# LSE Research Online

# Miguel A. Delgado and Peter M. Robinson
## Non-nested testing of spatial correlation

**Article (Published version)
(Refereed)**

This version available at: http://eprints.lse.ac.uk/61433/
Available in LSE Research Online: Online: June 2015

http://eprints.lse.ac.uk

# Non-nested testing of spatial correlation

Miguel A. Delgado [a], Peter M. Robinson [b,*]

[a] *Universidad Carlos III de Madrid, Spain*

[b] *London School of Economics, United Kingdom*

**ABSTRACT**

We develop non-nested tests in a general spatial, spatio-temporal or panel data context. The spatial aspect can be interpreted quite generally, in either a geographical sense, or employing notions of economic distance, or when parametric modelling arises in part from a common factor or other structure. In the former case, observations may be regularly-spaced across one or more dimensions, as is typical with much spatio-temporal data, or irregularly-spaced across all dimensions; both isotropic models and non-isotropic models can be considered, and a wide variety of correlation structures. In the second case, models involving spatial weight matrices are covered, such as "spatial autoregressive models". The setting is sufficiently general to potentially cover other parametric structures such as certain factor models, and vector-valued observations, and here our preliminary asymptotic theory for parameter estimates is of some independent value. The test statistic is based on a Gaussian pseudo-likelihood ratio, and is shown to have an asymptotic standard normal distribution under the null hypothesis that one of the two models is correct; this limit theory rests strongly on a central limit theorem for the Gaussian pseudo-maximum likelihood parameter estimates. A small Monte Carlo study of finite-sample performance is included.

© 2015 The Authors. Published by Elsevier B.V.
This is an open access article under the CC BY license
(http://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

Spatial and spatio-temporal data are liable to exhibit correlation, which will likely depend on locations of observations or distances between them. Knowledge of locations or distances can improve precision and is desirably employed in modelling and statistical inference. Regular spacing across a temporal dimension is likely, but intervals between observations across geographical space can be regular or irregular, while geographic distances between observations can be unavailable or less relevant than "economic distances", say. Models for regularly-spaced "lattice" data in two or more dimensions (see e.g. Whittle, 1954) can relatively straightforwardly extend time series ones, but statistical inference for irregularly spaced data is not well developed. For example, for irregularly spaced observations on a continuous Gaussian process, despite such work as Dunsmuir (1983), Matsuda and Yajima (2009) and Robinson (1977), there appear to exist no satisfactory set of regularity conditions for the central limit theorem for parametric maximum likelihood estimates which separate out the process generating the observations from that generating the locations, and this is the case even in the single dimension irregularly-spaced time series setting, which has attracted attention over the years. Partly as a result, models of "spatial autoregressive" type, first developed by Cliff and Ord (1972), have proved popular, especially in economics; these model correlations in terms of spatial weight matrices, often linearly in observations and unknown parameters, and possibly also in the weights, and are relatively convenient computationally. The elements of the weight matrices are pairwise inverse measures of distance, either economic distances or geographic ones, where the latter might not be Euclidean, allowing for example for natural barriers such as rivers. The philosophy of such models is quite different from that of spatial statistics models for observations whose argument is location.

The diversity of possible dependence models highlights the lack of a "generic" spatial data set, and motivates development of statistical inference that potentially covers a variety of the settings mentioned above, rather than being limited to a single model class. In the present paper we focus on justifying tests of non-nested hypotheses for spatial or spatio-temporal correlation. The rival models could be two members of the same general class, for example two different models of autoregressive moving average type in case of regularly-spaced "lattice" data, or a Matern and Markov model when irregularly-spaced locations are known, or two weight matrix type models such as a "spatial autoregressive" versus "spatial moving average" model, or they could be from

---

\* Corresponding author. Tel.: +44 20 7955 7516; fax: +44 20 7955 6592.
*E-mail address:* p.m.robinson@lse.ac.uk (P.M. Robinson).

different classes, given that the weight matrix models can in principle be employed in all these data settings. Nonparametric methods for estimating spatial correlation have been developed but in general are more problematic than in the time series setting where stationarity and regular spacing allow consistent estimation of autocovariances or spectral densities despite lack of replication. We thus focus on parametric models. Moreover the testing scenario is between models of covariances between observations, or much more likely, between unobservable disturbances, rather than between full statistical models.

In particular, for random variables $u_j, j = 1, 2, \ldots$, we consider the rival models

$$H_i : Cov\left(u_j, u_k\right) = \sigma_{i0}^2 \omega_{ijk}\left(\theta_{i0}\right), \quad j, k = 1, 2, \ldots; \; i = 1, 2, \quad (1)$$

where, for $i = 1, 2$, $\theta_{i0}$ is an unknown $p_i \times 1$ vector, $\sigma_{i0}^2$ is an unknown positive scalar, variation-free of $\theta_{i0}$, and $\omega_{ijk}(.)$ is a known function of its $p_i$-dimensional argument. Because inference will be based on implicitly-defined extremum estimates of parameters, the zero subscript is as usual used to denote true value. Though observable $u_j$ are covered, we motivate our focus on (1) in the context of a parametric model for the sequence of observations $y_j$:

$$f_j(y_j; \beta_0) = u_j, \quad j = 1, 2, \ldots, \quad (2)$$

where the $f_j$ are known functions of their arguments and possibly of observable explanatory variables varying with $j$, $\beta_0$ is an unknown $q \times 1$ parameter vector assumed variation-free of the $\theta_{i0}$, and $u_j$ is, thus, unobservable, but assumed to be a random variable with mean zero. For example, $f_j(y_j; \beta_0)$ may represent the deviation of $y_j$ from a linear or nonlinear regression function,

$$f_j(y_j; \beta_0) = y_j - g(z_j; \beta_0), \quad (3)$$

where $g$ is a known linear, partly linear or wholly nonlinear function of its arguments and $z_j$ is a vector of observable stochastic (but exogenous) or nonstochastic explanatory variables, including time trends in a spatio-temporal setting, or dummy variables. More generally, $f_j$ might be nonlinear in $y_j$, for example a parametric Box–Cox or arcsinh transformation. Correlation and heteroscedasticity in $y_j$ are thus supposed not to be fully accounted for by $z_j$.

Given $n$ observations on $y_j$ in (2), and writing $u = (u_1, \ldots, u_n)'$, there is interest in estimating the covariance matrix $E(uu')$, which has $(j, k)$th element $Cov\left(u_j, u_k\right)$, for the sake of robust and/or efficient inference on $\beta_0$. For example, given observations $y_1, \ldots, y_n$, the linear or nonlinear least squares estimate of $\beta_0$ in (3) is $\sqrt{n}$-consistent as $n \to \infty$ with a centred limiting normal distribution under regularity conditions on $g$ and the $z_j$, as well as conditions which suitably limit the extent of the correlation in the $u_j$, but the variance matrix in the limit distribution depends on the covariance structure of the $u_j$, and information on this is needed to consistently estimate this variance matrix and thereby provide robust inference on $\beta_0$, that is, asymptotically valid hypothesis tests and consistent interval estimates. Further, in the presence of dependence in the $u_j$ the least squares estimate of $\beta_0$ is generally asymptotically inefficient; efficient estimation via generalized linear or nonlinear least squares, and thence locally most powerful testing, will again require information on the covariance structure of $u_j$. The correlation in the $u_j$ is described in terms of the $n \times 1$ vector $u$ even though $n$ is regarded as increasing in asymptotic theory because, as mentioned previously, some spatial models are expressed in terms of one or more specified $n \times n$ spatial weight matrices: a generic such matrix $W$ has zero diagonal elements and typically satisfies some normalization restriction, e.g. that each of its rows sums to unity (though it need not necessarily be symmetric and it may have some negative elements). Consequently the $\omega_{ijk}(.)$, and thence the elements of $u$ and thus $y$, can be $n$-dependent, but we suppress this feature in

the notation. Of course since the $u_j$ are unobservable we would estimate the $\theta_{i0}$ in (1) after replacing each $u_j$ by its proxy $\widehat{u}_j = f_j(y_j; \widehat{\beta})$, where $\widehat{\beta}$ is a $\sqrt{n}$-consistent estimate of $\beta_0$, such as described above, and we suppose that, for $i = 1, 2$, $\theta_{i0}$, $\sigma_{i0}^2$ are variation-free of $\beta_0$ in (2). Given a $\sqrt{n}$-consistent estimate $\widehat{\beta}$ of $\beta_0$ in (2) we can proxy the $u_j$ by the $\widehat{u}_j = f_j(y_j; \widehat{\beta})$ in estimating the $\omega_{ijk}(\theta_{i0})$, in the usual way.

We test between the hypotheses in (1) by tests of Cox (1961, 1962) type. Non-nested tests between structures of "spatial autoregressive" form have been developed by several authors, see e.g. Anselin (1986), Burridge (2012), Burridge and Fingleton (2010), Han and Lee (2013), Jin and Lee (2013), Kelejian (2008), Kelejian and Piras (2011), Piras and Lozano-Gracia (2012), but mainly J-tests, though Jin and Lee (2013) also develop Cox tests, and mostly focusing on the issue of testing between different spatial weight matrix specifications. As indicated previously, our framework is designed to cover not only "spatial autoregressive" models, but also others, which do not involve weight matrices, as well as models for panel and spatio-temporal data which may or may not employ weight matrices; parametric modelling of heteroscedasticity can also be embraced. Cox tests may be more suitable than J-tests when only covariance structure is at issue. Kelejian and Piras (2011) note that J-tests are based on whether or not predictions based on alternative models add significantly to the explanatory power of the null model, and show that they cannot be used to construct a test which concerns only the structure of the error term in SAR models. A comparison between Cox tests and J-tests for SAR models, which includes a Monte Carlo study, can be found in Jin and Lee (2013). Formally, our methodology can also cover tests of nested hypotheses. An ancillary contribution of the paper is the justification of Gaussian pseudo-likelihood parameter estimates in a quite general setting. Our conditions do not assume stationarity of $u_j$ but are motivated by approximate stability. Inevitably, in view of the diversity of settings covered and the intrinsic issues with some of them, our conditions are high level, and some can be hard or impossible to satisfactorily check, but we provide some discussion. It would be possible to extend our work also to test between non-nested models for $y_j$ of type (2), for example between two regression models alongside non-nested models for $E(uu')$.

The following section describes a number of models that might feature as non-nested hypotheses. Our non-nested test is presented in Section 3, including versions that are robust with respect to departures from normality, and Section 4 contains a small Monte Carlo study of finite-sample performance, with Section 5 offering some concluding comments. Theoretical, large-sample, justification of the test is left to Appendices. Appendix A lists and discusses regularity conditions. Appendix B presents and proves several theorems: our test statistic is a function of Gaussian pseudo-maximum likelihood estimates of the parameter vectors $\theta_{10}$ and $\theta_{20}$ in (1), and the null (taken to be the hypothesis $H_1$) asymptotic distribution of the test statistic depends heavily on the null asymptotic distribution of the parameter estimates, so for these we provide consistency and asymptotic normality results which have some novelty in our general setting and represent by far the main technical contribution of the paper, given the breadth of models potentially covered.

## 2. Spatial correlation models

We consider first observations recorded on $d$-dimensional Euclidean space $\mathbb{R}^d$. For this purpose we introduce the location $t \in \mathbb{R}^d$. We proceed as if we have observations $u_j, j = 1, \ldots, n$, though as discussed above the $u_j$ are likely unobservable and replaced in estimation by observable proxies. An important theme

in the early spatial statistics literature concerned the foundations of modelling dependence in terms of distributional behaviour, with a debate between commencing from the joint distribution of the data $u_j, j = 1, \ldots, n$ (see e.g. Whittle, 1963), and from conditional distributions (see e.g. Bartlett, 1955). In particular, the specification of conditional distributions might not lead to a well defined joint distribution. The Hammersley–Clifford Theorem (see Hammersley and Clifford, unpublished manuscript, Besag, 1974) derived the most general form possible for a joint distribution, under regularity conditions. Both the joint distribution and conditional distribution approaches lead to a variety of statistical models, as does that of modelling the $u_j, j = 1, \ldots, n$, as a (possibly linear) filter of $n$ (or possibly infinitely many) independent unobservable random variables (see e.g. Tjostheim, 1978, 1983, Whittle, 1954). Because the rules of statistical inference we develop are based on the Gaussian pseudo-likelihood, the (parametric) specification of the covariance matrix is basic to our approach, with actual distributional form not of concern. Thus our discussion will stress covariance structure, though we will also give some space to connections with the approaches to modelling referred to above.

Given observations at $n$ distinct locations $t_1, \ldots, t_n$ on a scalar zero-mean process $U(t)$, we make the identification $U(t_j) = u_j$, $j = 1, \ldots, n$, where unlike with time series there is no natural ordering. However, in part motivated by the time series setting it is natural to consider the simplifying case that $U(t)$ is covariance stationary, so $EU(t) U(t + s) = \sigma_0^2 \gamma(s)$ for some function $\gamma(s)$ and unknown positive scalar $\sigma_0^2$, and all $t, s \in \mathbb{R}^d$. Consider a parameterization $\gamma(s; \phi), \phi \in \mathbb{R}^m$, such that $\gamma(s; \phi_0) = \gamma(s)$ for some $\phi_0 \in \mathbb{R}^m$. Here $\phi_0$ generically represents either $\theta_{01}$ or $\theta_{02}$ of the previous section. We thus take $\omega_{jk}(\phi) = \gamma(t_j - t_k; \phi)$, which generically represents $\omega_{1jk}(\theta_1)$ or $\omega_{2jk}(\theta_2)$ above, $\theta_i \in \mathbb{R}^{p_i}, i = 1, 2$.

When $t$ has integer-valued components, i.e. $t \in \mathbb{Z}^d$, there is an extension of the regularly-spaced time series setting, and thus extensions of typical time series models can be considered, for example, autoregressive moving averages, following Whittle (1954). To define these, introduce $L = (L_1, \ldots, L_d)$ such that $\Pi_{h=1}^d L_h^{l_h} U(t) = U(t - l), l = (l_1, \ldots, l_d) \in \mathbb{Z}^d$, and $a(L; \phi) = \sum_{l_1=-q_{L1}}^{q_{U1}} \cdots \sum_{l_d=-q_{Ld}}^{q_{Ud}} a_l(\phi) \Pi_{h=1}^d L_h^{l_h}, b(L; \phi) = \sum_{l_1=-r_{L1}}^{r_{U1}} \cdots \sum_{l_d=-r_{Ld}}^{r_{Ud}} b_l(\phi) \Pi_{h=1}^d L_h^{l_h}$ for given non-negative integers $q_{Lh}$, $q_{Uh}$, $r_{Lh}$, $r_{Uh}$, $h = 1, \ldots, d$, and given functions $a_l(\phi), b_l(\phi)$. Letting $\varepsilon(t), t \in \mathbb{Z}^d$, be independent and identically distributed (iid) random variables with zero mean and variance $\sigma_0^2$, under suitable conditions on $a(L; \phi)$ and $b(L; \phi)$, the process $U(t)$ generated by

$$a(L; \phi_0) U(t) = b(L; \phi_0) \varepsilon(t), \quad t \in \mathbb{Z}^d, \tag{4}$$

not only generalizes the time series stationary and invertible autoregressive moving average process to a general dimension $d$, but also allows for leads as well as lags, recognizing the lack of chronological ordering of spatial data. The $\gamma(s; \theta)$ and thus $\omega_{jk}(\theta)$ can be determined from (4). The model (4) potentially suffers seriously from the curse of dimensionality. This might be alleviated by, for example, replacing $a(L; \phi), b(L; \phi)$ by the product forms $\Pi_{h=1}^d \sum_{l_h=-q_{Lh}}^{q_{Uh}} a_{l_h}(\phi) L_h^{l_h}, \Pi_{h=1}^d \sum_{l_h=-r_{Lh}}^{r_{Uh}} b_{l_h}(\phi) L_h^{l_h}$, respectively. A parsimonious case of (4), $d = 2$ with $m = 1$ treated in the geography literature (see e.g. Hepple, 1976), is the first-order quadrilateral autoregression

$$\left(1 - \phi\left(L_1^{-1} + L_2^{-1} + L_1 + L_2\right)\right) U(t) = \varepsilon(t). \tag{5}$$

On the other hand Haining (1978) considered the corresponding moving average model

$$U(t) = \left(1 + \phi\left(L_1^{-1} + L_2^{-1} + L_1 + L_2\right)\right) \varepsilon(t). \tag{6}$$

Isotropy is another assumption that can produce parsimonious models. To define this we return to the previous more general

setting of $t \in \mathbb{R}^d$. We say $U(t)$ is isometric if for some function on $\mathbb{R}, \gamma(s) = \delta(|s|)$, where $|s|$ is the Euclidean distance of $s$ from the origin. Thus we consider parametric functions $\delta(|s|; \phi)$. One important class is the model of Matern (1986), which has various parameterizations (see Stein, 1999, pp. 48–51), one of which is

$$\delta(|s|; \phi) = \frac{1}{2^{\phi_1 - 1} \Gamma(\phi_1)} \left(\frac{(2\phi_1)^{1/2} |s|}{\phi_2}\right)^{\phi_1}$$
$$\times \mathcal{K}_{\phi_1}\left(\frac{(2\phi_1)^{1/2} |s|}{\phi_2}\right), \tag{7}$$

for $m = 2, \phi = (\phi_1, \phi_2)'$ with $\phi_j > 0, j = 1, 2$, and where $\mathcal{K}_{\phi_1}$ is the modified Bessel function of the second kind (see e.g. Gradshteyn and Ryzhik, 1994). Another parsimonious isotropic model with $m = 2$ has

$$\delta(|s|; \phi) = \exp\left(-|s/\phi_2|^{\phi_1}\right), \tag{8}$$

where $\phi_1 \in (0, 2], \phi_2 > 0$, (see e.g. Diggle et al., 1988, De Oliveira et al., 1997, Stein, 1999). When $\phi_1 = 0.5$, (7) reduces to the exponential covariance function $\exp(-|s/\phi_2|)$, which is identical to (8) with $\phi_1 = 1$, while as $\phi_1 \to \infty$, (7) converges to $\exp\left(-(s/\phi_2)^2/2\right)$, but non-nested tests can choose between (7) and (8). A number of other models, and their fitting to irregularly-spaced data, have been considered by, e.g., Vecchia (1988), Jones and Vecchia (1993), Handcock and Wallis (1994), Stein et al. (2004) and Fuentes (2007).

