξ_i under H_0 .

Theorem 3 Let Assumptions A1, A2, A6 and A8 hold. Then under H_0 ,

$$n^{\frac{1}{2}} \hat{\mu} / \sigma_\xi, \quad n^{\frac{1}{2}} \left(\hat{\sigma}^2 - \sigma_0^2 \right) / \left\{ E \left(\xi_1^4 \right) - \sigma_\xi^4 \right\}^{\frac{1}{2}}, \quad 4 \left[\text{tr}^{\frac{1}{2}} \{W(W + W')\} \right] \hat{\rho} / \sigma_\xi^2 \quad (6.4)$$

converge in distribution as $n \rightarrow \infty$ to independent standard normal variates.

Theorem 3 motivates the test statistic

$$s_1 = \frac{4tr^{\frac{1}{2}} \{W(W + W')\} \hat{\rho}}{\hat{\sigma}^2 + 4}. \quad (6.5)$$

A simpler one is

$$s_2 = \frac{\tilde{\xi}' W \tilde{\xi}}{\tilde{\sigma}_{\tilde{\xi}}^2 tr^{\frac{1}{2}} \{W(W + W')\}}, \quad (6.6)$$

where

$$\tilde{\xi} = (\tilde{\xi}_1, \dots, \tilde{\xi}_n)', \quad \tilde{\sigma}_{\tilde{\xi}}^2 = \frac{1}{n} \tilde{\xi}' \tilde{\xi}, \quad \tilde{\xi}_i = y_i - \tilde{\mu}, \quad \tilde{\mu} = n^{-1} \sum_{i=1}^n y_i. \quad (6.7)$$

Of course (6.6) is merely a standard statistical test for lack of spatial correlation but applied to the $\tilde{\xi}_i$: see Moran (1950), Pinkse (1999); also Robinson (2007) for a more general class.

Theorem 4 *Let Assumptions A1, A2, A6 and A8 hold. Then under H_0 , s_1 and s_2 both converge in distribution as $n \rightarrow \infty$ to standard normal variates.*

Both s_1 and s_2 can be used in one- or two-sided tests based on standard normal critical regions. A Pitman argument indicates that tests that reject for large positive (negative) values of s_1/s_2 have power against local, at rate $n^{-\frac{1}{2}}$, positive (negative) alternatives to H_0 . We can think of s_1^2 and s_2^2 as pseudo-Wald and pseudo-score statistics, respectively. A pseudo-log-likelihood-ratio test can also be developed, but for brevity, and as there is no one-sided version of it, we omit the details.

7 CONCLUDING REMARKS

We have established consistency and asymptotic normality of parameter estimates of a simple model that can explain spatial dependence in observations x_i in the absence of spatial correlation in the x_i , and has also presented related asymptotically justified tests for spatial dependence. One straightforward extension of the model would allow spatial correlation of observables, and perhaps include also explanatory variables; then test spatially uncorrelated inputs for spatial independence. It would also be worth examining both higher-order asymptotic properties, and finite-sample properties, of our various statistics. Higher-order asymptotics should be possible at least under Gaussian assumptions, on ξ_i and ν_i but presents a substantial additional challenge. There seems to be no higher-order asymptotic theory yet for even most basic statistics based on spatial weight matrices, and in the general statistical literature there is relatively little work covering implicitly defined estimates. Some finite-sample theory would be possible for s_2 under H_0 and Gaussianity of ξ_i and ν_i because it is then merely a ratio of quadratic forms of independent Gaussian variates. On a more mundane level, Monte Carlo simulations can also provide information about finite-sample properties, but given the limited nature of proposals

for modelling and inference of spatial dependence without spatial correlation, and for testing for spatial dependence as distinct from spatial correlation, it would be appropriate first to develop some further models, estimates and tests, with which ours can be compared. Other parametric models, including spatial moving averages, and spatial autoregressive moving averages, can be considered for our ξ , and along with the tests for independence suggested by such models there is considerable scope for developing nonparametric tests for independence in addition to those of Brett and Pinkse (1997). Bearing in mind the range of nonparametric independence tests available for time series data, there are clearly many possibilities in spatial settings.

ACKNOWLEDGEMENTS

This research was supported by ESRC Grant RES-062-23-0036.

