Non-Nested Testing of Spatial Correlation

Miguel A Delgado[†] and Peter M Robinson[*‡]

Universidad Carlos III and London School of Economics

November 20, 2013

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 even 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. A small Monte Carlo study of finite-sample performance is included.

JEL Classifications: C12, C21

1

Keywords: Non-nested test; spatial correlation; pseudo maximum likelihood estimation

1. INTRODUCTION

Spatial and spatio-temporal data are liable to exhibit dependence, 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 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" spa-

tial 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 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, $j = 1, 2, ...,$ we consider the rival models

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

where, for $i = 1, 2,$ θ_{i0} is an unknown $p_i \times 1$ vector, σ_{i0}^2 is an unknown positive scalar, variation-free of θ_{i0}, and $\omega_{ijk}\left(.\right)$ 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:

$$f_j(y_j; \beta_0) = u_j, \ j = 1, 2, ..., \tag{2}$$

where the f_j are known functions of their arguments and possibly of observable ex-

planatory variables varying with j, β_0 is an unknown $q \times 1$ parameter vector assumed variation-free of the θ_{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,

$$f_j(y_j; \beta_0) = y_j - g(z_j; \beta_0), \tag{3}$$

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 independent of the u_j) 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 $u = (u_1, ..., u_n)'$, there is interest in estimating the covariance matrix $E(uu')$, which has $(j, k)\,th$ element $Cov\,(u_j, u_k)$, for the sake of robust and/or efficient inference on β_0. For example, given observations $y_1..., y_n$, the linear or nonlinear least squares estimate of β_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 β_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 β_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 normalisation 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 θ_{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 β_0, such as described above, and we suppose that, for $i = 1, 2$, θ_{i0}, σ_{i0}^2 are variation-free of β_0 in (2). Given a $\sqrt{n}$-consistent estimate $\widehat{\beta}$ of β_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. Burridge (2012), Burridge and Fingleton (2010), Han and Lee (2013), Jin and Lee (2013), Kelejian (2008), Kelejian and Piras (2011), Piras and Lozano-Garcia (2012), but mainly J-tests, though Jin an Lee (2011) also develop Cox-type tests. As indicated previously, our framework is designed to cover such 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-type tests may be more suitable than J-tests when only covariance structure is at issue. 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 imposssible to satisfactorily check, but we provide some

6

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 1 lists and discusses regularity conditions. Appendix 2 presents and proves several theorems: our test statistic is a function of Gaussian pseudo-maximum likelihood estimates of the parameter vectors θ_{10} and θ_{20} in (1), and the null (taken to be the hypothesis H_1) asymptotic distribution of the test statistic depends on the null asymptotic distribution of the parameter estimates, so for these we provide consistency and asymptotic normality results which are needed in our proof of the null limit distribution of the test statistic (and which have some novelty in our general setting and represent by far the main theoretical contribution of the paper).

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, $j = 1, ..., n$, though as discussed above the u_j are likely unobservable and replaced in estimation by observable proxies. Given observations at n distinct locations $t_1, ..., t_n$ on a scalar zero-mean process $U(t)$, we make the identification $U(t_j) = u_j$, $j = 1, ..., n$, where unlike in the time series situation there is no natural ordering. It is natural to consider the 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 σ_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 ϕ_0 generically represents either θ_{01} or θ_{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 $L = (L_1, ..., L_d)$ such that $\Pi_{h=1}^d L_h^{l_h} U(t) = U(t - l)$, $l = (l_1, ..., l_d) \in \mathbb{Z}^d$, and $a(L;\phi) = \sum_{l_1=-q_{L1}}^{q_{U1}} ... \sum_{l_1=-q_{Ld}}^{q_{Ud}} a_l(\phi) \Pi_{h=1}^d L_h^{l_h}$, $b(L;\phi) = \sum_{l_1=-r_{L1}}^{r_{U1}} ... \sum_{l_1=-r_{Ld}}^{r_{Ud}} b_l(\phi) \Pi_{h=1}^d L_h^{l_h}$ for given non-negative integers q_{Lh}, q_{Uh}, r_{Lh}, r_{Uh}, $h = 1, ..., d$, 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 σ_0^2, under suitable conditions on $a(L;\phi)$ and $b(L;\phi)$, the process $U(t)$ generated by