Other examples entail one or more of the spatial weight matrices described in the previous section. Similarly to (4), these are most commonly expressed as a linear transformation of unobservable iid zero-mean random variables. Denoting by $\varepsilon$ an $n \times 1$ vector of these, we write

$$S(\phi_0) u = \varepsilon. \tag{9}$$

where the $n \times n$ matrix $S(\phi)$ has full rank for all relevant $\phi$, and suppressing reference to weight matrices. Thus $\Omega(\phi)$, the $n \times n$ matrix with $(j, k)$th element $\omega_{jk}(\phi)$, is given by

$$\Omega(\phi) = S(\phi)^{-1} S(\phi)^{-1\prime}. \tag{10}$$

Models of this type can be natural in, for example, a network setting. Consider first the $m$th order spatial autoregression (SAR($m$)), for $m \geq 1$, where

$$S(\phi) = I_n - \sum_{j=1}^m \phi_j W_j, \tag{11}$$

where $I_n$ is the $n \times n$ identity matrix and the $W_j$ are $n \times n$ weight matrices. By far the most frequently treated case of (11) in the theoretical and empirical literature is the SAR(1) (see e.g. Cliff and Ord, 1972; Arbia, 2006). Here, $W_1$ is sometimes chosen to be row-normalized such that the elements of each row sum to 1. The SAR($m$) might be compared in non-nested testing with the spatial moving average SMA($m$), where

$$S(\phi) = \left(I_n + \sum_{j=1}^m \phi_j W_j\right)^{-1}. \tag{12}$$

Both (11) and (12) are nested in

$$S(\phi) = \left(I_n + \sum_{j=m_a+1}^{m_a+m_b} \phi_j W_j\right)^{-1} \left(I_n - \sum_{j=1}^{m_a} \phi_j W_j\right), \tag{13}$$

denoting the spatial autoregressive moving average (SARMA($m_a$, $m_b$)), for $m_a \geq 1, m_b \geq 1, m_a + m_b = m$. In non-nested testing, the SARMA($m_a, m_b$) might be compared with the SARMA($m_b, m_a$), where either $m_a > m_b$ or $m_a < m_b$, or with the SAR($m$) or

SMA($m$). An alternative type of model is the matrix exponential spatial model MESS($m$), where

$$S(\phi) = \exp\left(-\sum_{j=1}^{m} \phi_j W_j\right) \tag{14}$$

and exp(.) is the matrix exponential function,
$\exp(A) = \sum_{j=0}^{\infty} A^j / j!$;
this model was proposed for $m = 1$ by LeSage and Pace (2009). The MESS($m$) might naturally be compared in non-nested testing with the SAR($m$) as in Han and Lee (2013) or with the SMA($m$). Other $S(\phi)$ that are non-linear functions of weight matrices might also be considered.

Advantages of the class (9) include the guaranteed non-negative definiteness of $\Omega(\phi)$ (10), the "lag" interpretation in (11)–(13), somewhat analogous to time series models, and the possibility of choosing weight matrices to be non-symmetric and to have some negative elements (though often they are symmetric with non-negative elements). However, given that the $(j, k)$th element $w_{jk}$ of a weight matrix can represent the inverse "distance" between agents $j$ and $k$, it is noticeable that for all of the cases of (9) presented in the previous paragraph $\omega_{jk}(\phi)$ does not depend only on $w_{jk}$. For example, for the SMA(1), $\omega_{jk}(\phi)$ depends on $w_{jl}$, $w_{lk}$, all $l = 1, \ldots, n$, while for the SAR(1) and MESS(1) it depends on the whole weight matrix. Such outcomes can be rationalized, but there is also a case for using a weight matrix in a simpler and more direct way in modelling $\Omega(\phi)$, which is arguably the most basic quantity of interest, indeed under Gaussianity it uniquely describes the distribution of $u$, apart from a scale factor. If we consider a weight matrix $V$ with rather different properties from before, being positive definite (and thus having positive elements on the diagonal), we might consider

$$\omega_{jk}(\phi) = \omega_{jk}(v_{jk}; \phi), \tag{15}$$

the notation stressing the dependence of $\omega_{jk}(\phi)$ on only the $(j, k)$th element $v_{jk}$ of $V$. As very simple examples, with $m = 1$ and $v_{jk} \geq 0$,

$$\omega_{jk}(\phi) = v_{jk}^{\phi}, \quad \phi > 0, \tag{16}$$

or

$$\omega_{jk}(\phi) = \phi^{1/v_{jk}}, \quad \phi \in (0, 1). \tag{17}$$

In both cases, $\omega_{jk}(\phi) \to 0$ as $v_{jk} \to 0$. However, we would require that $\Omega(\phi)$ is positive definite for all relevant values of $\phi$, and we are unable to give sufficient conditions for this property for either the models (16) or (17). Nevertheless, commencing modelling from the form (15) receives some further support from Section 4.2.2 of Besag (1974), who showed how a joint normal distribution for $u_s, s = 1, \ldots, n$, with covariance matrix $\sigma_0^2 \Omega(\phi_0)$ can arise from a certain specification of each of the $n$ (normal) conditional distributions of the $u_s$ (rather than from (10) and (9) with normal $\varepsilon$), with likelihood-based inference depending on positive definiteness of $\Omega(\phi)$ for all relevant values of $\phi$.

The setup in (1) is sufficiently general to cover also multivariate data (e.g. where $n = NK$, and we have $N$ observations on a $K$-dimensional vector, for fixed $K$ and $N \to \infty$) and panel data (where $n = NT$, and either or both the cross-sectional dimension $N$ and the time dimension $T$ are regarded as diverging in asymptotic theory); in each case a variety of dependence structures is possible.

## 3. Non-nested tests

Cox (1961, 1962) developed log-likelihood ratio type tests between non-nested probability densities for iid observations; White (1982) provided asymptotic justification in that setting. Our concern is to test between rival spatial correlation structures,

with precise distributional structure not of interest. Our tests are based on a Gaussian pseudo-log-likelihood ratio and thus share the robustness to non-Gaussianity property of the parameter estimates studied in the previous section. For a known, non-normal, parametric density for the $\varepsilon_j$ more efficient tests would be based on the appropriate maximum likelihood estimates. Indeed, the same efficiency could be achieved using adaptive estimates when the $\varepsilon_j$ have density of unknown, nonparametric form (as studied in a spatial autoregressive context by Robinson (2010)).

Our non-nested tests are based on parameter estimates of both models in (1). For $i = 1, 2$, denote by $\theta_i, \sigma_i^2$ respectively a $p_i \times 1$ vector and a scalar, representing any admissible values of $\theta_{i0}, \sigma_{i0}^2$ respectively, write $p = p_1 + p_2$, let $\Omega_i(\theta_i)$ be the $n \times n$ matrix with $(j, k)$th element $\omega_{ijk}(\theta_i)$, and define

$$L_i\left(\theta_i, \sigma_i^2\right) = \frac{1}{2}\log\sigma_i^2 + \frac{1}{2n}\log|\Omega_i(\theta_i)| + \frac{1}{2n\sigma_i^2}u'\Omega_i^{-1}(\theta_i)u, \tag{18}$$

which is minus the normalized Gaussian pseudo-maximum-likelihood based on (1), up to a constant. We do not assume normality, but base our parameter estimates and non-nested tests on (18). Our estimates of $\theta_{i0}, \sigma_{i0}^2$ minimize $L_i(\theta_i, \sigma_i^2)$. For given $\theta_i$, $L_i(\theta_i, \sigma_i^2)$ has a minimum

$$Q_i(\theta_i) = L_i\left(\theta_i, \overline{\sigma}_i^2(\theta_i)\right)$$

$$= \frac{1}{2}\log\overline{\sigma}_i^2(\theta_i) + \frac{1}{2n}\log|\Omega_i(\theta_i)| + \frac{1}{2}, \tag{19}$$

where

$$\overline{\sigma}_i^2(\theta_i) = \frac{1}{n}u'\Omega_i^{-1}(\theta_i)u = \frac{1}{n}u'\Omega_i^{-1}u, \tag{20}$$

writing $\Omega_i = \Omega_i(\theta_i)$. For $i = 1, 2$, define $R_i$ to be a given compact subset of $\mathbb{R}^{p_i}$ and

$$\widehat{\theta}_i = \underset{\theta_i \in R_i}{\arg\min}\, Q_i(\theta_i), \tag{21}$$

$$\hat{\sigma}_i^2 = \overline{\sigma}_i^2\left(\widehat{\theta}_i\right), \tag{22}$$

the Gaussian pseudo-maximum-likelihood estimates of $\theta_{i0}, \sigma_{i0}^2$.

From (18), (19), the Gaussian pseudo log-likelihood-ratio statistic for comparing the models in (1) is

$$2\left(Q_2\left(\widehat{\theta}_2\right) - Q_1\left(\widehat{\theta}_1\right)\right) = \log\frac{\overline{\sigma}_2^2\left(\widehat{\theta}_2\right)}{\overline{\sigma}_1^2\left(\widehat{\theta}_1\right)} + \frac{1}{n}\log\left|\Omega_2\left(\widehat{\theta}_2\right)\right|$$

$$- \frac{1}{n}\log\left|\Omega_1\left(\widehat{\theta}_1\right)\right|.$$

This converges in probability to a non-zero limit under $H_1$, where throughout our asymptotic theory conditions on locations, which might be regarded as stochastically generated in some situations. Defining, for $i = 1, 2$,

$$\tilde{\sigma}_i^2 = \tilde{\sigma}_i^2(\theta_i) = E_1\overline{\sigma}_i^2(\theta_i) = \sigma_{10}^2 n^{-1} tr\left(\Omega_i^{-1}\Omega_{10}\right) \tag{23}$$

and

$$\tilde{Q}_i = \tilde{Q}_i(\theta_i) = \frac{1}{2}\log\left\{\tilde{\sigma}_i^2(\theta_i)\right\} + \frac{1}{2n}\log|\Omega_i(\theta_i)|$$

$$= \frac{1}{2}\log\left\{\tilde{\sigma}_i^2\right\} + \frac{1}{2n}\log|\Omega_i|, \tag{24}$$

a centred statistic is

$$2\left(Q_2\left(\widehat{\theta}_2\right) - Q_1\left(\widehat{\theta}_1\right)\right) - 2\left(\tilde{Q}_2\left(\widehat{\theta}_2\right) - \tilde{Q}_1\left(\widehat{\theta}_1\right)\right)$$

$$= \log\frac{\overline{\sigma}_2^2\left(\widehat{\theta}_2\right)}{\overline{\sigma}_1^2\left(\widehat{\theta}_1\right)} - \log\frac{\tilde{\sigma}_2^2\left(\widehat{\theta}_2\right)}{\tilde{\sigma}_2^2\left(\widehat{\theta}_1\right)}.$$

This can be written, using (23), as

$$\log \frac{\overline{\sigma}_2^2 \left(\widehat{\theta}_2\right)}{\overline{\sigma}_1^2 \left(\widehat{\theta}_1\right)} - \log \frac{tr \left(\Omega_2^{-1} \left(\widehat{\theta}_2\right) \Omega_{10}\right)}{tr \left(\Omega_1^{-1} \left(\widehat{\theta}_1\right) \Omega_{10}\right)},$$

which can be estimated by

$$\log \frac{\overline{\sigma}_2^2 \left(\widehat{\theta}_2\right)}{\overline{\sigma}_1^2 \left(\widehat{\theta}_1\right)} - \log \frac{1}{n} tr \left(\Omega_2^{-1} \left(\widehat{\theta}_2\right) \Omega_1 \left(\widehat{\theta}_1\right)\right),$$

which we write as

$$LR = \log \frac{\overline{\sigma}_2^2 \left(\widehat{\theta}_2\right)}{\widetilde{\sigma}_2^2 (\widehat{\theta}_1, \widehat{\theta}_2)},$$

where

$$\overline{\sigma}_2^2(\theta_1, \theta_2) = \overline{\sigma}_1^2 (\theta_1) u (\theta_1, \theta_2),$$

with

$$u (\theta_1, \theta_2) = \frac{1}{n} tr \left(\Omega_2^{-1} (\theta_2) \Omega_1 (\theta_1)\right).$$

Under $H_1$ and conditions in Appendix A, $\overline{\sigma}_2^2 \left(\widehat{\theta}_2\right) - \overline{\sigma}_2^2(\widehat{\theta}_1, \widehat{\theta}_2) = o_p (1)$, but under $H_2$ $LR$ will generally have a non-zero probability limit, indicating that $LR$ is a basis for testing $H_1$. In Theorem 4 of Appendix B we justify the following large sample approximate null distributions:

$$LR \simeq \mathcal{N} \left(0, n^{-1} \widehat{e}' \widehat{M}^{-1} \widehat{N}_i \widehat{M}^{-1} \widehat{e}\right), \quad i = 1, 2, 3, 4, \qquad (25)$$

where $\widehat{e}$, $\widehat{M}$ and the $\widehat{N}_i$ are defined below and normality is assumed for $i = 1$ but robustness to departures from normality is afforded by the other choices of $i$, as will be explained. With level $\alpha \in (0, 1)$, and $z_\alpha$ such that the probability that a standard normal variate exceeds $z_\alpha$ is $\alpha$, it is proposed to reject $H_1$ in the direction of $H_2$ if $|LR| \geq \left(n^{-1} \widehat{e}' \widehat{M}^{-1} \widehat{N}_i \widehat{M}^{-1} \widehat{e}\right)^{1/2} z_{\alpha/2}$. In that event, as is common practice in non-nested testing, one can switch $H_1$ and $H_2$ and if there is a further rejection the test is deemed inconclusive. At the end of Appendix B we mention test statistics that are slightly simpler but valid less generally.

To define $\widehat{e}$ in (25), introduce $\tau_i = \left(\theta_i', \sigma_i^2\right)'$, $\widehat{\tau}_i = \left(\widehat{\theta}_i', \widehat{\sigma}_i^2\right)'$, $i = 1, 2$, $\tau = \left(\tau_1', \tau_2'\right)'$, $\widehat{\tau} = \left(\widehat{\tau}_1', \widehat{\tau}_2'\right)'$ and

$$c_j (\tau) = -\frac{\sigma_1^2}{n} tr \left(\Omega_2^{-1} \Omega_{1j}\right), \quad j = 1, \ldots, p_1;$$

$$d_j (\tau) = \frac{\sigma_1^2}{n} tr \left(\Omega_2^{-1} \Omega_{2j} \Omega_2^{-1} \Omega_1\right), \quad j = 1, \ldots, p_2,$$

and then

$$e (\tau) = \left(c_1 (\tau), \ldots, c_{p_1} (\tau), -u (\theta_1, \theta_2),\right.$$
$$\left. d_1 (\tau), \ldots, d_{p_2} (\tau), 1\right)' / \sigma_2^2, \qquad \widehat{e} = e (\widehat{\tau}).$$

To define $\widehat{M}$ in (25) denote, for $i = 1, 2$ and $j, k, l = 1, \ldots, p_i$,

$$\Omega_{ij} = \Omega_{ij} (\theta_i) = (\partial/\partial \theta_{ij}) \Omega_i,$$
$$\Omega_{ijk} = \Omega_{ijk} (\theta_i) = (\partial/\partial \theta_{ik}) \Omega_{ij}, \qquad (26)$$
$$\Omega_{ijkl} = \Omega_{ijkl} (\theta_i) = (\partial/\partial \theta_{il}) \Omega_{ijk},$$

where the existence of the derivatives is assured by Assumption 8 in Appendix A. For $i = 1, 2$, let $M_i = M_i (\tau)$ be the $(p_i + 1) \times (p_i + 1)$ symmetric matrix with $(j, k)$th element $M_{ijk}$ given by

$$M_{ijk} = \frac{\sigma_1^2}{2\sigma_i^2 n} tr \left(\Omega_i^{-1} \Omega_{ij} \Omega_i^{-1} \Omega_{ik} \Omega_i^{-1} \Omega_1\right)$$
$$+ \frac{1}{2n} tr \left(\Omega_i^{-1} \left(\Omega_{ij} \Omega_i^{-1} \Omega_{ik} - \Omega_{ijk}\right)\right.$$
$$\left. \times \left(\frac{\sigma_1^2}{\sigma_i^2} \Omega_i^{-1} \Omega_1 - I_n\right)\right), \quad j, k = 1, \ldots, p_i; \qquad (27)$$

$$M_{i,j,p_i+1} = \frac{\sigma_1^2}{2n\sigma_i^4} tr \left(\Omega_i^{-1} \Omega_{ij} \Omega_i^{-1} \Omega_1\right),$$

$$j = 1, \ldots, p_i; \qquad M_{i,p_2+1,p_2+1} = \frac{\sigma_1^2}{n\sigma_i^6} tr \left(\Omega_i^{-1} \Omega_1\right) - \frac{1}{2\sigma_i^4}.$$

Now take $\widehat{M}_i = M_i (\widehat{\tau})$, $i = 1, 2$, and

$$\widehat{M} = \begin{pmatrix} \widehat{M}_1 & 0 \\ 0 & \widehat{M}_2 \end{pmatrix}.$$

Note that $\widehat{M}$ results from deriving

$$\begin{pmatrix} p \lim \dfrac{\partial^2 L_1}{\partial \tau_1 \partial \tau_1'} & 0 \\ 0 & p \lim \dfrac{\partial^2 L_2}{\partial \tau_2 \partial \tau_2'} \end{pmatrix}$$

and then evaluating at $\widehat{\tau}$. The second derivative terms, involving $\Omega_{2jk}$, are present in some $M_{2jk}$ due to imposing $H_1$ on the $H_2$ model, but the second trace in (27) vanishes for $i = 1$.