REFERENCES

- Brett, C, and J. Pinkse (1997). These taxes are all over the map! A test for spatial independence of municipal tax rates in British Columbia. *International Regional Science Review* 20, 131-151.
- Cliff, A.D., and J.K. Ord (1973). *Spatial Autocorrelation*. London: Pion.
- Horn, R.A., and C.R. Johnson (1988). *Matrix Analysis*. Cambridge: Cambridge University Press.
- Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.
- Moran, P.A.P. (1950). Notes on continuous stochastic phenomena. *Biometrika* 37, 17-23.
- Pinkse, J. (1999). Asymptotic properties of Moran and related tests and testing for spatial correlation in probit models. Preprint, University of British Columbia.
- Robinson, P.M. (2007). Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics*, forthcoming.
- Taylor, S. (1986). *Modelling Financial Time Series*. New York: John Wiley.

APPENDIX: PROOFS OF THEOREMS

Proof of Theorem 1: We have

$$\begin{aligned}
 Q(\theta) &= \frac{1}{n} \log \det \{ \Omega(\sigma^2, \rho) \} + \frac{1}{n} (\mu - \mu_0)^2 \ell' \Omega(\sigma^2, \rho)^{-1} \ell \\
 &\quad - \frac{2}{n} (\mu - \mu_0) \ell' \Omega(\sigma^2, \rho)^{-1} \xi + \frac{1}{n} \xi' \Omega(\sigma^2, \rho)^{-1} \xi. \quad (\text{A.1})
 \end{aligned}$$

Then write

$$Q(\theta) - Q(\theta_0) = u(\theta) + v(\theta), \quad (\text{A.2})$$

where

$$u(\theta) = \frac{2}{n} (\mu_0 - \mu) \ell' \Omega(\sigma^2, \rho)^{-1} \xi + \frac{1}{n} \text{tr} \left[\left\{ \Omega(\sigma^2, \rho)^{-1} - \Omega(\sigma_0^2, \rho_0)^{-1} \right\} \{ \xi \xi' - \Omega(\sigma_0^2, \rho_0) \} \right], \quad (\text{A.3})$$

$$v(\theta) = \frac{1}{n} (\mu - \mu_0)^2 \ell' \Omega(\sigma^2, \rho)^{-1} \ell + w(\sigma^2, \rho). \quad (\text{A.4})$$

By Assumption A4 and a standard kind of argument for consistency of implicitly-defined extremum estimates, it thus suffices to show that

$$\sup_{\theta \in \Theta} |u(\theta)| \rightarrow_p 0, \quad \text{as } n \rightarrow \infty, \quad (\text{A.5})$$

and, for all $\delta_\theta > 0$, $\chi > 0$

$$\lim_{n \rightarrow \infty} \inf_{\{\|\theta - \theta_0\| > \delta_\theta\} \cap \Theta} v(\theta) > 0. \quad (\text{A.6})$$

To prove (A.6) we first consider the contribution to $u(\theta)$ from the first term in (A.4). This is uniformly $o_p(1)$ if

$$\frac{1}{n} \ell' \Omega(\sigma^2, \rho)^{-1} \xi \rightarrow_p 0, \quad \text{uniformly in } \Theta. \quad (\text{A.7})$$

We first show pointwise convergence, for any $\theta \in \Theta$. The left side of (A.7) has mean zero and variance

$$\begin{aligned} n^{-2} \ell' \Omega(\sigma^2, \rho)^{-1} \Omega(\sigma_0^2, \rho_0) \Omega(\sigma^2, \rho)^{-1} \ell &\leq n^{-2} \left(\frac{\ell' \ell}{\sigma^4} \right) \|\Omega(\sigma_0^2, \rho_0)\| \\ &\leq C n^{-1} \left(\sigma_0^2 + \|T(\rho_0)\|^2 \right). \end{aligned} \quad (\text{A.8})$$