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

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 $\Pi_{h=1}^d \sum_{l_h=-q_{Lh}}^{q_{Uh}} a_{l_h}(\phi)$, $\Pi_{h=1}^d \sum_{l_h=-r_{Lh}}^{r_{Uh}} b_{l_h}(\phi)$, 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

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

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

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

8

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

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

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, Tawn and Moyeed (1998), De Oliveira, Kedem and Short (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, Chi and Welty (2004), 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 ε an $n \times 1$ vector of these, we write

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

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 mth 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_r is the $r \times r$ 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

10

compared in non-nested testing with the $\text{SAR}(m)$ as in Han and Lee (2013) or with the $\text{SMA}(m)$. 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 of (11), (12) and (13), somewhat analogous to time series models, and the possibilty 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 noticable 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 $l = 1, ..., n$, while for the SAR(1) and MESS(1) it depends on the whole weight matrix. Such outcomes can be rationalised, 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

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

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 $v_{jk} \geq 0$,

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

or

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

In both cases, $\omega_{jk}(\phi) \to 0$ as $v_{jk} \to 0$.

11

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 (1981) 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 ε_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 ε_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 θ_i, σ_i^2 respectively a $p_i \times 1$ vector and a scalar, representing any admissible values of $\theta_{i0}, \sigma_{i0}^2$ respectively, let $\Omega_i(\theta_i)$ be the $n \times n$ matrix with $(j, k) th$ element $\omega_{ijk}(\theta_i)$, and define

$$L_i\left(\theta_i, \sigma_i^2\right) = \frac{1}{2} \log \sigma_i^2 + \frac{1}{2n} \log |\Omega_i(\theta_i)| + \frac{1}{2n\sigma_i^2} u' \Omega_i^{-1}(\theta_i) u, \qquad (15)$$

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 (15). Our estimates of $\theta_{i0}, \sigma_{i0}^2$ minimize $L_i(\theta_i, \sigma_i^2)$. For given

θ_i, $L_i\left(\theta_i, \sigma_i^2\right)$ has a minimum

$$Q_i\left(\theta_i\right) = L_i\left(\theta_i, \overline{\sigma}_i^2\left(\theta_i\right)\right) = \frac{1}{2}\log\overline{\sigma}_i^2\left(\theta_i\right) + \frac{1}{2n}\log\left|\Omega_i\left(\theta_i\right)\right| + \frac{1}{2}, \tag{16}$$

where

$$\overline{\sigma}_i^2\left(\theta_i\right) = \frac{1}{n}u'\Omega_i^{-1}\left(\theta_i\right)u = \frac{1}{n}u'\Omega_i^{-1}u, \tag{17}$$

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

$$\widehat{\theta}_i = \underset{\theta_i \in R_i}{\arg\min}\, Q_i\left(\theta_i\right), \tag{18}$$

$$\widehat{\sigma}_i^2 = \overline{\sigma}_i^2\left(\widehat{\theta}_i\right), \tag{19}$$

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

From (15), (16), the Gaussian pseudo log-likelihood-ratio statistic for testing 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. Defining, for $i = 1, 2$,

$$\widetilde{\sigma}_i^2 = \widetilde{\sigma}_i^2\left(\theta_i\right) = E_1\overline{\sigma}_i^2\left(\theta_i\right) = \sigma_{10}^2 n^{-1}tr\left(\Omega_i^{-1}\Omega_{10}\right) \tag{20}$$

and

$$\begin{aligned}\widetilde{Q}_i &= \widetilde{Q}_i\left(\theta_i\right) = \frac{1}{2}\log\left\{\widetilde{\sigma}_i^2\left(\theta_i\right)\right\} + \frac{1}{2n}\log\left|\Omega_i\left(\theta_i\right)\right| \\ &= \frac{1}{2}\log\left\{\widetilde{\sigma}_i^2\right\} + \frac{1}{2n}\log\left|\Omega_i\right|,\end{aligned} \tag{21}$$

a centred statistic is

$$2\left(Q_2\left(\widehat{\theta}_2\right) - Q_1\left(\widehat{\theta}_1\right)\right) - 2\left(\widetilde{Q}_2\left(\widehat{\theta}_2\right) - \widetilde{Q}_1\left(\widehat{\theta}_1\right)\right) = \log\frac{\overline{\sigma}_2^2\left(\widehat{\theta}_2\right)}{\overline{\sigma}_1^2\left(\widehat{\theta}_1\right)} - \log\frac{\widetilde{\sigma}_2^2\left(\widehat{\theta}_2\right)}{\widetilde{\sigma}_2^2\left(\widehat{\theta}_1\right)}.$$