To define $\widehat{N}_1$ in (25), let $N = N (\tau)$ be the $(p + 2) \times (p + 2)$ matrix

$$N (\tau) = \begin{pmatrix} N_{11} & N_{12} \\ N_{12}' & N_{22} \end{pmatrix},$$

where, for $h, i = 1, 2$, $N_{hi} = N_{hi} (\tau)$ is the $(p_h + 1) \times (p_i + 1)$ matrix with $(j, k)$th element $N_{hijk}$ given by

$$N_{hijk} = \frac{\sigma_1^4}{2n\sigma_h^2 \sigma_i^2} tr \left(\Omega_h^{-1} \Omega_{hj} \Omega_h^{-1} \Omega_1 \Omega_i^{-1} \Omega_{ik} \Omega_i^{-1} \Omega_1\right),$$

$$j = 1, \ldots, p_h, \ k = 1, \ldots, p_i;$$

$$N_{hij,p_i+1} = \frac{\sigma_1^4}{2n\sigma_h^2 \sigma_i^4} tr \left(\Omega_h^{-1} \Omega_{hj} \Omega_h^{-1} \Omega_1 \Omega_i^{-1} \Omega_1\right),$$

$$j = 1, \ldots, p_h;$$

$$N_{hi,p_h+1,k} = \frac{\sigma_1^4}{2n\sigma_h^4 \sigma_i^2} tr \left(\Omega_i^{-1} \Omega_{ik} \Omega_i^{-1} \Omega_1 \Omega_h^{-1} \Omega_1\right),$$

$$k = 1, \ldots, p_i;$$

$$N_{hi,p_h+1,p_i+1} = \frac{\sigma_1^4}{2n\sigma_h^4 \sigma_i^4} tr \left(\Omega_h^{-1} \Omega_1 \Omega_i^{-1} \Omega_1\right).$$

Note that $N_{11} = M_1$. Take $\widehat{N}_1 = N (\widehat{\tau})$, and note that $\widehat{N}_1$ results from deriving

$$E_1 \left\{ \begin{pmatrix} \dfrac{\partial L_1}{\partial \tau_1} \\ \dfrac{\partial L_2}{\partial \tau_2} \end{pmatrix} \begin{pmatrix} \dfrac{\partial L_1}{\partial \tau_1} \\ \dfrac{\partial L_2}{\partial \tau_2} \end{pmatrix}' \right\},$$

(with $E_1$ denoting expectation under $H_1$) with the assumption that the $u_i$ are Gaussian, and then evaluating at $\widehat{\tau}$.

Our $\widehat{N}_2$, $\widehat{N}_3$ and $\widehat{N}_4$ are robust to departures from Gaussianity, and are thus potentially less precise than $\widehat{N}_1$ when $u$ is actually Gaussian. We need to proxy the iid innovations $\varepsilon_s$, with zero mean and variance $\sigma_{10}^2$, that appear in the linear process representation for $u_j$,

$$u_j = \sum_{s=1}^{\infty} b_{js} \varepsilon_s, \quad j = 1, 2, \ldots, n, \ n = 1, 2, \ldots, \qquad (28)$$

described in Assumption 7 of Appendix A and discussed further immediately afterwards. The representation (28) is central to the proofs of Theorems 2–4 (on the limit distribution of $\widehat{\tau}$ and the null limit distribution of our test statistic) in Appendix B. Such a representation follows naturally if we commence from an $H_1$ model of form $S (\theta_{10}) u = \varepsilon$ (cf. (9)), with $\varepsilon_s$ the $s$th element of $\varepsilon$, or a model

of form (4). Given the functions $b_{js}(\theta_1)$, such that $b_{js} = b_{js}(\theta_{10})$, let the $n \times n$ matrix $B(\theta_1)$ have $(j,s)$th element $b_{js}(\theta_1)$, and define the $n \times 1$ vector $\widehat{\varepsilon} = B(\widehat{\theta}_1)^{-1} u$. Note that for a model of form $S(\theta_{10}) u$, we have $B(\theta_1) = S(\theta_1)^{-1}$, and we have an exact factorization $\Omega_1(\theta_1) = B(\theta_1) B(\theta_1)'$. In models involving more than $n$, including infinitely many, $\varepsilon_s$, such as (4), $B(\theta_1) B(\theta_1)'$ only approximately factorizes $\Omega_1(\theta_1)$, but the truncation in (28) entailed in using $\widehat{\varepsilon} = B(\widehat{\theta}_1)^{-1} u$ can be rigorously justified (given the summability properties of the $b_{js}$ in Assumption 7 of Appendix A). Denoting by $b_s(\theta_1)$ the $n \times 1$ vector with $k$th element $b_{ks}(\theta_1)$, for $i = 1, 2$ and $k = 1, \ldots, n$ let $a_{ist}(\tau)$ be the $(p_i + 1) \times 1$ vector with $j$th element $a_{ijst}(\tau) = -(2n\sigma_i^2)^{-1} b_s'(\theta_1) \Omega_i^{-1} \Omega_{ij} \Omega_i^{-1} b_t(\theta_1)$ for $j = 1, \ldots, p_i$, and $-(2n\sigma_i^4)^{-1} b_s'(\theta_1) \Omega_i^{-1} b_t(\theta_1)$ for $j = p_i + 1$, and put $a_{st}(\tau) = (a_{1st}(\tau)', a_{2st}(\tau)')'$; note that $a_{st}(\tau) = a_{ts}(\tau)$. Denote the $s$th element of $\widehat{\varepsilon}$ by $\widehat{\varepsilon}_s$; the $\widehat{\varepsilon}_s$ might also be used in bootstrap versions of our tests. Define

$$\widehat{N}_2 = n \sum_{s=1}^{n} a_{ss}(\widehat{\tau}) a_{ss}'(\widehat{\tau}) \left(\widehat{\varepsilon}_s^2 - \widehat{\sigma}_1^2\right)^2$$
$$+ 2n \sum_{s,t=1; s \neq t}^{n} a_{st}(\widehat{\tau}) a_{st}'(\widehat{\tau}) \widehat{\varepsilon}_s^2 \widehat{\varepsilon}_t^2, \qquad (29)$$

and, slightly more simply,

$$\widehat{N}_3 = n \sum_{s=1}^{n} a_{ss}(\widehat{\tau}) a_{ss}'(\widehat{\tau}) \left(\widehat{\varepsilon}_s^2 - \widehat{\sigma}_1^2\right)^2$$
$$+ 2\widehat{\sigma}_1^4 n \sum_{s,t=1; s \neq t}^{n} a_{st}(\widehat{\tau}) a_{st}'(\widehat{\tau}), \qquad (30)$$

$$\widehat{N}_4 = \sum_{s=1}^{n} \left(\widehat{\varepsilon}_s^2 - \widehat{\sigma}_1^2\right)^2 \sum_{s=1}^{n} a_{ss}(\widehat{\tau}) a_{ss}'(\widehat{\tau})$$
$$+ 2\widehat{\sigma}_1^4 n \sum_{s,t=1; s \neq t}^{n} a_{st}(\widehat{\tau}) a_{st}'(\widehat{\tau}). \qquad (31)$$

For Gaussian $\varepsilon_s$ we have $E(\varepsilon_s^2 - \sigma_{10}^2)^2 = 2\sigma_{10}^4$, and on replacing $\left(\widehat{\varepsilon}_s^2 - \widehat{\sigma}_1^2\right)^2$ by $2\widehat{\sigma}_1^4$, $\widehat{N}_3$ becomes $\widehat{N}_1$. Since each can be represented as a positively-weighted sum of non-negative definite matrices, $\widehat{N}_1$, $\widehat{N}_2$, $\widehat{N}_3$ and $\widehat{N}_4$ are desirably guaranteed non-negative definite. Note that unlike $\widehat{N}_1$ and $\widehat{N}_4$, $\widehat{N}_2$ and $\widehat{N}_3$ are also consistency-robust to variation in the fourth moment of $\varepsilon_s$.

Note that Theorem 3 of Appendix B justifies the large sample approximations

$$\widehat{\tau} \simeq \mathcal{N}\left(\tau, n^{-1} \widehat{M}^{-1} \widehat{N}_i \widehat{M}^{-1}\right), \quad i = 1, 2, 3, 4,$$

with Theorem 4 then using the delta method to justify (25).

## 4. Monte Carlo study of finite-sample performance

We generate designs as follows. First, we generate a random set of 2000 pairs $(r_1, r_2)$ iid as $(R_1, R_2)$, where $R_1$ and $R_2$ are two independent random variables uniformly distributed in the interval $[0, 100]$. Each pair $(r_1, r_2)$ is a coordinate of the square lattice $[0, 100] \times [0, 100]$. We then generate samples of size $n$ $(n < 2000)$, consisting of the $n$-nearest-neighbours to the centre of the square lattice (i.e. the point $(50, 50)$). The same coordinates are used in each Monte Carlo simulation.

We compare four alternative covariance specifications. On the one hand, we consider SAR $(1)$, SMA $(1)$ and MESS$(1)$ specifications, i.e (11), (12) and (14) respectively, with $m = 1$, all of which involve weight matrices. We also consider an isotropic covariance function (8) with $\phi_1 = 1$, or equivalently (7) with $\phi_1 = 0.5$,

i.e. the exponential covariance function $\exp(-|s|/\phi_2)$. In Tables 1, 2 and 4, we use the same parameter values for the different models when generating spatial data according to the different designs. On the other hand, we consider the same weight matrix $W_1$ for the non-isotropic specifications. The weights are constructed by the function "makeneighbors" taken from J. LeSage's MATLAB code (http://www.spatial-econometrics.com), which has been used before by Han and Lee (2013) in the context of non-nested testing of SAR vs MESS models. This function generates a row-normalized weight matrix $W_1 = [w_{ij}]_{i,j=1}^{n}$ based on $k$ nearest neighbours, i.e. $w_{ij} = w_{ij}^* / \sum_{j=1}^{n} w_{ij}^*$ where $w_{ij}^* = 1$ if the location $j$ is one of the $k$ nearest neighbours of the location $i$, $i \neq j$, and $w_{ij}^* = 0$ otherwise. The maximum eigenvalue of $W_1$ is 1. We chose $k = 5$, as in Han and Lee (2013). These weights produce covariance matrices satisfying Assumptions 2, 3 and 8 of Appendix A for the models and parameter values chosen. We compare results for alternative parameter values and weight functions in Table 3. These Monte Carlo experiments are based on 2000 replications.

Table 1 provides a comparison of the level accuracy under different kurtosis scenarios using the alternative estimates $\widehat{N}_i$. We provide the proportion of rejections under $H_1$ for SAR(1) and SMA(1) specifications with parameter $\phi_1 = 0.5$ and nearest neighbour weights with $k = 5$, generating innovations $\{\varepsilon_j\}_{j=1}^{n}$ with mean zero, variance one and varying kurtoses 0, 3 and 6, resulting from standardized versions of normal, centred Gamma with shape parameter 2 and scale parameter 1, and Student's $t$ with 5 degrees of freedom, respectively. Tests based on alternative $\widehat{N}_i$ behave very similarly under normality, though sometimes there is a cost to using the robust $\widehat{N}_2$, $\widehat{N}_3$ and $\widehat{N}_4$ when they are not needed, and more surprisingly, the test based on $\hat{N}_1$ still works fairly well under serious leptokurtosis, and generally is best under leptokurtic innovations. This outcome may be explainable by the imprecision of 4th moment estimates under leptokurtosis, in particular the 8th moment of a Gamma with shape parameter 2 is 9! and the 8th moment of Student's $t$ with 5 degrees of freedom does not exist, contradicting Assumption 7 of Appendix A. Amongst the three robust estimators, $\hat{N}_4$ is easiest to compute and behaves slightly better, possibly because, unlike the other two, it uses the information that the fourth moment is constant over observations.

Table 2 provides size and power comparisons of tests with the SAR, SMA, MESS and EXP specifications under $H_1$ (horizontal) in the direction of SAR, SMA and MESS under $H_2$ (vertical) using Gaussian $\varepsilon_j$ and tests based on $\hat{N}_4$ for sample sizes of 100, 200, 500 and 1000. We consider the SAR(1) and SMA(1) models with $\phi_1 = 0.5$ and the MESS(1) model with $\phi_1 = 0.65$ (prompted by a relationship between maximum row sums of $I_n - \phi_1 W_1$ and $\exp(-\phi_1 W_1)$ derived by LeSage and Pace (2007, p. 193), though these only become close as $\phi_1 \to 0$). The exponential isotropic model (EXP) given by (8) with $\phi_1 = 1$, is simulated with $\phi_2 = 1$ conditional on the fixed location points, using the lower–upper triangular decomposition of the covariance matrix, as suggested by Davis (1987) and implemented with a MATLAB routine (http://www.mathworks.com/matlabcentral/fileexchange/27613-random-field-simulation). The normal approximation is fairly good for the larger sample sizes (500 and 1000) except when testing SAR, SMA or MESS in the direction of EXP. The EXP likelihood under misspecification is badly behaved and the parameter estimates often fall on boundaries in many experiments. This is the case for various $\phi_1$ values we have tried. Performance under $H_1$ is very good when testing EXP in the direction of the other models. The EXP model is quite different from the others and it is not difficult to reject this specification in the direction of non-isotropic covariances. However, it is hard to discriminate between SAR, SMA and MESS for the smaller sample sizes, and MESS is difficult to reject in the direction of SMA even for large $n$.