But from (4.1)

$$\|T(\rho_0)\| \leq C, \quad (\text{A.9})$$

where C denotes throughout a generic finite constant. Thus pointwise convergence is established. The uniform convergence follows from an equicontinuity argument, as follows. Consider a neighbourhood \mathcal{N} of any, σ_*^2, ρ_* , such that $\mathcal{N} \subset \Theta_{\sigma^2} \times \Theta_\rho$. We have

$$\begin{aligned} &\sup_{(\sigma^2, \rho) \in \mathcal{N}} \left| \frac{\ell' \left\{ \Omega(\sigma^2, \rho)^{-1} - \Omega(\sigma_*^2, \rho_*)^{-1} \right\} \xi}{n} \right| \\ &\leq \left(\frac{\xi' \xi}{n} \right)^{\frac{1}{2}} \sup_{(\sigma^2, \rho) \in \mathcal{N}} \left\{ \frac{\ell' \left\{ \Omega(\sigma^2, \rho)^{-1} - \Omega(\sigma_*^2, \rho_*)^{-1} \right\}^2 \ell}{n} \right\}^{\frac{1}{2}}. \end{aligned} \quad (\text{A.10})$$

Now $E\xi'\xi/n = \text{tr}\{\Omega(\sigma_0^2, \rho_0)\}/n \leq C$, whereas the expression in braces is bounded by

$$\left\|\Omega(\sigma^2, \rho)^{-1}\right\|^2 \left\|\Omega(\sigma_*^2, \rho_*)^{-1}\right\|^2 \left\|\Omega(\sigma^2, \rho) - \Omega(\sigma_*^2, \rho_*)\right\|^2. \quad (\text{A.11})$$

The first two factors are bounded uniformly on \mathcal{N} , while the last one is

$$\left\|(\sigma^2 - \sigma_*^2)I + 4(TT' - T_*T'_*)\right\|^2, \quad (\text{A.12})$$

where $T_* = T(\rho_*)$. Now, with $S = S(\rho)$, $S_* = S(\rho_*)$,

$$TT' - T_*T'_* = T_*T'_*(S_*S'_* - SS')TT', \quad (\text{A.13})$$

where

$$\begin{aligned} S_*S'_* - SS' &= (I - \rho_*W)(I - \rho_*W)' - (I - \rho W)(I - \rho W)' \\ &= (\rho - \rho_*)(W + W') + (\rho_*^2 - \rho^2)WW'. \end{aligned} \quad (\text{A.14})$$

Then from (4.1), (A.14) is bounded by

$$C(\sigma^2 - \sigma_*^2)^2 + C(\rho - \rho_*)^2. \quad (\text{A.15})$$

This can be made arbitrarily small uniformly on \mathcal{N} by choosing \mathcal{N} small enough. Since any open cover of $\Theta_{\sigma^2} \times \Theta_{\rho}$ has a finite subcover, the proof of (A.9) is completed. The second term in $u(\theta)$ can be dealt with in a similar way, using the fourth moment conditions in Assumption A1. We omit the details.

Now looking at $v(\theta)$, in view of Assumption A3 it suffices to show that, for any $\delta_\mu > 0$, that

$$\liminf_{n \rightarrow \infty} \inf_{\substack{|\mu - \mu_0| > \delta_\mu \\ \sigma^2 \in \Theta_{\sigma^2}, \rho \in \Theta_{\rho}}} (\mu - \mu_0)^2 \frac{\ell' \Omega(\sigma^2, \rho)^{-1} \ell}{n} > 0. \quad (\text{A.16})$$

But

$$\begin{aligned} \frac{\ell' \Omega(\sigma^2, \rho)^{-1} \ell}{n} &\geq n (\ell' \Omega(\sigma^2, \rho) \ell)^{-1} \\ &= \left(\sigma^2 + \frac{\ell' TT' \ell}{n} \right)^{-1} \\ &\geq \left(\sigma^2 + \|T\|^2 \right)^{-1} \\ &\geq \left(\sigma^2 + (1 - |\rho|)^{-2} \right)^{-1} \\ &\geq \left(c_4 + (1 - \max(c_5, c_6))^{-2} \right)^{-1}, \end{aligned} \quad (\text{A.17})$$

so (A.16) is established, to complete the proof. \blacksquare

Proof of Theorem 2: By the mean value theorem,

$$0 = n^{\frac{1}{2}} \frac{\partial Q(\hat{\theta})}{\partial \theta} = n^{\frac{1}{2}} \frac{\partial Q(\theta_0)}{\partial \theta} + \tilde{A} n^{\frac{1}{2}} (\hat{\theta} - \theta_0), \quad (\text{A.18})$$