13

This can be written, using (20), 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

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

where

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

with

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

Under H_1 and conditions in Appendix 1, $\overline{\sigma}_2^2\left(\widehat{\theta}_2\right) - \overline{\sigma}_2^2(\widehat{\theta}_1,\widehat{\theta}_2) = o_p\left(1\right)$, but under H_2 LR will generally have a non-zero probability limit, indicating that LR is a basis for testing H_1. After studentization, it will be found to have a limiting standard normal distribution. The studentization depends directly on an estimate of the covariance matrix in the normal approximation to the distribution of our parameter estimates. To discuss this, first denote, for $i = 1, 2$ and $j, k, l = 1, ..., p_i$,

$$\begin{aligned}\Omega_{ij} &= \Omega_{ij}\left(\theta_i\right) = \left(\partial/\partial\theta_{ij}\right)\Omega_i, \ \Omega_{ijk} = \Omega_{ijk}\left(\theta_i\right) = \left(\partial/\partial\theta_{ik}\right)\Omega_{ij}, &(22)\\ \Omega_{ijkl} &= \Omega_{ijkl}\left(\theta_i\right) = \left(\partial/\partial\theta_{il}\right)\Omega_{ijk,}\end{aligned}$$

where the existence of the derivatives is assured by Assumption 8 in Appendix 2. Write $\tau_i = \left(\theta_i', \sigma_i^2\right)'$, $\widehat{\tau}_i = \left(\widehat{\theta}_i', \widehat{\sigma}_i^2\right)'$, $i = 1, 2$, $\tau = \left(\tau_1', \tau_2'\right)'$, $\widehat{\tau} = \left(\widehat{\tau}_1', \widehat{\tau}_2'\right)'$. The large

14

sample covariance matrix of $\widehat{\tau}$ under H_1 is approximately

$$\left(\begin{array}{cc} p\lim\frac{\partial^2 L_1}{\partial\tau_1\partial\tau_1'} & 0 \\ 0 & p\lim\frac{\partial^2 L_2}{\partial\tau_2\partial\tau_2'} \end{array}\right)^{-1} E_1\left\{\left(\begin{array}{c}\frac{\partial L_1}{\partial\tau_1}\\\frac{\partial L_2}{\partial\tau_2}\end{array}\right)\left(\begin{array}{c}\frac{\partial L_1}{\partial\tau_1}\\\frac{\partial L_2}{\partial\tau_2}\end{array}\right)'\right\}\left(\begin{array}{cc} p\lim\frac{\partial^2 L_1}{\partial\tau_1\partial\tau_1'} & 0 \\ 0 & p\lim\frac{\partial^2 L_2}{\partial\tau_2\partial\tau_2'} \end{array}\right)^{-1},$$

$$(23)$$

evaluated at $\widehat{\tau}$, E_1 denoting expectation under H_1. We can approximate (23) by one of $n^{-1}\widehat{M}^{-1}\widehat{N_i}\widehat{M}^{-1}$, for $i = 1, 2, 3, 4$ with the following definitions.

Write

$$\widehat{M} = \left(\begin{array}{cc}\widehat{M_1} & 0 \\ 0 & \widehat{M_2}\end{array}\right),$$

where for $i = 1, 2$, $\widehat{M_i} = M_i(\widehat{\tau})$, $M_i = M_i(\tau)$ being the $(p_i + 1) \times (p_i + 1)$ symmetric matrix with $(j, k)th$ element M_{ijk} given by

$$\begin{aligned} M_{ijk} &= \frac{\sigma_1^2}{2\sigma_i^2 n}tr\left(\Omega_i^{-1}\Omega_{ij}\Omega_i^{-1}\Omega_{ik}\Omega_i^{-1}\Omega_1\right) \\ &+ \frac{1}{2n}tr\left(\Omega_i^{-1}\left(\Omega_{ij}\Omega_i^{-1}\Omega_{ik} - \Omega_{ijk}\right)\left(\frac{\sigma_1^2}{\sigma_i^2}\Omega_i^{-1}\Omega_1 - I_n\right)\right), \\ j, k &= 1, .., p_i; \end{aligned}$$

$$(24)$$

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

The second derivative terms, involving Ω_{2jk}, are present in some M_{2jk} due to imposing H_1 on the H_2 model, but the second trace in (24) vanishes for $i = 1$.