**Table 1**

Size comparisons for alternative hypotheses and $\hat{N}_i$. Percentage of rejections with nominal 5% level.

| $\varepsilon_j \sim N(0, 1)$, Kurtosis Excess = 0 | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $H_1/H_2$ | SAR/SMA | | | SAR/MESS | | | SMA/SAR | | | SMA/MESS | | |
| | n | | | | | | | | | | | |
| $\hat{N}_i$ | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 |
| $\hat{N}_1$ | 0.60 | 4.80 | 5.00 | 1.30 | 3.35 | 4.75 | 1.40 | 4.15 | 4.55 | 3.35 | 4.60 | 4.75 |
| $\hat{N}_2$ | 1.00 | 4.90 | 4.95 | 1.60 | 3.25 | 4.75 | 1.35 | 3.70 | 4.30 | 3.40 | 4.50 | 4.50 |
| $\hat{N}_3$ | 0.65 | 4.85 | 4.95 | 1.65 | 3.30 | 4.65 | 1.40 | 4.05 | 4.55 | 3.20 | 4.70 | 4.75 |
| $\hat{N}_4$ | 0.70 | 4.85 | 4.95 | 1.75 | 3.40 | 4.75 | 1.45 | 4.25 | 4.50 | 3.20 | 4.55 | 4.75 |
| $\varepsilon_j \sim (\Gamma(2, 1) - 2)/\sqrt{2}$, Kurtosis Excess = 3 | | | | | | | | | | | |
| | n | | | | | | | | | | | |
| $\hat{N}_i$ | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 |
| $\hat{N}_1$ | 1.05 | 5.25 | 6.35 | 2.10 | 5.90 | 6.70 | 2.55 | 4.35 | 4.40 | 3.55 | 4.60 | 4.70 |
| $\hat{N}_2$ | 0.45 | 3.15 | 4.00 | 0.95 | 4.10 | 4.70 | 0.85 | 2.35 | 3.20 | 3.25 | 3.30 | 3.45 |
| $\hat{N}_3$ | 0.30 | 3.25 | 4.35 | 0.95 | 4.05 | 4.90 | 0.85 | 2.80 | 3.30 | 3.30 | 3.05 | 3.45 |
| $\hat{N}_4$ | 0.35 | 3.85 | 4.60 | 0.95 | 4.60 | 5.15 | 0.85 | 3.10 | 3.30 | 3.45 | 3.15 | 3.60 |
| $\varepsilon_j \sim t_5/\sqrt{5/3}$, Kurtosis Excess = 6 | | | | | | | | | | | |
| | n | | | | | | | | | | | |
| $\hat{N}_i$ | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 |
| $\hat{N}_1$ | 1.05 | 6.10 | 5.60 | 2.30 | 6.40 | 5.10 | 1.85 | 5.25 | 5.60 | 3.20 | 4.50 | 6.30 |
| $\hat{N}_2$ | 0.40 | 3.70 | 3.25 | 1.20 | 3.85 | 3.35 | 0.70 | 3.10 | 2.85 | 2.60 | 2.35 | 3.65 |
| $\hat{N}_3$ | 0.25 | 3.95 | 3.50 | 1.10 | 3.75 | 3.55 | 0.85 | 3.20 | 3.05 | 2.75 | 2.60 | 3.90 |
| $\hat{N}_4$ | 0.25 | 4.20 | 3.55 | 1.15 | 4.05 | 3.65 | 0.90 | 3.65 | 3.20 | 2.60 | 2.50 | 4.15 |

**Table 2**

Size and power comparison using Gaussian $u_j$ and $\hat{N}_4$.

| $H_2 \setminus H_1$ | n | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | % Rejections under $H_1$ | | | | % Rejections under $H_2$ | | | |
| | | SAR | SMA | MESS | EXP | SAR | SMA | MESS | EXP |
| SAR | 100 | | 1.45 | 1.40 | 0.90 | | 1.12 | 2.70 | 86.15 |
| | 200 | | 2.75 | 2.95 | 3.20 | | 23.60 | 8.00 | 99.15 |
| | 500 | | 4.25 | 3.65 | 4.70 | | 52.80 | 15.65 | 100 |
| | 1000 | | 4.50 | 4.60 | 5.15 | | 82.45 | 31.05 | 100 |
| SMA | 100 | 0.70 | | 1.20 | 2.05 | 1.20 | | 1.10 | 39.70 |
| | 200 | 2.75 | | 2.55 | 3.65 | 6.40 | | 3.00 | 89.30 |
| | 500 | 4.85 | | 3.90 | 4.65 | 19.25 | | 7.35 | 100 |
| | 1000 | 4.95 | | 4.55 | 5.10 | 35.70 | | 13.45 | 100 |
| MESS | 100 | 2.50 | 3.20 | | 1.40 | 3.80 | 5.65 | | 6.40 |
| | 200 | 3.80 | 3.10 | | 3.60 | 7.85 | 9.15 | | 98.55 |
| | 500 | 4.70 | 4.55 | | 4.65 | 19.60 | 21.10 | | 100 |
| | 1000 | 5.40 | 4.75 | | 5.30 | 37.35 | 38.55 | | 100 |

Of course, the discriminating ability of the tests depends greatly on the distance between the competing models. This is illustrated in the following Monte Carlo experiments.

Table 3 demonstrates how power depends on the underlying processes. Testing MESS(1) against SMA(1) performs comparatively worse than the other tests in Table 2. We investigate behaviour under $H_2$, i.e. for SMA(1), with $\phi_1 = 0.5, 0.6, 0.7, 0.8$ and 0.9. We also consider tests using $W_1$ computed with different numbers $k$ of nearest neighbours. We also use symmetrized nearest neighbour weights based on J. LeSage's MATLAB routine "fsym_neighbors2" for different $\phi_1$. It uses $W_1 = A^{-1/2}C_k A^{-1/2}$, with $C_k = \sum_{\ell=1}^{k} \rho^\ell S_{(\ell)}$, where $\rho \in (0, 1)$, $S_{(\ell)} = \left[ s_{(\ell)}^{(i,j)} \right]_{i,j=1}^{n}$ with $s_{(\ell)}^{(i,j)} = 1$ if location $i$ is the $\ell$th nearest neighbour of $j$ or $j$ is the $\ell$th nearest neighbour of $i$, and $s_{(\ell)}^{(i,j)} = 0$ otherwise, and $A = diag\left\{ \sum_{i=1}^{n} C_k^{(1,i)}, \ldots, \sum_{i=1}^{n} C_k^{(n,i)} \right\}$ with $C_k = \left[ C_k^{(i,j)} \right]_{i,j=1}^{n}$. The maximum eigenvalue of the resulting $W_1$ is 1, and the corresponding covariance matrices satisfy Assumptions 2, 3 and 8 of Appendix A. We took $\rho = 0.8$ and $k = 5$. The symmetrized nearest

**Table 3**

Power: % rejections under $H_2$. $H_1 : MESS$ vs $H_2 : SMA$. Tests using Gaussian $u_j$ and $\hat{N}_4$.

| $\phi_1$ | NN, $k = 5$ | | | | $k$ | ANN | |
|---|---|---|---|---|---|---|---|
| | ANN | | SNN | | | $\phi_1 = 0.5$ | |
| | n | | | | | n | |
| | 500 | 1000 | 500 | 1000 | | 500 | 1000 |
| 0.5 | 7.35 | 13.45 | 6.95 | 13.45 | 4 | 7.85 | 14.70 |
| 0.6 | 9.70 | 18.90 | 9.05 | 16.50 | 5 | 7.35 | 13.45 |
| 0.7 | 12.9 | 28.15 | 12.05 | 25.95 | 6 | 9.25 | 11.65 |
| 0.8 | 19.0 | 40.75 | 15.45 | 31.10 | 10 | 6.65 | 7.60 |
| 0.9 | 23.5 | 52.80 | 17.55 | 36.80 | 20 | 4.75 | 5.30 |

neighbours are denoted as SNN and the asymmetric ones, used in Tables 1 and 2, are denoted as ANN. Power very much depends on $W_1$ and $\phi_1$.

Table 4 provides size and power for tests comparing the same models as in Table 1 but where the $u_j$ are unobserved and tests are based on least squares residuals $\hat{u}_j$ for (3) with $g(z_j; \beta_0) = \beta_{10} + \beta_{20}z_j$, $\beta_0' = (\beta_{10}, \beta_{20}) = (1, 1)$. There is some effect of

estimating the nuisance parameters $\beta_{10}$ and $\beta_{20}$, but it seems to disappear as sample size increases.

As in many other circumstances a bootstrap can improve finite sample accuracy. A residual naive bootstrap resampling mimics the behaviour of the test under the null hypothesis. A random sample with replacement $\{\hat{\varepsilon}_j^*\}_{j=1}^n$ from $\{\hat{\varepsilon}_j\}_{j=1}^n$, with $\hat{\varepsilon} = B(\hat{\theta}_1)^{-1}\hat{u}$, forms a basis for a bootstrap resample $\hat{u}^* = B(\hat{\theta}_1)\hat{\varepsilon}^*, j = 1, \ldots, n$, which imposes the restriction under the null $H_1$. Critical values of the asymptotically pivotal test statistic $\hat{\eta} = \sqrt{n}LR/\left(\hat{e}'\hat{M}^{-1}\hat{N}_4\hat{M}^{-1}\hat{e}\right)$ are approximated by its bootstrap analogs, which are expected to be more accurate than the standard normal counterparts. Bootstrap critical values are approximated by Monte Carlo: we generate $m$ bootstrap resamples $\left\{\hat{u}_j^{*(l)}\right\}_{l=1}^m$ and the corresponding test statistics $\left\{\hat{\eta}^{*(l)}\right\}_{l=1}^m$, then $H_1$ is rejected at the $\alpha 100\%$ level in the direction of $H_2$ when $\hat{\eta} \geq c_{\alpha/2}^*$ or $\hat{\eta} \leq c_{1-\alpha/2}^*$, where $c_\alpha^* = \inf\left\{c \in \mathbb{R}^+ : m^{-1}\sum_{l=1}^m 1_{\{\hat{\eta}^{*(l)} \geq c\}} \leq \alpha\right\}$. Table 4 provides sizes with SAR(1) $H_1$ in the direction of SMA(1) $H_2$ with innovations generated as a standard normal and as leptokurtic Student's $t$ with 5 degrees of freedom. Here, we use only 1000 Monte Carlo experiments and 500 resamples to approximate the bootstrap critical values. The bootstrap tests exhibit excellent accuracy even for $n$ as small as 50, and even in the leptokurtic case. One can save the trouble of computing the scale $\hat{e}'\hat{M}^{-1}\hat{N}_4\hat{M}^{-1}\hat{e}$, at the price of worse accuracy, by implementing the bootstrap test directly on $\sqrt{n}LR$.

We examined sensitivity of the test to the way locations are derived. Rather than taking the $n$ locations as the $n$ nearest neighbours of the centre of the lattice (50, 50), we consider three other cases, where locations are placed at each quadrant of the lattice with centre (50, 50) at different proportions. That is we consider $n_i$ nearest neighbours of the centre of each quadrant $i$, $i = 1, 2, 3, 4$, with $n = n_1 + n_2 + n_3 + n_4$. The centres of quadrants I, II, III and IV are (75, 75), (75, 25), (25, 25) and (24, 75), respectively.

Figure 1 plots the locations for the three cases considered. Case 0 is the case considered in the previous simulations where the $n$ locations are nearest neighbours of the centre (50, 50); in Case 1, $n_1 = n/2, n_2 = n_3 = n/5$, and $n_4 = n/10$; in Case 2, $n_1 = 7n/10$, and $n_2 = n_3 = n_4 = n/10$; in Case 3, $n_1 = 8n/10, n_3 = n/10$, and $n_4 = 5n/100$. Table 6 reports percentage of rejections under the null and alternative hypotheses, in each of these scenarios, when testing SMA against SAR, using the four alternative $N$ estimators and with standard normal $\varepsilon_j$. The test is not particularly sensitive to the different location structures. We also employed $\varepsilon_j$ with different kurtosis excess, as in Table 1, and again the proportions of rejections under $H_0$ and $H_1$ are very similar across the four cases so the results are not displayed.

## 5. Final comments

In line with Table 4 of the previous section, under regularity conditions our tests remain valid when the $u_j$ are unobservable disturbances in a parametric model such as (2) and estimates of the correlation and scale parameters of the $u_j$ for the $H_i$ are based instead on residuals, as discussed in Section 1. In (2), the preliminary estimate of $\beta_0$, likely one motivated by uncorrelated and homoscedastic $u_j$, would need to be shown to be $\sqrt{n}$-consistent in the presence of possible correlation and heteroscedasticity, and this is relatively straightforward to establish, especially in (3), compared to the asymptotic theory for kernel nonparametric regression estimates under (28) in Robinson (2011). The rest of the verification that the $u_j$ can be replaced by residuals is lengthy but straightforward, under standard additional conditions. Table 5 of the previous section suggested that improved level accuracy can be achieved by bootstrapping, and theoretical justification could be sought. It may be of value to extend our focus on correlation to test between models that also entail different parameterization of the means of observations, for example different choices of $g$ in (3), such as testing between a linear and a nonlinear model or between linear models involving non-nested selections of explanatory variables.

**Table 4**

Size and power using Gaussian $u_j$ and $\hat{N}_4$. Tests based on residuals of simple linear regression.

| $H_2 \setminus H_1$ | $n$ | Size | | | | Power | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | % Rejections under $H_1$ | | | | % Rejections under $H_2$ | | | |
| | | SAR | SMA | MESS | EXP | SAR | SMA | MESS | EXP |
| | 100 | | 1.80 | 1.35 | 1.80 | | 6.60 | 4.20 | 80.00 |
| SAR | 200 | | 2.50 | 2.80 | 3.60 | | 20.55 | 8.40 | 98.00 |
| | 500 | | 4.00 | 4.05 | 4.35 | | 47.75 | 18.60 | 100 |
| | 1000 | | 4.45 | 4.10 | 4.80 | | 61.15 | 33.70 | 100 |
| | 100 | 1.15 | | 0.90 | 1.20 | 3.35 | | 1.60 | 35.25 |
| SMA | 200 | 2.80 | | 2.70 | 3.15 | 8.00 | | 4.05 | 87.60 |
| | 500 | 4.70 | | 3.95 | 3.80 | 17.85 | | 7.80 | 99.90 |
| | 1000 | 4.75 | | 4.55 | 4.65 | 32.15 | | 12.70 | 100 |
| | 100 | 3.45 | 2.50 | | 1.40 | 5.70 | 4.15 | | 8.95 |
| MESS | 200 | 3,75 | 2.65 | | 3.15 | 10.20 | 8.25 | | 97.50 |
| | 500 | 3.80 | 3.85 | | 3.85 | 20.40 | 18.15 | | 100 |
| | 1000 | 4.05 | 5.55 | | 4.70 | 36.20 | 37.20 | | 100 |

**Table 5**

$H_1$ : SAR vs $H_2$ : SMA. Bootstrap and asymptotic tests. Size: % Rejections under $H_1$.

| | $\varepsilon_j \sim N(0, 1)$ | | | | | | $\varepsilon_j \sim t_5$ | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Bootstrap | | | Asymptotic | | | Bootstrap | | | Asymptotic | | |
| $n \setminus 100\alpha\%$ | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% | 1% | 5% | 10% |
| 50 | 1.00 | 5.10 | 8.00 | 0.00 | 0.20 | 1.50 | 0.70 | 5.70 | 11.40 | 0.00 | 0.00 | 0.90 |
| 100 | 0.90 | 5.50 | 10.20 | 0.00 | 0.50 | 4.20 | 0.70 | 5.60 | 11.20 | 0.00 | 0.50 | 3.20 |
| 200 | 1.30 | 6.70 | 13.40 | 0.10 | 2.50 | 7.60 | 1.60 | 6.10 | 11.00 | 0.10 | 0.80 | 5.00 |

**Table 6**
Size & Power comparisons for testing $H_0$ : $SMA$ versus $H_1$ : $SAR$. Alternative locations with different density of points at each quadrant. Nominal 5% level. $\varepsilon_j \sim N(0, 1)$.

| | n | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\hat{N}_i$ | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 |

Percentage of rejections under $H_0$

| $\hat{N}_i$ | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 | 100 | 500 | 1000 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\hat{N}_1$ | 1.40 | 4.15 | 4.55 | 1.60 | 4.20 | 4.50 | 1.80 | 3.80 | 4.80 | 1.55 | 4.05 | 4.05 |
| $\hat{N}_2$ | 1.35 | 3.70 | 4.30 | 1.45 | 4.20 | 4.35 | 1.70 | 3.85 | 4.65 | 1.45 | 4.15 | 3.90 |
| $\hat{N}_3$ | 1.40 | 4.05 | 4.55 | 1.70 | 4.10 | 4.40 | 2.10 | 3.75 | 4.80 | 1.80 | 4.05 | 4.05 |
| $\hat{N}_4$ | 1.45 | 4.25 | 4.50 | 1.45 | 3.90 | 4.15 | 1.50 | 3.65 | 4.55 | 1.65 | 3.90 | 3.90 |

Percentage of rejections under $H_1$

| $\hat{N}_1$ | 10.8 | 53.2 | 82.3 | 12.6 | 50.0 | 78.3 | 11.0 | 54.5 | 80.5 | 11.0 | 53.7 | 81.8 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\hat{N}_2$ | 11.2 | 53.4 | 82.3 | 12.8 | 49.6 | 78.5 | 11.5 | 54.3 | 80.9 | 11.4 | 53.8 | 81.6 |
| $\hat{N}_3$ | 10.5 | 53.1 | 82.4 | 12.9 | 49.8 | 78.5 | 11.3 | 54.6 | 80.5 | 11.3 | 53.9 | 81.8 |
| $\hat{N}_4$ | 10.9 | 52.9 | 82.5 | 12.5 | 49.5 | 78.1 | 11.5 | 54.1 | 80.7 | 11.5 | 54.0 | 81.2 |
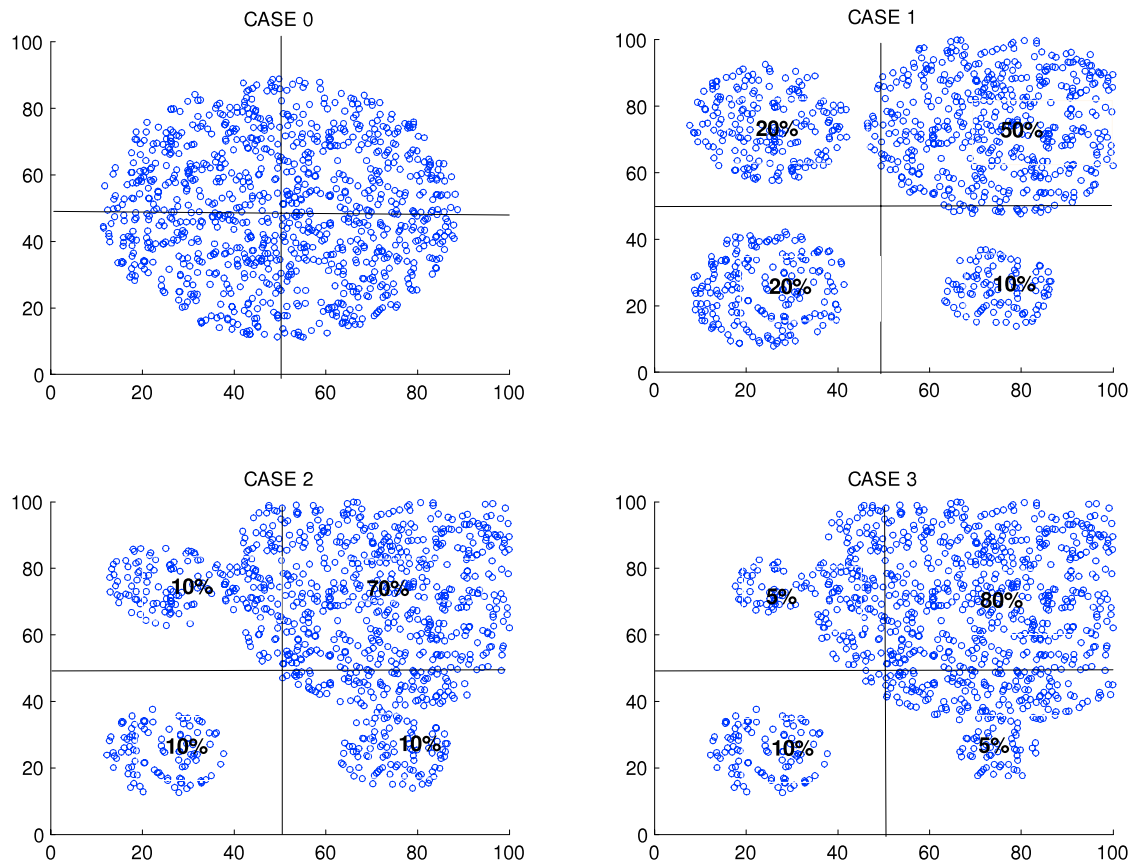


**Fig. 1.** Simulated locations for the four cases, $n = 1, 000$.

## Acknowledgements

## Appendix A. Regularity conditions and discussion

The first five assumptions are imposed for consistency of our parameter estimates (Theorem 1 in Appendix B).