where \tilde{A} is formed by evaluating each row of $\partial^2 Q(\theta)/\partial \theta \partial \theta'$ at (possibly different) $\tilde{\theta}^{(i)}$, $i = 1, 2, 3$, such that $\|\tilde{\theta}^{(i)} - \theta_0\| \leq \|\hat{\theta} - \theta_0\|$. To evaluate the derivatives, for notational convenience write $\Omega(\sigma^2, \rho)$, and $Q(\theta)$ as, respectively, Ω , and Q . Note first that

$$\frac{\partial \Omega}{\partial \sigma^2} = I, \quad \frac{\partial^2 \Omega}{\partial (\sigma^2)^2} = 0, \quad \frac{\partial^2 \Omega}{\partial \rho \partial \sigma^2} = 0. \quad (\text{A.19})$$

Noting also that

$$\frac{\partial S}{\partial \rho} = -W, \quad \frac{\partial T}{\partial \rho} = TWT, \quad (\text{A.20})$$

we have

$$\frac{\partial \Omega}{\partial \rho} = 4T(WT + T'W')T', \quad (\text{A.21})$$

$$\frac{\partial^2 \Omega}{\partial \rho^2} = 8T(WTT'W' + WTWT + T'W'T'W')T'. \quad (\text{A.22})$$

Then

$$\frac{\partial Q}{\partial \mu} = -\frac{2}{n} \ell' \Omega^{-1} (y - \mu \ell), \quad (\text{A.23})$$

$$\frac{\partial Q}{\partial \sigma^2} = \frac{1}{n} \text{tr} \{ \Omega^{-1} \} - \frac{1}{n} (y - \mu \ell)' \Omega^{-2} (y - \mu \ell), \quad (\text{A.24})$$

$$\frac{\partial Q}{\partial \rho} = 8 \frac{1}{n} \text{tr} \{ \Omega^{-1} TWT'T' \} \quad (\text{A.25})$$

$$- \frac{8}{n} (y - \mu \ell)' \Omega^{-1} TWT'T' \Omega^{-1} (y - \mu \ell), \quad (\text{A.26})$$

$$\frac{\partial^2 Q}{\partial \mu^2} = \frac{2}{n} \ell' \Omega^{-1} \ell, \quad (\text{A.27})$$

$$\frac{\partial^2 Q}{\partial \mu \partial \sigma^2} = -\frac{2}{n} \ell' \Omega^{-2} (y - \mu \ell), \quad (\text{A.28})$$

$$\frac{\partial^2 Q}{\partial \mu \partial \rho} = -\frac{8}{n} \ell' \Omega^{-1} T(WT + T'W')T' \Omega^{-1} (y - \mu \ell), \quad (\text{A.29})$$

$$\frac{\partial^2 Q}{\partial (\sigma^2)^2} = -\frac{1}{n} \text{tr} \{ \Omega^{-2} \} + \frac{2}{n} (y - \mu \ell)' \Omega^{-3} (y - \mu \ell), \quad (\text{A.30})$$

$$\begin{aligned} \frac{\partial^2 Q}{\partial \sigma^2 \partial \rho^2} &= -\frac{8}{n} \text{tr} \{ \Omega^{-2} TWT'T' \} \\ &\quad + \frac{16}{n} (y - \mu \ell)' \{ \Omega^{-1} TWT'T' \Omega^{-2} \} (y - \mu \ell), \end{aligned} \quad (\text{A.31})$$

$$\begin{aligned} \frac{\partial^2 Q}{\partial \rho^2} &= \frac{1}{n} \text{tr} \left[\Omega^{-1} \left(\frac{\partial^2 \Omega}{\partial \rho^2} - \frac{\partial \Omega}{\partial \rho} \Omega^{-1} \frac{\partial \Omega}{\partial \rho} \right) \Omega^{-1} \{ (y - \mu \ell)(y - \mu \ell)' - \Omega \} \right] \\ &\quad + \frac{1}{n} \text{tr} \left[\Omega^{-1} \frac{\partial \Omega}{\partial \rho} \Omega^{-1} \frac{\partial \Omega}{\partial \rho} \Omega^{-1} (y - \mu \ell)(y - \mu \ell)' \right], \end{aligned} \quad (\text{A.32})$$