Our proposed $\widehat{N_1}$ is based on taking the expectation in (23) under the assumption that the u_i are Gaussian and then evaluating at $\widehat{\tau}$. Take $\widehat{N} = N(\widehat{\tau})$, where $N = N(\tau)$ is the $(p + 2) \times (p + 2)$ matrix

$$N(\tau) = \left(\begin{array}{cc} N_{11} & N_{12} \\ N_{12}' & N_{22} \end{array}\right),$$

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

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

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

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

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

Note that $N_{11} = M_1$.

Our proposed $\widehat{N}_2$, $\widehat{N}_3$ 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 ε_s appearing in the linear process representation for u_j described further in Assumption 7 of Appendix 1, which is discussed there and employed in Theorems 2-4 (on the limit distribution of $\widehat{\tau}$ and the null limit distribution of our test statistic),

$$u_j = \sum_{s=1}^{\infty} b_{js}\varepsilon_s, \ j = 1, 2, ...n, \ n = 1, 2, ... \tag{25}$$

Such a representation follows naturally if we commence from an H_1 model of form $S(\theta_{10})u = \varepsilon$ (cf. (9)), with ε_s the sth element of ε, or a model of form (4), but if we start from a model for Ω_1 we have in effect to postulate a parameterization $b_{js}(\theta_1)$ in (25), such that $b_{js} = b_{js}(\theta_{10})$, to lead to an approximate factorisation, $\Omega_1(\theta_1) \simeq B(\theta_1)B(\theta_1)'$, where the $n \times n$ matrix $B(\theta_1)$ has $(j, s)th$ element $b_{js}(\theta_1)$ and the approximation refers to the truncation, after n terms, of the series in (25) when it is non-trivially infinite or otherwise contains more than n terms. Denoting by $b_s(\theta_1)$ the $n \times 1$ vector with kth element $b_{ks}(\theta_1)$, for $i = 1, 2$ and $k, l = 1, ..., n$ let $a_{ist}(\tau)$ be the $(p_1 + 1) \times 1$ vector with jth 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 $j = 1, ..., p_i$, 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 $a_{st}(\tau) = a_{ts}(\tau)$. De-

fine the $n \times 1$ vector $\widehat{\varepsilon} = B\left(\widehat{\theta}_1\right)^{-1} u$, and denote its sth element by $\widehat{\varepsilon}_s$; the $\widehat{\varepsilon}_s$ might also be used in bootstrap versions of our tests. One alternative robust estimate is

$$\widehat{N}_2 = n\sum_{s=1}^{n} a_{ss}\left(\widehat{\tau}\right) a'_{ss}\left(\widehat{\tau}\right)\left(\widehat{\varepsilon}_s^2 - \widehat{\sigma}_1^2\right)^2 + 2n\sum_{s,t=1;s\neq t}^{n} a_{st}\left(\widehat{\tau}\right) a'_{st}\left(\widehat{\tau}\right)\widehat{\varepsilon}_s^2\widehat{\varepsilon}_t^2. \tag{26}$$

Slightly simpler ones are

$$\widehat{N}_3 = n\sum_{s=1}^{n} a_{ss}\left(\widehat{\tau}\right) a'_{ss}\left(\widehat{\tau}\right)\left(\widehat{\varepsilon}_s^2 - \widehat{\sigma}_1^2\right)^2 + 2\widehat{\sigma}_1^4 n\sum_{s,t=1;s\neq t}^{n} a_{st}\left(\widehat{\tau}\right) a'_{st}\left(\widehat{\tau}\right), \tag{27}$$

$$\widehat{N}_4 = \sum_{s=1}^{n}\left(\widehat{\varepsilon}_s^2 - \widehat{\sigma}_1^2\right)^2 \sum_{s=1}^{n} a_{ss}\left(\widehat{\tau}\right) a'_{ss}\left(\widehat{\tau}\right) + 2\widehat{\sigma}_1^4 n\sum_{s,t=1;s\neq t}^{n} a_{st}\left(\widehat{\tau}\right) a'_{st}\left(\widehat{\tau}\right). \tag{28}$$

For Gaussian ε_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 $2\widehat{\sigma}_1^4$, $\widehat{N}_3$ is seen to reduce to $\widehat{N}_1$. 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 ε_s.