**Assumption 1.** Under $H_1$, for all sufficiently large $n$, the $u_j$ have uniformly bounded fourth moment, and, denoting by $\kappa_{1jklm}$ the fourth cumulant of $u_j, u_k, u_l, u_m$,

$$\lim_{n\to\infty} n^{-2} \sum_{j,k,l,m=1}^{n} \kappa_{1jklm}^2 = 0.$$

This condition of weak dependence with respect to fourth cumulants holds trivially on the one hand if $u_j$ is Gaussian, and on the other if the $u_j$ are independent. It will also hold under the linear process assumption imposed later for the central limit theorem, indeed there $\sum_{j,k,l,m=1}^{n} \kappa_{1jklm}^2 = O(n)$.

For a real matrix $A$, denote by $\|A\|$ the spectral norm of $A$, i.e. the square root of the largest eigenvalue of $A'A$. In view of the Gaussian pseudo-likelihood employed, the Euclidean norm $\|A\|_2 = (tr(A'A))^{1/2}$ arises naturally, and as well as the standard

norm inequality $\|AB\| \leq \|A\| \|B\|$ our proofs use the inequality

$$\|AB\|_2 \leq \|A\|_2 \|B\|. \tag{32}$$

**Assumption 2.** For $i = 1, 2$

$$\overline{\lim_{n \to \infty}} \sup_{\theta_i \in R_i} \left( \|\Omega_i(\theta_i)\| + \left\|\Omega_i^{-1}(\theta_i)\right\| \right) < \infty.$$

**Assumption 3.** For $i = 1, 2$, for any $\theta_i^\dagger \in R_i$ and any $\eta > 0$, there exists $\varepsilon > 0$ such that

$$\overline{\lim_{n \to \infty}} \sup_{\theta_i: \left\|\theta_i - \theta_i^\dagger\right\| < \varepsilon; \theta_i \in R_i} \left\| \Omega_i(\theta_i) - \Omega_i\left(\theta_i^\dagger\right) \right\| < \eta. \tag{33}$$

Notice that Assumptions 2 and 3 imply that (33) holds with $\Omega_i(\theta_i) - \Omega_i\left(\theta_i^\dagger\right)$ replaced by $\Omega_i^{-1}(\theta_i) - \Omega_i^{-1}\left(\theta_i^\dagger\right)$. There is interest in checking Assumptions 2 and 3 under more primitive conditions, given the specifications of the $\Omega_i$. To place the assumptions in perspective, for equally-spaced time series, when $H_i$ implies stationarity $\Omega_i$ is a Toeplitz matrix and Assumption 2 is satisfied if the (spectral density) function $f(\lambda; \theta_i) = (2\pi)^{-1} \sum_{j,k:|j-k|=l} \omega_{ijk}(\theta_i) \cos l\lambda$ is bounded and bounded away from zero on $\lambda \in (-\pi, \pi]$, uniformly in $\theta_i \in R_i$, while Assumption 3 is satisfied by continuity of $f(\lambda; \theta_i)$ in $\theta_i$. These observations are straightforwardly extended in case of regular spatial or spatio-temporal lattices. For irregularly-spaced data, there is less scope for finding comprehensible sufficient conditions for Assumptions 2 and 3, because the properties of both the underlying process (denoted $U$ in the Section 2) and the regime generating the observation points are generally entwined in a complicated way in the $\Omega_i(\theta_i)$. However, a combination of stationary weak dependence in $U$ and a degree of regularity (lack of trending in the degree of sparseness of observations) would be expected to suffice. An advantage of Assumptions 2 and 3 is their relative simplicity. When the $H_i$ model can be naturally factored as $\Omega_i = B_i B_i'$, where $B_i$ is a known matrix function of $\theta_i$, Assumptions 2 and 3 (and subsequent assumptions) can be written in terms of $B_i$. This is the case in (4), where in each case a particular inversion must generally be selected from several possibilities, as well as in models of form (9) and (10), where $B_i = S_i^{-1}$. However, such models are readily covered also by our assumptions on $\Omega_i$, whereas for some other models (e.g. (7) and (8)), though of course $\Omega_i$ admits a factorization for any $\theta_i$, the factors need not have a simple closed form representation as functions of $\theta_i$.

With $\tilde{Q}_{10} = \tilde{Q}_1(\theta_{10})$,

$$\tilde{Q}_1 - \tilde{Q}_{10} = \frac{1}{2} \log \left\{ \frac{1}{n} tr\left(\Omega_1^{-1} \Omega_{10}\right) \Big/ \left|\Omega_1^{-1} \Omega_{10}\right|^{1/n} \right\}, \tag{34}$$

which is guaranteed to be non-negative by the inequality between arithmetic and geometric means. An identifiability condition for $\theta_{10}$ is:

**Assumption 4.** $\theta_{10} \in R_1$ and for all $\theta_1 \in R_1 \setminus \theta_{10}$,

$$\lim_{n \to \infty} \frac{1}{n} tr\left(\Omega_1^{-1} \Omega_{10}\right) \Big/ \left|\Omega_1^{-1} \Omega_{10}\right|^{1/n} > 1, \tag{35}$$

where the limit is assumed to exist.

For the SAR(1) special case, write $S = S(\theta) = I_n - \theta W$, $S_0 = S(\theta_0)$. Then (35) can be written

$$\lim_{n \to \infty} \frac{1}{n} tr\left(SS_0^{-1} S_0^{-1'} S'\right) \Big/ \left|SS_0^{-1} S_0^{-1'} S'\right|^{1/n} > 1. \tag{36}$$

Since $SS_0^{-1} = I_n + (\theta_0 - \theta) W$, and thus
$\left|\Omega_1^{-1} \Omega_{10}\right| = |I_n + (\theta_0 - \theta) W|^2$,
$tr\left(SS_0^{-1} S_0^{-1'} S'\right) = tr\left(I_n + (\theta_0 - \theta) W\right) \left(I_n + (\theta_0 - \theta) W\right)'$,
the strict inequality in (36) holds for all $\theta \neq \theta_0$ except under some degeneracy in the limit behaviour of $W$. The corresponding condition of Lee (2004) is given in his equation (4.2) as

$$\lim_{n \to \infty} \left( \frac{1}{n} \log \left|\sigma_0^2 S_0^{-1} S_0^{-1'}\right| - \frac{1}{n} \log \left|\sigma^2(\theta) S^{-1} S^{-1'}\right| \right) \neq 0, \tag{37}$$

where $\sigma^2(\theta) = \sigma_0^2 tr\left(SS_0^{-1} S_0^{-1'} S'\right)/n$ and our notation differs somewhat from his. The left side of (37) can thus be written

$$\lim_{n \to \infty} \left( \frac{1}{n} \log \left|S_0^{-1} S_0^{-1'}\right| - \frac{1}{n} \log \left|S^{-1} S^{-1'}\right| - \frac{1}{n} \log tr\left(SS_0^{-1} S_0^{-1'} S'\right) \right)$$

$$= \lim_{n \to \infty} \left( \log \left|SS_0^{-1} S_0^{-1'} S'\right|^{1/2} - \frac{1}{n} \log tr\left(SS_0^{-1} S_0^{-1'} S'\right) \right)$$

$$= -\lim_{n \to \infty} \log \left( \frac{1}{n} tr\left(SS_0^{-1} S_0^{-1'} S'\right) \Big/ \left|SS_0^{-1} S_0^{-1'} S'\right|^{1/n} \right).$$

Thus, our identification condition (36) is the same as (37) of Lee (2004), except that ours indicate that the left side of his (37) cannot be positive so "$\neq$" in (37) can be replaced by "$<$" (and our presentation recognizes the elimination of $\sigma_0^2$).

Denote by $\theta_{2*} = \theta_{2*n}$ a *sequence* of pseudo-true values under $H_1$:

$$\theta_{2*} = \arg \min_{\theta_2 \in R_2} \tilde{Q}_2(\theta_2),$$

and write $\tilde{Q}_{2*} = \tilde{Q}_2(\theta_{2*})$, $\Omega_{2*} = \Omega_2(\theta_{2*})$. Define also

$$\sigma_{2*}^2 = \tilde{\sigma}_2^2(\theta_{2*}) = \sigma_{10}^2 n^{-1} tr\left(\Omega_{2*}^{-1} \Omega_{10}\right). \tag{38}$$

Define, for all $n$ and $\varepsilon > 0$, the neighbourhoods
$\mathcal{N}_{2\varepsilon} = \{\theta_2 : \|\theta_2 - \theta_{2*}\| < \varepsilon\}$,
and let $\bar{\mathcal{N}}_{2\varepsilon} = R_2 \setminus \mathcal{N}_{2\varepsilon}$. We have

$$\tilde{Q}_2 - \tilde{Q}_{2*} = \frac{1}{2} \log \left\{ \frac{tr\left(\Omega_2^{-1} \Omega_{10}\right)}{tr\left(\Omega_{2*}^{-1} \Omega_{10}\right)} \left|\Omega_{2*}^{-1} \Omega_2\right|^{1/n} \right\} \tag{39}$$

where $\Omega_{2*} = \Omega_2(\theta_{2*})$. Because $\theta_{2*}$ need not be constant over $n$, we identify it by the condition:

**Assumption 5.** For all sufficiently large $n$ and any $\eta > 0$, $\theta_{2*} \in R_2$ and there exists $\varepsilon > 0$ such that

$$\lim_{n \to \infty} \inf_{\theta_2 \in \bar{\mathcal{N}}_{2\varepsilon}} \left\{ \frac{tr\left(\Omega_2^{-1} \Omega_{10}\right)}{tr\left(\Omega_{2*}^{-1} \Omega_{10}\right)} \left|\Omega_{2*}^{-1} \Omega_2\right|^{1/n} \right\} > 1.$$

Our remaining assumptions are needed in asymptotic normality results for the parameter estimates (Theorems 2 and 3 in Appendix B) and for the non-nested test statistics.

**Assumption 6.** $\theta_{10}$ is an interior point of $R_1$ and, for all sufficiently large $n$, $\theta_{2*}$ is an interior point of $R_2$.

**Assumption 7.** The representation (28) holds, where $\varepsilon_s$ is a sequence of iid random variables with zero mean, variance $\sigma_{10}^2$, and finite eighth moment, $b_{js}$ can depend on $n$, $b_{js} = b_{jsn}$, and, defining

$$c_{js} = c_{jsn} = b_{js}/\omega_{1jj0}^{1/2}, \quad j = 1, \ldots, n; \ n = 1, 2, \ldots; \ s = 1, 2, \ldots,$$

we have

$$\overline{\lim_{n \to \infty}} \sup_{1 \leq j \leq n} \sum_{s=1}^{\infty} |c_{js}| + \overline{\lim_{n \to \infty}} \sup_{s \geq 1} \sum_{j=1}^{n} |c_{js}| < \infty. \tag{40}$$

The representation (28) was previously used in a spatial context by Robinson (2011), where its relevance was discussed. It implies that

$$\omega_{1jk0} = \sum_{s=1}^{\infty} b_{js} b_{ks}, \quad j, k = 1, 2, \ldots, n, \tag{41}$$

where Assumption 2 implies the $\omega_{1jj*}$ are uniformly bounded and bounded away from zero, and thus

$$\sum_{s=1}^{\infty} c_{js}^2 = 1, \quad j = 1, 2, \ldots, n. \tag{42}$$

The normalized $c_{js}$ can be compared with moving average weights in the stationary time series setting where $c_{js} = c_{j-s}$, when (40) reduces to a standard weak dependence summability condition; the eighth moment condition automatically holds under Gaussianity and is needed only to check a Lyapunov condition, otherwise finite fourth moments suffice. In models of the form (9) we can choose $b_{js}$ to be the $(j, s)$th element of $S(\theta_{10})^{-1}$, $s = 1, 2, \ldots, n$, and $b_{js} = c_{js} = 0, j \geq n + 1$. More generally, the latter equality can be satisfied if the $u_j$ are Gaussian, since they can be represented as a linear transformation of $n$ iid normal variables. In particular, this is the case for Gaussian irregularly-spaced $u_j$, with the $b_{js}, s = 1, \ldots, n$, depending on the locations as well as the underlying continuous process $U(t)$ described in Section 2; see also the discussion in Robinson (2011). If the $u_j$ are non-Gaussian, the infinite series representation is generally required to cover models such as (4), (7) and (8); at least in the regularly spaced case under broad conditions there exist linear representations of $u_s$ in terms of uncorrelated and homoscedastic innovations, with these innovations then assumed to be independent.

In much asymptotic theory for estimation of spatial weight matrix models (9) (see e.g. Lee (2004)), two other norms are used: the absolute row sum norm $\|A\|_r = \max_i \sum_j |a_{ij}|$ and the $l_\infty$ or maximum element norm $\|A\|_e = \max_{i,j} |a_{ij}|$, for a matrix $A = (a_{ij})$. Noting that for symmetric $A$, $\|A\| \leq \|A\|_r$ and $\|A\| \leq \|A\|_e$, it was desirable for Theorem 1 to rely only on spectral norm assumptions, but our central limit theorem needs $\|.\|_r$ and $\|.\|_e$. Using the definitions (26), introduce:

**Assumption 8.** For $i = 1, 2$ and $j, k, l = 1, \ldots, p_i$ and all sufficiently large $n$, on an arbitrarily small neighbourhood $\mathcal{N}_i$ of $\theta_{i*}$, the elements of $\Omega_i$ are thrice boundedly differentiable,

$$\varlimsup_{n \to \infty} \sup_{\theta_i \in \mathcal{N}_i} \left( \left\| \Omega_i^{-1} \right\|_r + \left\| \Omega_{ij} \right\|_r + \left\| \Omega_{ijk} \right\|_r + \left\| \Omega_{ijkl} \right\|_r \right) < \infty, \tag{43}$$

and for a positive sequence $h = h_n$ such that either

$$h \leq C \tag{44}$$

or

$$h^{-1} + h/n \to 0 \quad \text{as } n \to \infty, \tag{45}$$

we have

$$\varlimsup_{n \to \infty} \sup_{\theta_i \in \mathcal{N}_i} h \left( \left\| \Omega_{ij} \right\|_e + \left\| \Omega_{ijk} \right\|_e + \left\| \Omega_{ijkl} \right\|_e \right) < \infty, \tag{46}$$

In spatial statistics models such as (4), (7) and (8), the $h$ bounded case (44), where $h$ can be chosen to be constant, is appropriate, when (46) is implied by (43). The allowance for (slower-than-$n$) divergent $h$ (45) is motivated by spatial weight matrix models such as (9) and (10), where, as in Lee (2004), weight matrices are assumed to have all elements that uniformly converge to zero as $n \to \infty$. For example in the SMA(1), see (11), $\Omega_i = (I_n - \theta_{i1}W)(I_n - \theta_{i1}W)'$, and $W$ is often assumed to

satisfy $h \|W\|_e + \|W\|_r + \|W'\|_r \leq C$. Thus $\Omega_{i1} = 2\theta_{i1}WW' - W - W'$ satisfies (43) and (46), and also $\left\| \Omega_{ij} \right\|_2^2 \leq n \left\| \Omega_{ij} \right\|_e \left\| \Omega_{ij} \right\|_r$ implies that $\sup_{\theta_i \in \mathcal{R}_i} \left\| \Omega_{ij} \right\|_2 = O\left((n/h)^{1/2}\right)$. Notice that divergent $h$ is tantamount to a form of persistence, and will be reflected in slower-than-$\sqrt{n}$ convergence rates for the $\widehat{\theta}_i$.