where we omit the complicated expression for the latter in terms of W and T . From (A.23)-(A.25)

$$\frac{\partial Q_0}{\partial \mu} = -\frac{2}{n} \ell' \Omega_0^{-1} \xi, \quad (\text{A.33})$$

$$\frac{\partial Q_0}{\partial \sigma^2} = -\frac{1}{n} \text{tr} \{ (\xi \xi' - \Omega_0) \Omega_0^{-2} \}, \quad (\text{A.34})$$

$$\frac{\partial Q_0}{\partial \rho} = -\frac{8}{n} \text{tr} \{ (\xi \xi' - \Omega_0) \Omega_0^{-1} T_0 W T_0 T_0' \Omega_0^{-1} \}. \quad (\text{A.35})$$

Then we deduce that

$$\lim_{n \rightarrow \infty} nE \left\{ \frac{\partial Q_0}{\partial \theta} \frac{\partial Q_0}{\partial \theta'} \right\} = B. \quad (\text{A.36})$$

Also evaluating second derivative at θ_0 we deduce via (A.33)-(A.35)

$$\frac{\partial^2 Q_0}{\partial \theta \partial \theta'} \rightarrow_p A, \text{ as } n \rightarrow \infty, \quad (\text{A.37})$$

by establishing convergence in probability to zero of zero mean quantities, via techniques as in the proof of Theorem 1.

Then using Theorem 1 it is straightforward to show also that

$$\tilde{A} - \frac{\partial^2 Q_0}{\partial \theta \partial \theta'} \rightarrow_p 0, \quad (\text{A.38})$$

and hence, from (A.37), $\tilde{A} \rightarrow_p A$. It remains to show that

$$n^{\frac{1}{2}} \frac{\partial Q_0}{\partial \theta} \rightarrow_d \mathcal{N}(0, B). \quad (\text{A.39})$$

This follows if all suitably normalized linear combinations of the left hand side of (A.39) are asymptotically standard normal. To achieve this, a linear combination is written as a sum of martingale differences, and a martingale central limit theorem is applied. Many of the details are standard and straightforward, and the aspect that most warrants attention pertains to (A.35), in view of its dependence on W , so we simply consider the asymptotic normality of $n^{\frac{1}{2}} \partial Q_0 / \partial \rho$. First denote $P = \Omega_0^{-1} T_0 W T_0 T_0' \Omega_0^{-1}$, so

$$n^{\frac{1}{2}} \frac{\partial Q_0}{\partial \rho} = -\frac{8}{n^{\frac{1}{2}}} \text{tr} \{ (\xi \xi' - \Omega_0) P \}. \quad (\text{A.40})$$

Then writing $\tau = (\nu'/\sigma_0, \varepsilon)'$ and $R = (\sigma_0 I_n, 2T_0)$, we have $\xi = R\tau$, and thence (A.40) becomes

$$\begin{aligned} & \frac{-8}{n^{\frac{1}{2}}} \text{tr} \{(\tau\tau' - I_{2n}) M\} \\ &= \frac{-8}{n^{\frac{1}{2}}} \left\{ \sum_{i=1}^{2n} (\tau_i^2 - 1) m_{ii} + \sum_{i=1}^{2n} \tau_i \sum_{j<i} \tau_j (m_{ij} + m_{ji}) \right\}, \end{aligned} \quad (\text{A.41})$$

where τ_i is the i -th element of τ and m_{ij} is the (i, j) -th element of $M = R'PR$. In view of Assumption A1, a martingale central limit thus holds if, as $n \rightarrow \infty$

$$\max_i \frac{m_{ii}}{n^{\frac{1}{2}}} + \max_i E \left| \frac{\sum_{j<i} \tau_j (m_{ij} + m_{ji})}{n^{\frac{1}{2}}} \right|^{2+\delta} \rightarrow 0, \quad (\text{A.42})$$

for some $\delta > 0$. Using Burkholder and von Bahr/Esseen inequalities, the expectation is bounded by

$$C \left(\frac{\sum_{j<i} (m_{ij} + m_{ji})^2}{n} \right)^{1+\delta/2}, \quad (\text{A.43})$$

so it suffices to show that as $n \rightarrow \infty$

$$\max_i \frac{m_{ii}}{n^{\frac{1}{2}}} + \max_i \frac{\sum_{j<i} (m_{ij}^2 + m_{ji}^2)}{n} \rightarrow 0. \quad (\text{A.44})$$