Now define

$$c_j\left(\tau\right) = -\frac{\sigma_1^2}{n}tr\left(\Omega_2^{-1}\Omega_{1j}\right), \ j = 1, ..., p_1; \ d_j\left(\tau\right) = \frac{\sigma_1^2}{n}tr\left(\Omega_2^{-1}\Omega_{2j}\Omega_2^{-1}\Omega_1\right), \ j = 1, ..., p_2,$$

and

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

We have the following choice of large sample approximate null distributions (see Theorem 4 of Appendix 2):

$$LR \simeq \mathcal{N}\left(0, n^{-1}\widehat{e}'\widehat{M}^{-1}\widehat{N}_i\widehat{M}^{-1}\widehat{e}\right), \ i = 1, 2, 3, 4. \tag{29}$$

where Gaussianity of ε_s is assumed when $i = 1$. With level $\alpha \in (0, 1)$, and z_α such that the probability that a standard normal variate exceeds z_α is α, 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 2 we mention test statistics that are slightly simpler but valid less generally.

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 $\exp\left(-|s|/\phi_2\right)$. In Tables I, II and IV, 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 "makeneighbours" 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 neighbors, 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 neighbors of the location i, $i \neq j$, and $w_{ij}^* = 0$ otherwise. The maximum eigenvalue of W_1 is 1. We have chosen $k = 5$; the same choice in Han and Lee (2013). These weights produce covariance matrices satisfying Assumptions 2, 3 and 8 of Appendix 1 for the models and parameter values chosen. We compare results for alternative parameter values and weight functions in Table III. These Monte Carlo experiments are based on 2000 simulations.

Table I 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 a Student's t with 5 degrees of freedom doesn't exist, contradicting Assumption 7 of Appendix 1. Amongst the three robust estimators, $\hat{N}_4$ is easiest to compute and behaves slightly better, possibly because it uses the most information.

$$\boxed{TABLE\ I\ ABOUT\ HERE}$$

Table II provides size and power comparisons of tests for 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 ε_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 $\phi_1 = 1 - \exp(0.5)$. These models are quite similar for these parameter values: $\|I_n - \phi W_1\|_r = 1 - \phi$ and $\|\exp(\phi W_1)\|_r = \exp(\phi)$, where $\|\cdot\|_r$ is the maximum sum row norm, and the the smaller is ϕ_1, the closer are the SAR(1) and SMA(1). See LeSage and Pace (2007, pp 193). The exponential isotropic model (EXP) given by (8) with $\phi_1 = 1$, is simulated with $\phi_2 = 1$ condi-

tional on the fixed location points, using the lower-upper triangular decomposition of the covariance matrix, as suggested by Davis (1987) and implemented with the MATLAB routine (http://www.mathworks.com/matlabcentral/fileexchange/27613-random-field-simulation). The normal approximation of the test statistic 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 mispecification is badly behaved and the parameter estimates often fall on boundaries in many experiments. This is the case for various ϕ_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 is not difficult to reject this specification in the direction of non-isotropic covariance specifications. 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. 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.

$$\boxed{TABLE\ II\ ABOUT\ HERE}$$

Table III demonstrates how power depends on the underlying processes. We focus on testing the MESS(1) specification in the direction of a SMA(1), which performs comparatively worse than the other tests in Table II. 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 neighbors. We also use symmetrized nearest neighbour weights based on J. LeSage's MATLAB routine "fsym_neighbors2" for different ϕ_1; it uses $W_1 = AC_kA$ with $C_k = \left[C_k^{(i,j)} \right]_{i,j=1}^n$, $C_k^{(i,j)} = \sum_{i=1}^k \rho^i S_{(i)}$, where ρ is a parameter and $S_{(i)}$, $i = 1, ..., k$, are k individual binary weight matrices with 0 and 1 indicating whether the observations are one of the $i's$ nearest neighbours, and $A = diag\left\{ \sum_{i=1}^n C_k^{(1,i)}, ..., \sum_{i=1}^n C_k^{(n,i)} \right\}$. The maximum eigenvalue of this W_1 is 1, and