Denote $M_{i*} = M_i(\tau_*), i = 1, 2$ and

$$M_* = \begin{pmatrix} M_{1*} & 0 \\ 0 & M_{2*} \end{pmatrix},$$

$$N_* = N(\theta_*) = n \sum_{s=1}^{n} a_{ss*} a_{ss*}' E \left(\varepsilon_s^2 - \sigma_{10}^2\right)^2 + 2\sigma_{10}^4 n \sum_{s,t=1; s \neq t}^{n} a_{st*} a_{st*}',$$

where $a_{st*} = a_{st*}(\tau_*)$ and the first expectation depends also on the 4th cumulant of $\varepsilon_j$, reference to which is suppressed. Write

$$D_i = \begin{pmatrix} I_{p_i} h^{1/2} & 0 \\ 0 & 1 \end{pmatrix}, \quad i = 1, 2, \quad D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}.$$

**Assumption 9.** The matrices

$$\Phi = \lim_{n \to \infty} DM_*D,$$

$$\Psi = \lim_{n \to \infty} DN_*D$$

exist and are positive definite.

**Appendix B. Theorems and proofs**

**Theorem 1.** *Under Assumptions 1–5 and $H_1$, as $n \to \infty$*

$$\widehat{\tau}_1 \to_p \tau_{10}, \qquad \widehat{\tau}_2 - \tau_{2*} \to_p 0.$$

**Proof of Theorem 1.** Write $\tilde{Q}_{1*} = \tilde{Q}_{10}$ for ease of notation. The following arguments apply for $i = 1, 2$ except where otherwise specified. For $\varepsilon > 0$ define the neighbourhood $\mathcal{N}_{i\varepsilon} = \{\theta_i : \|\theta_i - \theta_{i*}\| < \varepsilon\}$, and let $\bar{\mathcal{N}}_{i\varepsilon} = R_i \setminus \mathcal{N}_{i\varepsilon}$. Denoting $P_1$ probability under $H_1$

$$P_1 \left( \widehat{\theta}_i \in \bar{\mathcal{N}}_{i\varepsilon} \right) \leq P_1 \left( \inf_{\bar{\mathcal{N}}_{i\varepsilon}} Q_i \leq Q_{i*} \right)$$

$$\leq P_1 \left( \sup_{R_i} \left| Q_i - \tilde{Q}_i \right| \geq \inf_{\bar{\mathcal{N}}_{i\varepsilon}} \left( \tilde{Q}_i - \tilde{Q}_{i*} \right) \right).$$

The result follows if

$$\inf_{\bar{\mathcal{N}}_{i\varepsilon}} \left( \tilde{Q}_i - \tilde{Q}_{i*} \right) > \eta, \quad \text{all sufficiently large } n \text{ and any } \eta > 0, \tag{47}$$

and if

$$\sup_{R_i} \left| Q_i - \tilde{Q}_i \right| \to_p 0, \quad \text{as } n \to \infty. \tag{48}$$

The left side of (48) is bounded by

$$\frac{1}{2} \sup_{R_i} \log \bar{\sigma}_i^2 / \tilde{\sigma}_i^2 \leq \frac{1}{2} \sup_{R_i} \left| \bar{\sigma}_i^2 - \tilde{\sigma}_i^2 \right| / \inf_{R_i} \tilde{\sigma}_i^2.$$

By the inequality (32),

$$\tilde{\sigma}_i^2 = \frac{1}{n} tr \left( \Omega_i^{-1} \Omega_{10} \right) = \frac{1}{n} \left\| \Omega_i^{-1/2} \Omega_{10}^{1/2} \right\|_2^2$$

$$\geq \left\| \Omega_i^{1/2} \right\|^{-2} \frac{1}{n} \left\| \Omega_{10}^{1/2} \right\|_2^2 \geq \|\Omega_i\|^{-1} \left\| \Omega_{10}^{-1} \right\|^{-1},$$

so by Assumption 2,

$$\lim_{n \to \infty} \inf_{R_i} \tilde{\sigma}_i^2 > 0. \tag{49}$$

On the other hand, for given $\theta_i$, $\overline{\sigma}_i^2 - \tilde{\sigma}_i^2$ has variance under $H_1$

$$\frac{2}{n^2} \left\| \Omega_i^{-1} \Omega_{10} \right\|_2^2 + \frac{1}{n^2} \sum_{j,k,l,m=1}^n \omega_i^{jk} \omega_i^{lm} \kappa_{1jklm}$$

$$\leq \frac{2}{n^2} \left\| \Omega_i^{-1} \right\|_2^2 \| \Omega_{10} \|^2 + \frac{1}{n^2} \left( \left\| \Omega_i^{-1} \right\|_2^4 \sum_{j,k,l,m=1}^n \kappa_{1jklm}^2 \right)^{1/2}$$

$$\leq \frac{2}{n} \left( \| \Omega_{10} \|^2 + \left( \frac{1}{n^2} \sum_{j,k,l,m=1}^n \kappa_{1jklm}^2 \right)^{1/2} \right) \to 0 \quad \text{as } n \to \infty$$

by Assumptions 1 and 2 and (32), establishing pointwise convergence in probability of $\overline{\sigma}_i^2 - \tilde{\sigma}_i^2$ to zero. Uniform convergence follows from compactness of $R_i$ and noting that for any $\theta_i^\dagger \in R_i$ and small enough $\eta > 0$, we can choose $\varepsilon > 0$ such that for

$$\mathcal{N}_{i\dagger\eta} = \left\{ \theta_i : \left\| \theta_i - \theta_i^\dagger \right\| < \varepsilon \right\}$$

$$E_1 \sup_{\theta_i : \| \theta_i - \theta_i^\dagger \| < \varepsilon; \theta_i \in R_i} \left| tr \left( \left( \Omega_i (\theta_i)^{-1} - \Omega_i \left( \theta_i^\dagger \right)^{-1} \right) (uu' - \Omega_{10}) \right) \right|$$

$$\leq \left( E_1 \| u \|^2 + tr (\Omega_{10}) \right)$$

$$\times \sup_{\theta_i : \| \theta_i - \theta_i^\dagger \| < \varepsilon} \left\| \Omega_i (\theta_i)^{-1} - \Omega_i \left( \theta_i^\dagger \right)^{-1} \right\| = O(\eta n),$$

by (32) and Assumptions 2 and 3. This proves (48). Next, for $i = 2$, (47) is Assumption 5 in view of (39). For $i = 1$, by compactness $R_i$ has a finite subcover and fixing $\theta_1^\dagger \in R_1 \setminus \theta_{10}$, and for any $\varepsilon > 0$

$$\inf_{\theta_1 : \| \theta_1 - \theta_1^\dagger \| < \varepsilon; \theta_1 \in R_1} \left( \tilde{Q}_1 - \tilde{Q}_{1*} \right)$$

$$\geq \left( \tilde{Q}_{1\dagger} - \tilde{Q}_{1*} \right) - \sup_{\theta_1 : \| \theta_1 - \theta_1^\dagger \| < \varepsilon; \theta_1 \in R_1} \left| \tilde{Q}_1 - \tilde{Q}_{1\dagger} \right|, \tag{50}$$

where

$$\tilde{Q}_1 - \tilde{Q}_{1\dagger} = \frac{1}{2} \log \left\{ tr \left( \Omega_1^{-1} \Omega_{10} \right) / \left| \Omega_1^{-1} \right|^{1/n} \right\}$$

$$- \frac{1}{2} \log \left\{ tr \left( \Omega_{1\dagger}^{-1} \Omega_{10} \right) / \left| \Omega_{1\dagger}^{-1} \right|^{1/n} \right\}$$

$$= \frac{1}{2} \log \left\{ tr \left( \Omega_1^{-1} \Omega_{10} \right) / tr \left( \Omega_{1\dagger}^{-1} \Omega_{10} \right) \right\} + \frac{1}{2n} \log \left| \Omega_1^{-1} \Omega_{1\dagger} \right|$$

$$= \frac{1}{2} \log \left( 1 + \frac{tr \left( \left( \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right) \Omega_{10} \right)}{tr \left( \Omega_{1\dagger}^{-1} \Omega_{10} \right)} \right)$$

$$+ \frac{1}{2n} \log \left| I_n + \left( \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right) \Omega_{1\dagger} \right|.$$

Denoting by $\lambda_j$ and $v_j$ the $j$th eigenvalues of $\left( \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right) \Omega_{10}$ and $\left( \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right) \Omega_{1\dagger}$ respectively, by Assumption 2 the last expression is bounded by

$$\frac{1}{2} \left| \sum_{j=1}^n |\lambda_j| \right| \bigg/ tr \left( \Omega_{1\dagger}^{-1} \Omega_{10} \right) + \frac{1}{2n} \sum_{j=1}^n |v_j|$$

$$\leq C n^{-1/2} \left\{ \left( \sum_{j=1}^n |\lambda_j|^2 \right)^{1/2} + \left( \sum_{j=1}^n |v_j|^2 \right)^{1/2} \right\}$$

$$\leq C n^{-1/2} \left\| \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right\|_2$$

$$\leq C \left\| \Omega_1 - \Omega_{1\dagger} \right\|,$$

where $C$ denotes a positive generic constant and we use Assumption 2, (32) and

$$\sum_{j=1}^n \lambda_j^2 = \left\| \left( \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right) \Omega_{10} \right\|_2^2 \leq \left\| \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right\|_2^2,$$

$$\sum_{j=1}^n v_j^2 = \left\| \left( \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right) \Omega_{1\dagger} \right\|_2^2 \leq \left\| \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right\|_2^2.$$

By Assumption 3, for any $\eta > 0$ we can choose $\varepsilon$ such that for all sufficiently large $n$ the last displayed expression is bounded by $C\eta$, uniformly on $\left\{ \theta_1 : \left\| \theta_1 - \theta_1^\dagger \right\| < \varepsilon \right\}$, as therefore is $\left| \tilde{Q}_1 - \tilde{Q}_{1\dagger} \right|$. In view of (50) the proof of (47) for $i = 1$ is completed by noting that (34) and Assumption 4 imply that for some $c_\dagger > 0$, $\tilde{Q}_{1\dagger} - \tilde{Q}_{1*} \to c_\dagger$ as $n \to \infty$.

**Theorem 2.** *Under Assumptions 1–9 and $H_1$, as $n \to \infty$,*

$$n^{1/2} D^{-1} (\widehat{\tau} - \tau_*) \to_d \mathcal{N} \left( 0, \Phi^{-1} \Psi \Phi^{-1} \right).$$

**Proof of Theorem 2.** We record some preliminary calculations. For $i = 1, 2$ write

$$L_i = \frac{1}{2} \log \sigma_i^2 + \frac{1}{2n} \log |\Omega_i| + \frac{1}{2n\sigma_i^2} u' \Omega_i^{-1} u.$$

For $j = 1, \ldots, p_i$,

$$\frac{\partial}{\partial \theta_{ij}} \log |\Omega_i| = tr \left( \Omega_i^{-1} \Omega_{ij} \right), \qquad \frac{\partial}{\partial \theta_{ij}} \Omega_i^{-1} = -\Omega_i^{-1} \Omega_{ij} \Omega_i^{-1}.$$

Thus

$$\frac{\partial L_i}{\partial \theta_{ij}} = -\frac{1}{2n\sigma_i^2} tr \left\{ \Omega_i^{-1} \Omega_{ij} \Omega_i^{-1} \left( uu' - \sigma_i^2 \Omega_i \right) \right\},$$

$$\frac{\partial L_i}{\partial \sigma_i^2} = \frac{1}{2\sigma_i^2} - \frac{u' \Omega_i^{-1} u}{2n\sigma_i^4}.$$

For $i = 1$, evaluating at $\theta_1 = \theta_{10}$, $\sigma_1^2 = \sigma_{10}^2$ and under $H_1$,

$$\frac{\partial L_{10}}{\partial \theta_{1j}} = -\frac{1}{2n\sigma_{10}^2} tr \left\{ \Omega_{10}^{-1} \Omega_{1j0} \Omega_{10}^{-1} \left( uu' - \sigma_{10}^2 \Omega_{10} \right) \right\},$$

$$\frac{\partial L_{10}}{\partial \sigma_1^2} = -\frac{1}{2n\sigma_{10}^4} tr \left\{ \Omega_{10}^{-1} \left( uu' - \sigma_{10}^2 \Omega_{10} \right) \right\}.$$

For $i = 2$, evaluating at $\theta_2 = \theta_{2*}$, $\sigma_2^2 = \sigma_{2*}^2$ and under $H_1$,

$$\frac{\partial L_{2*}}{\partial \theta_{2j}} = -\frac{1}{2n\sigma_{2*}^2} tr \left\{ \Omega_{2*}^{-1} \Omega_{2j*} \Omega_{2*}^{-1} \left( uu' - \sigma_{2*}^2 \Omega_{2*} \right) \right\}$$

$$= -\frac{1}{2n\sigma_{2*}^2} tr \left\{ \Omega_{2*}^{-1} \Omega_{2j*} \Omega_{2*}^{-1} \left( uu' - \sigma_{10}^2 \Omega_{10} \right) \right\},$$

$$\frac{\partial L_{2*}}{\partial \sigma_2^2} = -\frac{1}{2n\sigma_{2*}^4} tr \left\{ \Omega_{2*}^{-1} \left( uu' - \sigma_{2*}^2 \Omega_{2*} \right) \right\}$$

$$= -\frac{1}{2n\sigma_{2*}^4} tr \left\{ \Omega_{2*}^{-1} \left( uu' - \sigma_{10}^2 \Omega_{10} \right) \right\},$$

since

$$0 = E_1 \frac{\partial L_{2*}}{\partial \theta_{2j}} = -\frac{\sigma_{10}^2}{2n\sigma_{2*}^2} tr \left\{ \Omega_{2*}^{-1} \Omega_{2j*} \Omega_{2*}^{-1} \Omega_{10} \right\} + \frac{1}{2n} tr \left\{ \Omega_{2*}^{-1} \Omega_{2j*} \right\},$$

$$0 = E_1 \frac{\partial L_{2*}}{\partial \sigma_2^2} = -\frac{\sigma_{10}^2}{2n\sigma_{2*}^4} tr \left\{ \Omega_{2*}^{-1} \Omega_{10} \right\} + \frac{1}{2\sigma_{2*}^2},$$

that is,

$$tr \left\{ \Omega_{2*}^{-1} \Omega_{2j*} \right\} = \frac{\sigma_{10}^2}{\sigma_{2*}^2} tr \left\{ \Omega_{2*}^{-1} \Omega_{2j*} \Omega_{2*}^{-1} \Omega_{10} \right\},$$

$$\sigma_{2*}^2 = \frac{\sigma_{10}^2}{n} tr\left\{\Omega_{2*}^{-1}\Omega_{10}\right\}.$$

We thus have, denoting $d_* = (\partial L_{10}/\partial \tau_1', \partial L_{2*}/\partial \tau_2')'$,

$$d_* = \sum_{s,t=1}^{\infty} a_{st*}\left(\varepsilon_s\varepsilon_t - \sigma_{10}^2\delta_{st}\right),$$

where $a_{st*} = a_{st}(\tau_*)$. Now $n^{1/2}d_*$ has mean zero and variance matrix $N_*$, and we wish to show that

$$n^{1/2}Dd_* \to_d \mathcal{N}(0, \Psi), \quad \text{as } n \to \infty. \tag{51}$$

The proof begins similarly to that of Theorem 4 of Robinson (2011), but there a linear rather than quadratic function of the $\varepsilon_i$ was involved. Since $a_{st*} = a_{ts*}$, we rewrite $d_*$ as

$$d_* = \sum_{s=1}^{\infty} a_{ss*}\left(\varepsilon_s^2 - \sigma_{10}^2\right) + 2\sum_{s=1}^{\infty} 1(s \geq 2)\sum_{t=1}^{s-1} a_{st*}\varepsilon_s\varepsilon_t$$

$$= \sum_{s=1}^{\infty} v_s, \tag{52}$$

where $1(.)$ is the indicator function and

$$v_s = \left(\varepsilon_s^2 - \sigma_{10}^2\right)a_{ss*} + 21(s \geq 2)\varepsilon_s\sum_{t=1}^{s-1} a_{st*}\varepsilon_t.$$

For a positive integer sequence $J = J_n$, increasing with $n$, write

$$d_{*a} = \sum_{s=1}^{J} v_s, \qquad d_{*b} = d_* - d_{*a}.$$

On proving that, for some $J$ sequence,

$$n^{1/2}Dd_{*b} \to_p 0, \tag{53}$$

it suffices to focus on $d_{*a}$, leading to consideration of

$$T = nE\left(Dd_{*a}d_{*a}'D\right) = n\sum_{s=1}^{J} DE\left(v_s v_s'\right)D.$$

Introduce a square matrix $Z$ such that $T = ZZ'$. For large enough $J$, $T$ is positive definite under our conditions (see (55)). For a vector $\zeta$ such that $\|\zeta\| = 1$, write

$$r_* = n^{1/2}\zeta'Z^{-1}Dd_{*a} = n^{1/2}\sum_{s=1}^{J}\zeta'Z^{-1}Dv_s.$$

Now $r_*$ has zero mean and unit variance for all $n$, and the property

$$r_* \to_d \mathcal{N}(0, 1), \quad \text{as } n \to \infty, \tag{54}$$

will follow by checking the conditions of a martingale central limit theorem, because the elements of the $v_k$, and thus the summands of $r_*$, are martingale differences. If also

$$T \to N_* \quad \text{as } J \to \infty, \tag{55}$$

the proof of (51) is completed; we omit proof of (55) as it is straightforward given our other proofs.