Now

$$\sum_{j=1}^{2n} m_{ij}^2 = r_i' P R R' P' r_i, \quad (\text{A.45})$$

where r_i is the i -th column of R . Thus

$$\max_i \sum_{j=1}^{2n} m_{ij}^2 \leq \left(\max_i \|r_i\|^2 \right) \|PR\|^2. \quad (\text{A.46})$$

But

$$\max_i \|r_i\| \leq \|R\| \leq \sigma_0 + 2\|T\| \leq C, \quad (\text{A.47})$$

$$\|P\| \leq C\|W\| \leq C, \quad (\text{A.48})$$

whence (A.42) follows. \blacksquare

Proof of Theorem 3: It is straightforward to deduce under H_0 that

$$a_{11} = \frac{2}{\sigma_\xi^2}, \quad (\text{A.49})$$

$$a_{22} = \frac{1}{\sigma_\xi^2}, \quad (\text{A.50})$$

$$a_{23} = \lim_{n \rightarrow \infty} (8/\sigma_\xi^2) \operatorname{tr}(W) = 0, \quad (\text{A.51})$$

$$a_{33} = \frac{32}{\sigma_\xi^4} \lim_{n \rightarrow \infty} \frac{1}{n} \operatorname{tr} \{W(W + W')\}, \quad (\text{A.52})$$

$$b_{11} = \frac{4}{\sigma_\xi^2}, \quad (\text{A.53})$$

$$b_{12} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{E(\xi_i^3)}{\sigma_\xi^2} = \frac{E(\xi_i^3)}{\sigma_\xi^2}, \quad (\text{A.54})$$

$$b_{13} = \lim_{n \rightarrow \infty} \frac{1}{n} E \left\{ \left(\sum_{i=1}^n \frac{\xi_i}{\sigma_\xi^2} \right) \sum_{j=1}^n \frac{(\xi_j^2 - \sigma_\xi^2)}{\sigma_\xi^2} \right\} = 0, \quad (\text{A.55})$$

$$b_{22} = \lim_{n \rightarrow \infty} \frac{1}{n} E \left\{ \frac{1}{\sigma_\xi^2} \sum_{i=1}^n (\xi_i^2 - \sigma_\xi^2) \right\}^2 = \frac{E(\xi_i^4) - \sigma_\xi^4}{\sigma_\xi^4}, \quad (\text{A.56})$$

$$b_{23} = -8 \lim_{n \rightarrow \infty} \frac{1}{n} E \left\{ \frac{1}{\sigma_\xi^2} \sum_{i=1}^n (\xi_i^2 - \sigma_\xi^2) \frac{1}{\sigma_\xi^2} \sum_{j=1}^n \sum_{k=1}^n \xi_j \xi_k w_{jk} \right\} = 0, \quad (\text{A.57})$$

$$\begin{aligned} b_{33} &= 64 \lim_{n \rightarrow \infty} \frac{1}{n} E \left\{ \frac{1}{\sigma_\xi^4} \sum_{j=1}^n \sum_{k=1}^n \xi_j \xi_k w_{jk} \right\}^2 \\ &= \frac{64}{\sigma_\xi^4} \lim_{n \rightarrow \infty} \frac{1}{n} \operatorname{tr} \{W(W + W')\}, \end{aligned} \quad (\text{A.58})$$

where w_j is the (j, k) -th element of W . Thus A and B are diagonal matrices. Then the theorem is proved when $n^{-1} \operatorname{tr} \{W(W + W')\}$ converges to a positive limit. But if in fact $\operatorname{tr} \{W(W + W')\} = o(n)$, while Assumption A8 holds, then a modified proof leads to the statement of the present theorem. For brevity we omit the details. ■

Proof of Theorem 4: The result for s_1 follows directly from Theorems 1 and 3. The proof for s_2 proceeds by noting that $\tilde{\xi}_i = \xi_i + (\mu_0 - \tilde{\mu})$, and then using standard arguments with a simplified version of the martingale central limit arguments in the proof of Theorem 2.