20

the corresponding covariance matrices satisfy Assumptions 2, 3 and 8 of Appendix 1. We took $\rho = 0.8$ and $k = 5$. The symmetrized nearest neighbors are denoted as SNN and the asymmetric ones, used in Tables I and II, are denoted as ANN. Power very much depends on W_1 and ϕ_1.

$$\boxed{TABLE\ III\ ABOUT\ HERE}$$

Table IV provides size and power for tests comparing the same models as in Table I 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 β_{10} and β_{20}, but it seems to dissapear as sample size increases.

$$\boxed{TABLE\ IV\ ABOUT\ HERE}$$

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 $\hat{u}^* = B(\hat{\theta}_1)\hat{\varepsilon}^*$, $j = 1, ..., n$, 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. That is, 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 $c_\alpha^* = \inf\left\{c \in \mathbb{R}^+ : m^{-1}\sum_{l=1}^m 1_{\left\{\hat{\eta}^{*(lsize} \geq c\right\}} \leq \alpha\right\}$. Table IV 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 t with 5 degrees of freedoom. 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 $\sqrt{n}LR$.

$$\boxed{TABLE\ V\ ABOUT\ HERE}$$

5. FINAL COMMENTS

In line with Table IV 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 β_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 (25) 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 V 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.

APPENDIX 1 : Regularity Conditions and Discussion

The first five assumptions are imposed for consistency of our parameter estimates (Theorem 1 in Appendix 2).

Assumption 1: *Under H_1, for all sufficiently large n, the u_j have uniformly bounded fourth moment, and, denoting by κ_{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 is condition of weak dependence with respect to fourth moments, which would hold 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 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

$$\|AB\|_2 \leq \|A\|_2 \|B\|. \tag{30}$$

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

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

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

$$\overline{\lim_{n\to\infty}} \sup_{\theta_i : \|\theta_i - \theta_i^\dagger\| < \varepsilon;\ \theta_i \in R_i} \left\| \Omega_i(\theta_i) - \Omega_i\left(\theta_i^\dagger\right) \right\| < \eta. \tag{31}$$

Notice that Assumptions 2 and 3 imply that (31) 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 Ω_i. To place the assumptions in perspective, for equally-spaced time series, when H_i implies stationarity Ω_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

23

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 θ_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 previous section) and the regime generating the observation points are generally entwined in a complicated way. 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 θ_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 Ω_i, whereas for some other models (e.g. (7) and (8)), though of course Ω_i admits a factorisation for any θ_i, the factors need not have a simple closed form representation as functions of θ_i.

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

$$\tilde{Q}_1 - \tilde{Q}_{10} = \frac{1}{2} \log \left\{ \frac{1}{n} tr \left(\Omega_1^{-1} \Omega_{10} \right) / \left| \Omega_1^{-1} \Omega_{10} \right|^{1/n} \right\}, \tag{32}$$

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

Assumption 4: $\theta_{10} \in R_1$ and for all $\theta_1 \in R_1 \backslash \theta_{10}$,

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

where the limit is assumed to exist.

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

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

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

$$\sigma_{2*}^2 = \tilde{\sigma}_2^2(\theta_{2*}) = \sigma_{10}^2 n^{-1} tr\left(\Omega_{2*}^{-1}\Omega_{10}\right). \tag{33}$$

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

$$\tilde{Q}_2 - \tilde{Q}_{2*} = \frac{1}{2}\log\left\{\frac{tr\left(\Omega_2^{-1}\Omega_{10}\right)}{tr\left(\Omega_{2*}^{-1}\Omega_{10}\right)}\left|\Omega_{2*}^{-1}\Omega_2\right|^{1/n}\right\}. \tag{34}$$

where $\Omega_{2*} = \Omega_2(\theta_{2*})$. Because θ_{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*

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

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

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

Assumption 7: *The representation (25) holds, where ε_s is a sequence of iid random variables with zero mean, variance σ_{10}^2, and finite eighth moment, b_{js} can depend on n, $b_{js} = b_{jsn}$, and, defining*