The details for checking (53) and (54) differ considerably from those of Robinson (2011), mainly because our $v_k$ is quadratic in the $\varepsilon_s$. First, (53) follows on showing that as $J \to \infty$,

$$E\left\|n^{1/2}Dd_{*b}\right\|^2 \to 0. \tag{56}$$

From Assumption 7 the $v_s$ are uncorrelated and the left side of (56) is bounded by

$$Cn\sum_{s=J+1}^{\infty} E\|Dv_s\|^2, \tag{57}$$

where, from (52)

$$E\|Dv_s\|^2 \leq C\|Da_{ss*}\|^2 + C1(s \geq 2)\sum_{t=1}^{s-1}\|Da_{st*}\|^2$$

$$\leq C\sum_{t=1}^{s}\|Da_{st*}\|^2.$$

$a_{ist}(\tau)$ be the $(p_i + 1) \times 1$ vector with $j$th element $a_{ijst}(\tau) = -\left(2n\sigma_i^2\right)^{-1}b_s'(\theta_1)\Omega_i^{-1}\Omega_{ij}\Omega_i^{-1}b_t(\theta_1)$ for $j = 1, \ldots, p_i$, and $-\left(2n\sigma_i^4\right)^{-1}b_s'(\theta_1)\Omega_i^{-1}b_t(\theta_1)$ for $j = p_i + 1$, and put $a_{st}(\tau) = (a_{1st}(\tau)', a_{2st}(\tau)')'$; note that $a_{st}(\tau) = a_{ts}(\tau)$.

$$D_i = \begin{pmatrix} I_{p_i}h^{1/2} & 0 \\ 0 & 1 \end{pmatrix}, \quad i = 1, 2, \quad D = \begin{pmatrix} D_1 & 0 \\ 0 & D_2 \end{pmatrix}.$$

The $(p + 2) \times 1$ vector $Da_{st*}$ has $m$th element of form $b_s'R_m b_t/n$, where $b_s = b_s(\theta_{10})$ and
$R_m = -h^{1/2}\left(2n\sigma_{10}^2\right)^{-1}\Omega_{10}^{-1}\Omega_{1m0}\Omega_{10}^{-1}, m = 1, \ldots, p_1$;
$R_{p_1+1} = -\left(2n\sigma_{10}^4\right)^{-1}\Omega_{10}^{-1}$;
$R_{p_1+1+m} = -h^{1/2}\left(2n\sigma_{20}^2\right)^{-1}\Omega_{20}^{-1}\Omega_{2m0}\Omega_{20}^{-1}, m = 1, \ldots, p_2$;
$R_{p+2} = -h^{1/2}\left(2n\sigma_{20}^4\right)^{-1}\Omega_{20}^{-1}$.
Now (57) is bounded by

$$Cn\sum_{s=J+1}^{\infty}\sum_{t=1}^{s}\|Da_{st*}\|^2 \leq \frac{C}{n}\sum_{m=1}^{p_1+p_2+2}\sum_{s=J+1}^{\infty} b_s'R_m\sum_{t=1}^{s} b_t b_t'R_m'b_s$$

$$\leq \frac{C}{n}\sum_{m=1}^{p_1+p_2+2}\left(\sum_{s=J+1}^{\infty} b_s'R_mR_m'b_s\right), \tag{58}$$

since (28) implies $\left\|\sum_{t=1}^{s} b_t b_t'\right\| \leq \|\Omega_1\| \leq C$ by Assumption 2. Denote by $r_{mjk}$ the $(j, k)$th element of $R_m$. We deduce from Assumption 8 that for $1 \leq m \leq p_1$ and $p_1 + 2 \leq m \leq p_1 + p_2 + 1$,

$$|r_{mjk}| \leq \frac{C}{h^{1/2}}, \qquad \sum_{k}|r_{mjk}| \leq Ch^{1/2},$$

while for $m = p_1 + 1$ and $m = p_1 + 1 + 2$

$$\sum_{k}|r_{mjk}| \leq C.$$

The bracketed term in (58) is

$$\sum_{s=J+1}^{\infty}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} b_{js}r_{mjk}r_{mlk}b_{ls}$$

$$\leq C\sum_{s=J+1}^{\infty}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}|c_{js}||r_{mjk}||r_{mlk}||c_{ls}|$$

$$\leq C\sum_{s=J+1}^{\infty}\sum_{j=1}^{n}\sum_{l=1}^{n}\left(\max_k|r_{mjk}|\right)|c_{js}|\left(\sum_{k=1}^{n}|r_{mlk}|\right)|c_{ls}|$$

$$\leq C\sum_{j=1}^{n}\sum_{s=J+1}^{\infty}|c_{js}|\sum_{l=1}^{n}|c_{ls}|$$

$$\leq Cn\max_j\sum_{s=J+1}^{\infty}|c_{js}| \tag{59}$$

from Assumption 7 and elementary inequalities. Also that assumption implies that, for $j = 1, \ldots, n$, for any sequence $\eta_n \downarrow 0$ as $n \to \infty$ we may choose $J_{jn}$ such that $\sum_{s=J_{jn}+1}^{\infty}|c_{js}| < \eta_n$. Thus taking $J = J_n = \max(J_{1n}, \ldots, J_{nn})$, (59) $\leq Cn\eta_n = o(n)$ as $n \to \infty$. This completes the proof of (56).

The proof of (54) follows (see e.g. Scott (1973)) on checking a Lyapunov type condition

$$\sum_{s=1}^{J} E \left\| n^{1/2} Z^{-1} D v_s \right\|^4 \to 0 \tag{60}$$

and

$$n \sum_{s=1}^{J} \left( E \left( D v_s v_s' D \mid \varepsilon_t, t \le s - 1 \right) - E \left( D v_s v_s' D \right) \right) \to_p 0. \tag{61}$$

To check (60) note first that by straightforward evaluation and the inequality $\sum z_s^2 \le \left( \sum |z_s| \right)^2$

$$E \left\| Z^{-1} D v_s \right\|^4 \le C \left\| D a_{ss*} \right\|^4 + CE \left\| \sum_{t=1}^{s-1} D a_{st*} \varepsilon_t \right\|^4$$

$$\le C \sum_{t=1}^{s} \| D a_{st*} \|^4 + C \left( \sum_{t=1}^{s-1} \| D a_{st*} \|^2 \right)^2$$

$$\le C \left( \sum_{t=1}^{s} \| D a_{st*} \|^2 \right)^2.$$

Now

$$\sum_{t=1}^{s} \| D a_{st*} \|^2 \le \frac{C}{n^2} \sum_{m=1}^{p_1+p_2+2} \left( b_s' R_m \sum_{t=1}^{s} b_t b_t' R_m' b_s \right)$$

$$\le \frac{C}{n^2} \| b_s \|^2.$$

Thus the left side of (60) is bounded by

$$\frac{C}{n^2} \sum_{s=1}^{J} \| b_s \|^4 \le \frac{C}{n^2} \sum_{s=1}^{J} \left( \sum_{j=1}^{n} c_{js}^2 \right)^2 \le \frac{C}{n^2} \sum_{s=1}^{J} \left( \sum_{j=1}^{n} |c_{js}| \right)^4$$

$$\le \frac{C}{n^2} \sum_{j=1}^{n} \left( \sum_{s=1}^{J} |c_{js}| \right) \le \frac{C}{n}$$

on applying both parts of (40) of Assumption 7 and the inequality $\sum z_s^2 \le \left( \sum |z_s| \right)^2$, to prove (60).

To prove (61), note first that $E \left( v_s v_s' \mid \varepsilon_t, t \le s - 1 \right)$ is

$$\left( 2\sigma_{10}^4 + \kappa_* \right) a_{ss*} a_{ss*}' + E \left( \varepsilon_s^3 \right) 1 (s \ge 2) \sum_{t=1}^{s-1} \left( a_{st*} a_{ss*}' + a_{ss*} a_{st*}' \right) \varepsilon_t$$

$$+ \sigma_{10}^2 1 (s \ge 2) \left( \sum_{t=1}^{s-1} a_{st*} \varepsilon_t \right) \left( \sum_{t=1}^{s-1} a_{st*} \varepsilon_t \right)',$$

and its expectation $E \left( v_s v_s' \right)$ is

$$\left( 2\sigma_{10}^4 + \kappa_* \right) a_{ss*} a_{ss*}' + \sigma_{10}^4 1 (s \ge 2) \sum_{t=1}^{s-1} a_{st*} a_{st*}'.$$

Thus the Euclidean norm of the left side of (61) is bounded by

$$n \left| E \left( \varepsilon_s^3 \right) \right| \left\| \sum_{s=2}^{J} \sum_{t=1}^{s-1} \left( A_{st} + A_{st}' \right) \varepsilon_t \right\|_2 \tag{62}$$

$$+ n\sigma_{10}^2 \left\| \sum_{s=2}^{J} \left\{ \left( \sum_{t=1}^{s-1} a_{st*} \varepsilon_t \right) \left( \sum_{t=1}^{s-1} a_{st*} \varepsilon_t \right)' - \sigma_{10}^2 \sum_{t=1}^{s-1} a_{st*} a_{st*}' \right\} \right\|_2, \tag{63}$$

writing $A_{st} = a_{st*} a_{ss*}'$. Since $\sum_{s=2}^{J} \sum_{t=1}^{s-1} A_{st} \varepsilon_t = \sum_{t=1}^{J-1} \sum_{s=t+1}^{J} A_{st} \varepsilon_t$, the square of (62) has expectation bounded by

$$Cn^2 E \left\| \sum_{t=1}^{J-1} \left( \sum_{s=t+1}^{J} A_{st} \right) \varepsilon_t \right\|_2^2 \le Cn^2 \sum_{t=1}^{J-1} \left\| \sum_{s=t+1}^{J} A_{st} \right\|_2^2,$$

where

$$\left\| \sum_{s=t+1}^{J} A_{st} \right\|_2^2 = \frac{1}{n^4} \sum_{j,k=1}^{p_1+p_2+2} \sum_{r=t+1}^{J} \sum_{s=t+1}^{J} b_r' R_j b_r b_r' R_k' b_t b_s' R_j b_s b_s' R_k' b_t$$

$$\le \frac{C}{n^4} \sum_{k=1}^{p_1+p_2+2} \left( \sum_{r=t+1}^{J} |b_t' R_k b_r| \right)^2$$

$$\le \frac{C}{n^4} \sum_{k=1}^{p_1+p_2+2} \left( \sum_{r=t+1}^{J} \sum_{l=1}^{n} \sum_{m=1}^{n} |c_{lt}| |r_{klm}| |c_{mr}| \right)^2.$$

Now for $1 \le m \le p_1$ and $p_1 + 2 \le m \le p_1 + p_2 + 1$, on the one hand

$$\sum_{r=t+1}^{J} \sum_{l=1}^{n} \sum_{m=1}^{n} |c_{lt}| |r_{klm}| |c_{mr}| \le C \sum_{l=1}^{n} \sum_{m=1}^{n} |c_{lt}| |r_{klm}|$$

$$\le Ch^{1/2} \sum_{l=1}^{n} |c_{lt}|,$$

while on the other,

$$\sum_{r=t+1}^{J} \sum_{l=1}^{n} \sum_{m=1}^{n} |c_{lt}| |r_{klm}| |c_{mr}| \le C \sum_{l=1}^{n} |c_{lt}| \sum_{m=1}^{n} |r_{klm}|$$

$$\le Ch^{-1/2} \sum_{l=1}^{n} |c_{lt}| \le Ch^{-1/2}$$

while for $m = p_1 + 1$ and $m = p_1 + 1 + 2$, these bounds hold without the respective $h^{1/2}$ and $h^{-1/2}$ factors. Thus $(62) = O_p \left( n^{-1/2} \right)$.

To deal with (63), note that

$$\sum_{s=2}^{J} \sum_{t=1}^{s-1} a_{st*} a_{st*}' \left( \varepsilon_t^2 - \sigma_{10}^2 \right)$$

$$+ \sum_{s=2}^{J} \sum_{t=1}^{s-1} \sum_{u=1}^{t-1} \left( a_{st*} a_{su*}' + a_{su*} a_{st*}' \right) \varepsilon_t \varepsilon_u$$

$$= \sum_{t=1}^{J-1} \sum_{s=t+1}^{J} a_{st*} a_{st*}' \left( \varepsilon_t^2 - \sigma_{10}^2 \right)$$

$$+ . \sum_{t=1}^{J-1} \sum_{u=1}^{t-1} \left\{ \sum_{s=t+1}^{J} \left( a_{st*} a_{su*}' + a_{su*} a_{st*}' \right) \right\} \varepsilon_t \varepsilon_u. \tag{64}$$

Now by elementary inequalities

$$n^2 E \left\| . \sum_{t=1}^{J-1} \sum_{u=1}^{t-1} \left\{ \sum_{s=t+1}^{J} \left( a_{st*} a_{su*}' + a_{su*} a_{st*}' \right) \right\} \varepsilon_t \varepsilon_u \right\|_2^2$$

$$\le Cn^2 . \sum_{t=1}^{J-1} \sum_{u=1}^{t-1} \left\| \sum_{s=t+1}^{J} a_{st*} a_{su*}' \right\|_2^2$$

$$\le \frac{C}{n^2} . \sum_{t=1}^{J-1} \sum_{u=1}^{t-1} \sum_{s=t+1}^{J} \sum_{r=t+1}^{J} b_s' R_j b_t b_s' R_k b_u b_r' R_j b_t b_r' R_k b_u$$

$$\le \frac{C}{n^2} . \sum_{t=1}^{J} \sum_{s=1}^{J} \sum_{r=1}^{J} \sum_{u=1}^{J} \sum_{j} \sum_{k} \sum_{l} \sum_{m} |c_{ls}| |r_{jlm}| |c_{mt}|$$

$$\times \sum_{l}\sum_{m}|c_{ls}||r_{klm}||c_{mu}|$$

$$\times \sum_{l}\sum_{m}|c_{lr}||r_{jlm}||c_{mt}| \sum_{l}\sum_{m}|c_{lr}||r_{jlm}||c_{mu}|$$

$$\leq \frac{C}{n^2}\cdot\sum_{t=1}^{J}\sum_{s=1}^{J}\sum_{r=1}^{J}\sum_{j}\sum_{k}\left(\sum_{l}\sum_{m}|c_{ls}||r_{jlm}||c_{mt}|\right)$$

$$\times \left(\sum_{l}\sum_{m}|c_{ls}||r_{klm}|\left|\sum_{u=1}^{J}c_{mu}\right|\right)$$

$$\times \left(\sum_{l}\sum_{m}|c_{lr}||r_{jlm}||c_{mt}|\right)\left(\sum_{l}\sum_{m}|c_{lr}||c_{mu}|\,h^{-1/2}\right)$$

$$\leq \frac{Ch^{-1/2}}{n^2}\cdot\sum_{t=1}^{J}\sum_{s=1}^{J}\sum_{j}\left(\sum_{l}\sum_{m}|c_{ls}||r_{jlm}||c_{mt}|\right)$$

$$\times \sum_{l}|c_{ls}|\,h^{1/2}\left(\sum_{l}\sum_{m}\sum_{r=1}^{J}|c_{lr}||r_{jlm}||c_{mt}|\right)$$

$$\leq \frac{C}{n^2}\cdot\sum_{j}\sum_{l}\sum_{m}\left(\sum_{s=1}^{J}|c_{ls}|\right)|r_{jlm}|$$

$$\times \left(\sum_{t=1}^{J}|c_{mt}|\right)\sum_{m}\left(\sum_{l}|r_{jml}|\right)|c_{mt}|$$

$$\leq \frac{Ch}{n},$$

for $1\leq m\leq p_1$ and $p_1+2\leq m\leq p_1+p_2+1$, the penultimate step using symmetry of $R_k$. Clearly for $m=p_1+1$ and $m=p_1+1+2$ the bound is $C/n$.

Finally, for the second part of (64),

$$n^2 E\left\|\sum_{t=1}^{J-1}\sum_{s=t+1}^{J}a_{st*}a_{st*}'\left(\varepsilon_t^2-\sigma_{10}^2\right)\right\|_2^2 \leq Cn^2\sum_{t=1}^{J-1}\left\|\sum_{s=t+1}^{J}a_{st*}a_{st*}'\right\|_2^2.$$

The $(j.k)$th element of $a_{st*}a_{st*}'$ is $b_s'R_j b_t\, b_s'R_k b_t/n^2$ so, since

$$\sum_{s=t+1}^{J}|b_s'R_j b_t| \leq C\sum_{s=t+1}^{J}\sum_{l=1}^{n}\sum_{m=1}^{n}|c_{ls}||r_{jlm}||c_{mt}|$$

$$\leq C\sum_{l=1}^{n}\sum_{m=1}^{n}|r_{jlm}||c_{mt}| \leq Ch^{1/2}\sum_{m=1}^{n}|c_{mt}|,$$

we have

$$n^2\sum_{t=1}^{J-1}\left\|\sum_{s=t+1}^{J}a_{st*}a_{st*}'\right\|_2^2 \leq \frac{Ch}{n^2}\sum_{t=1}^{J-1}\left(\sum_{m=1}^{n}|c_{mt}|\right)^2$$

$$\leq \frac{Ch}{n^2}\sum_{m=1}^{n}\sum_{t=1}^{J-1}|c_{mt}| \leq \frac{Ch}{n}.$$

This completes the proof of (51).