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

we have

$$\varlimsup_{n\to\infty}\sup_{1\le j\le n}\sum_{s=1}^{\infty}|c_{js}| + \varlimsup_{n\to\infty}\sup_{s\ge 1}\sum_{j=1}^{n}|c_{js}| < \infty. \tag{35}$$

The representation (25) was previously used in a spatial context by Robinson (2011), where its relevance is discussed. It implies that

$$\omega_{1jk0} = \sum_{s=1}^{\infty} b_{js} b_{ks}, \ \ j, k = 1, 2, ...n, \tag{36}$$

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

$$\sum_{s=1}^{\infty} c_{js}^2 = 1, \ \ j = 1, 2, ...n. \tag{37}$$

The normalized c_{js} can be compared with moving average weights in the stationary time series setting where $c_{js} = c_{j-s}$, when (35) reduces to a standard weak dependence summability condition; the eighth moment conditiom automatically holds under Gaussianity and is needed only to check a Lyapounov condition, otherwise finite fourth moments suffice. Note that in models of the form (9) we can choose b_{js} to be the $(j, s) th$ element of $S\left(\theta_{10}\right)^{-1}$, $s = 1, 2, ...n$, 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, implying indeed that such a representation holds quite generally in Gaussian settings, including with irregularly$-$spaced observations, as noted by 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).

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_∞ or maximum element norm $\|A\|_e = \max_{i,j} |a_{ij}|$, for a matrix $A = \left(a_{ij}\right)$. 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 (22), introduce:

Assumption 8: *For $i = 1, 2$ and all sufficiently large n, on an arbitrarily small*

neighbourhood $\mathcal{N}_i$ of θ_{i*}, the elements of Ω_i are thrice boundedly differentiable,

$$\overline{\lim_{n\to\infty}}\sup_{\theta_i\in\mathcal{N}_i}\left(\left\|\Omega_i^{-1}\right\|_r + \left\|\Omega_{ij}\right\|_r + \left\|\Omega_{ijk}\right\|_r + \left\|\Omega_{ijkl}\right\|_r\right) < \infty, \tag{38}$$

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

$$h \leq C \tag{39}$$

or

$$h^{-1} + h/n \to 0 \ as \ n \to \infty, \tag{40}$$

we have

$$\overline{\lim_{n\to\infty}}\sup_{\theta_i\in\mathcal{N}_i} h\left(\left\|\Omega_{ij}\right\|_e + \left\|\Omega_{ijk}\right\|_e + \left\|\Omega_{ijkl}\right\|_e\right) < \infty, \tag{41}$$

In spatial statistics models such as (4), (7) and 98), the h bounded case (39) is appropriate, when (41) is implied by (38). The allowance for (slower-than-n) divergent h (40) 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), $\Omega_i = (I_n - \theta_{i1}W)(I_n - \theta_{i1}W)'$, where it is sometimes assumed that $h\left\|W\right\|_e + \left\|W\right\|_r + \left\|W'\right\|_r \leq C$. Thus $\Omega_{i1} = 2\theta_{i1}WW' - W - W'$ satisfies (38) and (41), and also $\left\|\Omega_{ij}\right\|_2^2 \leq n\left\|\Omega_{ij}\right\|_e\left\|\Omega_{ij}\right\|_r$ implies that $\sup_{\theta_i\in\mathcal{R}_i}\left\|\Omega_{i1}\right\|_2 = O\left((n/h)^{1/2}\right)$. 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

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

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

27

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

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

Assumption 9: *The matrices*

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

exist and are positive definite.

APPENDIX 2: Theorems and Proofs

Theorem 1: *Under Assumptions 1-5 and H_1, as $n \to \infty$*

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

Proof of Theorem 1: Write $\tilde{Q}_{1*} = Q_{10}$ 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 $\bar{\mathcal{N}}_{i\varepsilon} = R_i \backslash \mathcal{N}_{i\varepsilon}$. Denoting P_1 probability under H_1

$$\begin{aligned} P_1\left(\widehat{\theta}_i \in \bar{\mathcal{N}}_{i\varepsilon}\right) &\leq P_1\left(\inf_{\bar{\mathcal{N}}_{i\varepsilon}} Q_i \leq Q_{i*}\right) \\ &\leq P_1\left(\sup_{R_i} \left|Q_i - \tilde{Q}_i\right| \geq \inf_{\bar{\mathcal{N}}_{i\varepsilon}}\left(\tilde{Q}_i - \tilde{Q}_{i*}\right)\right). \end{aligned}$$