Next consider

$$\frac{\partial^2}{\partial\theta_{ij}\partial\theta_{ik}}\log|\Omega_i| = tr\left(\Omega_i^{-1}\Omega_{ijk}\right) - tr\left(\Omega_i^{-1}\Omega_{ik}\Omega_i^{-1}\Omega_{ij}\right),$$

$$\frac{\partial^2}{\partial\theta_{ij}\partial\theta_{ik}}\Omega_i^{-1} = \Omega_i^{-1}\Omega_{ik}\Omega_i^{-1}\Omega_{ij}\Omega_i^{-1} - \Omega_i^{-1}\Omega_{ijk}\Omega_i^{-1}$$

$$+ \Omega_i^{-1}\Omega_{ij}\Omega_i^{-1}\Omega_{ik}\Omega_i^{-1}.$$

Thus

$$\frac{\partial^2 L_i}{\partial\theta_{ij}\partial\theta_{ik}}$$

$$= -\frac{1}{2n\sigma_i^2}tr\left\{\left(\Omega_i^{-1}\Omega_{ijk}\Omega_i^{-1} - 2\Omega_i^{-1}\Omega_{ij}\Omega_i^{-1}\Omega_{ik}\Omega_i^{-1}\right)uu'\right\}$$

$$+ \frac{1}{2n}tr\left\{\Omega_i^{-1}\Omega_{ijk} - \Omega_i^{-1}\Omega_{ij}\Omega_i^{-1}\Omega_{ik}\right\},\quad j,k=1,\ldots,p_i,$$

$$\frac{\partial^2 L_i}{\partial\theta_{ij}\partial\sigma_i^2} = \frac{1}{2n\sigma_i^4}tr\left\{\Omega_i^{-1}\Omega_{ij}\Omega_i^{-1}uu'\right\},\quad j=1,\ldots,p_i,$$

$$\frac{\partial^2 L_i}{\partial\sigma_i^4} = \frac{1}{n\sigma_i^6}tr\left\{\Omega_i^{-1}uu'\right\} - \frac{1}{2\sigma_i^4}.$$

It is then readily seen that

$$E_1\frac{\partial^2 L_{10}}{\partial\tau_1\partial\tau_1'} = M_{10},\qquad E_1\frac{\partial^2 L_{2*}}{\partial\tau_2\partial\tau_2'} = M_{2*}. \tag{65}$$

Now denote

$$F = F(\tau) = \begin{pmatrix} \dfrac{\partial^2 L_1}{\partial\tau_1\partial\tau_1'} & 0 \\ 0 & \dfrac{\partial^2 L_2}{\partial\tau_2\partial\tau_2'} \end{pmatrix}.$$

We have

$$n^{1/2}N_*^{-1/2}M_*\left(\widehat{\tau}-\tau_*\right) = n^{1/2}N_*^{-1/2}D^{-1}\left(DM_*D\right)D^{-1}\left(\widehat{\tau}-\tau_*\right)$$

$$\rightarrow_d \mathcal{N}\left(0,I_{p+2}\right),$$

where, for a positive definite matrix $A$, $A^{1/2}$ denotes the unique positive definite matrix such that $A^{1/2}A^{1/2}=A$. By the mean value theorem,

$$0 = d_* + \widetilde{F}\left(\widehat{\tau}-\tau_*\right),$$

where $\widetilde{F}$ is derived from the matrix $F(\tau)$ by evaluating each row at a possibly different $\overline{\tau}$ such that $\|\overline{\tau}-\tau_*\| \leq \|\widehat{\tau}-\tau_*\|$. Thus

$$0 = Dd_* + D\widetilde{F}DD^{-1}\left(\widehat{\tau}-\tau_*\right),$$

and so

$$n^{1/2}D^{-1}\left(\widehat{\tau}-\tau_*\right) = -n^{1/2}\left(D\widetilde{F}D\right)^{-1}Dd_*.$$

It may be readily verified that

$$D\left(\widetilde{F}-F(\tau_*)\right)D \rightarrow_p 0,\qquad D\left(F(\tau_*)-M_*\right)D \rightarrow_p 0,$$

where the first step uses consistency of $\widehat{\tau}$ and the implied regularity of $F(\tau)$, and the second entails a law of large numbers in view of (65). Because of (51) the result readily follows.

Our next theorem justifies the feasible large sample approximations to the distribution of $\widehat{\tau}$:

$$\widehat{\tau}-\tau_* \simeq \mathcal{N}\left(0,n^{-1}\widehat{M}^{-1}\widehat{N}_i\widehat{M}^{-1}\right),\quad i=1,2,3.$$

**Theorem 3.** *Under Assumptions 1–9 and $H_1$, and with the $\varepsilon_s$ assumed Gaussian for $i=1$, as $n\rightarrow\infty$,*

$$\widehat{M}M_*^{-1} \rightarrow_p I_{p+2},\qquad \widehat{N}_iN_*^{-1} \rightarrow_p I_{p+2},$$

$$n^{1/2}\widehat{N}_i^{-1/2}\widehat{M}\left(\widehat{\tau}-\tau_*\right) \rightarrow_d \mathcal{N}\left(0,I_{p+2}\right),$$

*for $i=1,2,3$.*

The proof is lengthy but straightforward given previous results and is thus omitted.

**Theorem 4.** *Under Assumptions 1–9 and $H_1$, and with the $\varepsilon_s$ assumed Gaussian for $i=1$, as $n\rightarrow\infty$,*

$$\frac{LR}{\left(n^{-1}\widehat{e}'\widehat{M}^{-1}\widehat{N}_i\widehat{M}^{-1}\widehat{e}\right)^{1/2}} \rightarrow_d \mathcal{N}(0,1),\quad i=1,2,3,4. \tag{66}$$

**Proof of Theorem 4.** Writing, as in (22), $\hat{\sigma}_2^2 = \overline{\sigma}_2^2(\widehat{\theta}_i)$, $i = 1, 2$, we have as $n \to \infty$

$$LR = \log \frac{\hat{\sigma}_2^2}{\sigma_{2*}^2} - \log \frac{\overline{\sigma}_2^2(\widehat{\theta}_1, \widehat{\theta}_2)}{\sigma_{2*}^2}$$

$$= \log\left(1 + \frac{\hat{\sigma}_2^2 - \sigma_{2*}^2}{\sigma_{2*}^2}\right) - \log\left(1 + \frac{\overline{\sigma}_2^2(\widehat{\theta}_1, \widehat{\theta}_2) - \sigma_{2*}^2}{\sigma_{2*}^2}\right)$$

$$= \frac{\widehat{\sigma}_2^2 - \sigma_{2*}^2}{\sigma_{2*}^2} - \frac{\overline{\sigma}_2^2(\widehat{\theta}_1, \widehat{\theta}_2) - \sigma_{2*}^2}{\sigma_{2*}^2}$$

$$+ O_p\left(\left(\frac{\widehat{\sigma}_2^2 - \sigma_{2*}^2}{\sigma_{2*}^2}\right)^2 + \left(\frac{\overline{\sigma}_2^2(\widehat{\theta}_1, \widehat{\theta}_2) - \sigma_{2*}^2}{\sigma_{2*}^2}\right)^2\right), \quad (67)$$

from the inequality $|\log(1 + x) - x| \le Cx^2$ for $|x| < 1$. From calculations below and since $\underline{\lim}_{n \to \infty} \sigma_{2*}^2 > 0$, the remainder term in (67) can be neglected. Now

$$\overline{\sigma}_2^2(\widehat{\theta}_1, \widehat{\theta}_2) - \sigma_{2*}^2 = \widehat{\sigma}_1^2 u(\widehat{\theta}_1, \widehat{\theta}_2) - \sigma_{10}^2 u(\theta_{10}, \theta_{2*}),$$

which may be written

$$(\widehat{\sigma}_1^2 - \sigma_{10}^2) u(\theta_{10}, \theta_{2*}) + \sigma_{10}^2(u(\widehat{\theta}_1, \widehat{\theta}_2) - u(\theta_{10}, \theta_{2*}))$$

$$+ (\widehat{\sigma}_1^2 - \sigma_{10}^2)(u(\widehat{\theta}_1, \widehat{\theta}_2) - u(\theta_{10}, \theta_{2*})),$$

where, by the mean value theorem,

$$u(\widehat{\theta}_1, \widehat{\theta}_2) - u(\theta_{10}, \theta_{2*})$$

$$= \sum_{j=1}^{p_1} c_j(\overline{\theta})(\widehat{\theta}_{1j} - \theta_{1j0}) + \sum_{j=1}^{p_2} d_j(\overline{\theta})(\widehat{\theta}_{2j} - \theta_{2j*}),$$

where $\|\overline{\theta} - \theta_*\| \le \|\widehat{\theta} - \theta\|$. Thus as $n \to \infty$,

$$n^{1/2} LR - n^{1/2} e'_*(\widehat{\tau} - \tau_*) \to_p 0,$$

where $e'_* = e(\tau_*)$. But by Assumption 8 and Theorem 2.

$$n^{1/2} e'_*(\widehat{\tau} - \tau_*) = e'_* M_*^{-1} N_*^{1/2} n^{1/2} N_*^{-1/2} M_*(\widehat{\tau} - \tau_*)$$

$$= e'_* D(DM_*D)^{-1} DN_*^{1/2} n^{1/2} N_*^{-1/2} M_*(\widehat{\tau} - \tau_*)$$

$$\to_d \mathcal{N}\left(0, \zeta' \Phi^{-1} \Psi \Phi^{-1} \zeta\right),$$

where

$$\zeta = \lim_{n \to \infty} De_*$$

and using $N_*^{-1/2} M_* n^{1/2}(\widehat{\tau} - \tau_*) \to_d \mathcal{N}(0, I_{p+2})$. Equivalently

$$\frac{n^{1/2} e'_*(\widehat{\tau} - \tau_*)}{(e'_* M_*^{-1} N_* M_*^{-1} e_*)^{1/2}} \to_d \mathcal{N}(0, 1),$$

and since it is straightforwardly verified that

$$D(\widehat{e} - e_*) \to_p 0,$$

the result follows from Theorem 3.

Note that all elements of $M_*$ and $N_*$ are $O(h^{-1})$ except for the $(p_1 + 1, p_1 + 1)$th and $(p + 2, p + 2)$th, which are $O(1)$, explaining the normalizations in Assumption 9 and indicating that when $h$ diverges the $(j, p_1 + 1)$th, $j = 1, \ldots, p_1$, and $(j + p_1 + 1, p + 2)$th, $j = 1, \ldots, p_2$, elements of $\Phi$ and $\Psi$ are zero. Thus on the assumption of divergent $h$ a somewhat simpler test statistic can be justified. We have $c_j(\widehat{\tau}) = O_p(h^{-1})$, $d_j(\widehat{\tau}) = O_p(h^{-1})$, for all $j$, so taking account of the normalizations involved it is relevant that $h^{1/2} c_j(\widehat{\tau}) \to_p 0$, $h^{1/2} d_j(\widehat{\tau}) \to_p 0$, for all $j$. Thus, defining

$$e_-(\tau) = \left(0'_{p_1}, -u(\theta_1, \theta_2), 0'_{p_2}, 1\right)' / \sigma_2^2, \qquad \widehat{e}_- = e_-(\widehat{\tau}),$$

where $0_k$ is the $k \times 1$ vector of zeros, we have

$$\frac{LR}{\left(n\widehat{e}'_- \widehat{M}^{-1} \widehat{N}_i \widehat{M}^{-1} \widehat{e}_-\right)^{1/2}} \to_d \mathcal{N}(0, 1), \quad i = 1, 2, 3, 4 \text{ as } n \to \infty$$

when $h \to \infty$. However, the statistic in (66) is valid for both bounded and divergent $h$.

## References

Anselin, L., 1986. Non-nested tests on the weight structure in spatial autoregressive models.: some Monte Carlo results. J. Reg. Sci. 26.

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

Bartlett, M.S., 1955. An Introduction to Stochastic Processes. Cambridge University Press, Cambridge.

Besag, J., 1974. Spatial interaction and the statistical analysis of lattice systems (with Discussion). J. Roy. Statist. Soc. Ser. B 36, 192–236.

Burridge, P., 2012. Improving the J test in the SARAR model by likelihood-based estimation. Spat. Econ. Anal. 7, 75–107.

Burridge, P., Fingleton, B., 2010. Bootstrap inference in spatial econometrics: the J test. Spat. Econ. Anal. 5, 93–119.

Cliff, A., Ord, C., 1972. Testing for spatial autocorrelation among regression residuals. Geogr. Anal. 4, 267–284.

Cox, D.R., 1961. Tests of separate families of hypotheses. In: Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 105–123.

Cox, D.R., 1962. Further results on tests of separate families of hypotheses. J. Roy. Statist. Soc. B 24, 406–424.

Davis, M.W., 1987. Production of conditional simulation via the LU triangular decomposition of the covariance matrix. Math. Geol. 2, 91–98.

De Oliveira, V., Kedem, B., Short, D.A., 1997. Bayesian prediction of transformed Gaussian fields. J. Amer. Statist. Assoc. 92, 1422–1433.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1988. Model-based geostatistics (with discussion). Appl. Stat. 47, 299–350.

Dunsmuir, W., 1983. A central limit theorem for estimation in Gaussian stationary time series observed at unequally spaced times. Stochastic Process. Appl. 14, 279–295.

Fuentes, M., 2007. Approximate likelihood for large irregularly spaced spatial data. J. Amer. Statist. Assoc. 102, 321–331.

Gradshteyn, I.S., Ryzhik, I.M., 1994. Tables of Integrals, Series and Products, A. Jeffrey (Ed.). Academic Press, New York.

Haining, R.P., 1978. The moving average model of spatial interaction. Trans. Inst. Br. Geogr. 3, 202–225.

Hammersley, J.M., Clifford, P., 1971. Markov fields on finite graphs and lattices, unpublished manuscript.

Han, X., Lee, L.-F., 2013. Model selection using J-test for the spatial autoregressive model. Reg. Sci. Urban Econ. 43, 250–271.

Handcock, M.S., Wallis, J.R., 1994. An approach to statistical spatial–temporal modeling of meteorological fields (with discussion). J. Amer. Statist. Assoc. 89, 368–390.

Hepple, L.W., 1976. A maximum likelihood model for econometric estimation with spatial series. In: Masser, I. (Ed.), Theory and Practice in Regional Science. Pion, London, pp. 90–104.

Jin, F., Lee, L.-F., 2013. Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. Reg. Sci. Urban Econ. 63, 590–616.

Jones, R.H., Vecchia, A.V., 1993. Fitting continuous ARMA models to unequally spaced spatial data. J. Amer. Statist. Assoc. 88, 947–954.

Kelejian, H.H., 2008. A spatial J-test for model specification against a single or a set of non-nested alternatives. Lett. Spat. Res. Sci. 1, 3–11.

Kelejian, H.H., Piras, G., 2011. An extension of Kelejian's J-test for non-nested spatial models. Reg. Sci. Urban Econ. 41, 281–292.

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

LeSage, J., Pace, R.K., 2007. A matrix exponential spatial specification. J. Econometrics 140, 190–214.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC Press, New York.

Matern, B., 1986. Spatial Variation. Almaenna Foerlaget, Stockholm.

Matsuda, Y., Yajima, Y., 2009. Fourier analysis of irregularly spaced data on $R^d$. J. Roy. Statist. Soc. B 71, 191–217.

Piras, G., Lozano-Gracia, N., 2012. Spatial J-test: some Monte Carlo evidence. Stat. Comput. 22, 169–183.

Robinson, P.M., 1977. Estimation of a time series model from unequally spaced data. Stochastic Process. Appl. 6, 9–24.

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

Robinson, P.M., 2011. Asymptotic theory for nonparametric regression with spatial data. J. Econometrics 165, 5–19.

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.

Stein, M.L., 1999. Interpolation of Spatial Data. Springer-Verlag, New York.

Stein, M.L., Chi, Z., Welty, L.J., 2004. Approximating likelihoods for large spatial data sets. J. Roy. Statist. Soc. B 66, 275–296.

Tjostheim, D., 1978. Statistical spatial series modelling. Adv. Appl. Probab. 10, 130–154.

Tjostheim, D., 1983. Statistical spatial series modelling II: some further results on unilateral lattice processes. Adv. Appl. Probab. 15, 562–584.

Vecchia, A.V., 1988. Estimation and model identification for continuous spatial processes. J. Roy. Statist. Soc. B 50, 297–312.

White, H., 1982. Regularity conditions for Cox's tests of non-nested hypotheses. J. Econometrics 19, 301–318.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 49, 305–314.

Whittle, P., 1963. Stochastic processes in several dimensions. Bull. Int. Stat. Inst. 40, 974–994.