The result follows if

$$\inf_{\bar{\mathcal{N}}_{i\varepsilon}}\left(\tilde{Q}_i - \tilde{Q}_{i*}\right) > \eta, \ \text{all sufficiently large } n \text{ and any } \eta > 0, \tag{42}$$

and if

$$\sup_{R_i}\left|Q_i - \tilde{Q}_i\right| \to_p 0, \ \text{as } n \to \infty. \tag{43}$$

The left side of (43) is bounded by

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

By the inequality (30),

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

so by Assumption 2,

$$\lim_{n\to\infty}\inf_{R_i}\tilde{\sigma}_i^2 > 0. \tag{44}$$

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

$$\begin{aligned}
&\frac{2}{n^2}\left\|\Omega_i^{-1}\Omega_{10}\right\|_2^2 + \frac{1}{n^2}\sum_{j,k,l,m=1}^{n}\omega_i^{jk}\omega_i^{lm}\kappa_{1jklm} \\
&\le \frac{2}{n^2}\left\|\Omega_i^{-1}\right\|_2^2\|\Omega_{10}\|^2 + \frac{1}{n^2}\left(\left\|\Omega_i^{-1}\right\|_2^4\sum_{j,k,l,m=1}^{n}\kappa_{1jklm}^2\right)^{1/2} \\
&\le \frac{2}{n}\left(\|\Omega_{10}\|^2 + \left(\frac{1}{n^2}\sum_{j,k,l,m=1}^{n}\kappa_{1jklm}^2\right)^{1/2}\right) \to 0 \text{ as } n \to \infty
\end{aligned}$$

by Assumptions 1 and 2, 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 $\mathcal{N}_{i\dagger\eta} = \left\{\theta_i : \left\|\theta_i - \theta_i^\dagger\right\| < \varepsilon\right\}$

$$\begin{aligned}
&E_1\sup_{\theta_i:\left\|\theta_i-\theta_i^\dagger\right\|<\varepsilon;\theta_i\in R_i}\left|tr\left(\left(\Omega_i\left(\theta_i\right)^{-1} - \Omega_i\left(\theta_i^\dagger\right)^{-1}\right)\left(uu' - \Omega_{10}\right)\right)\right| \\
&\le \left(E_1\|u\|^2 + tr\left(\Omega_{10}\right)\right)\sup_{\theta_i:\left\|\theta_i-\theta_i^\dagger\right\|<\varepsilon}\left\|\Omega_i\left(\theta_i\right)^{-1} - \Omega_i\left(\theta_i^\dagger\right)^{-1}\right\| = O(\eta n),
\end{aligned}$$

by (30) and Assumptions 2 and 3. This proves (43). Next, for $i = 2$, (42) is Assumption 5 in view of (34). For $i = 1$, by compactness R_i has a finite subcover and

29

fixing $\theta_1^\dagger \in R_1 \backslash \theta_{10}$, and for any $\varepsilon > 0$

$$\inf_{\theta_1: \|\theta_1 - \theta_1^\dagger\| < \varepsilon; \theta_1 \in R_1} \left(\tilde{Q}_1 - \tilde{Q}_{1*} \right) \geq \left(\tilde{Q}_{1\dagger} - \tilde{Q}_{1*} \right) - \sup_{\theta_1: \|\theta_1 - \theta_1^\dagger\| < \varepsilon; \theta_1 \in R_1} \left| \tilde{Q}_1 - \tilde{Q}_{1\dagger} \right|, \quad (45)$$

where

$$
\begin{aligned}
\tilde{Q}_1 - \tilde{Q}_{1\dagger} &= \frac{1}{2} \log \left\{ tr \left(\Omega_1^{-1} \Omega_{10} \right) / \left| \Omega_1^{-1} \right|^{1/n} \right\} - \frac{1}{2} \log \left\{ tr \left(\Omega_{1\dagger}^{-1} \Omega_{10} \right) / \left| \Omega_{1\dagger}^{-1} \right|^{1/n} \right\} \\
&= \frac{1}{2} \log \left\{ tr \left(\Omega_1^{-1} \Omega_{10} \right) / tr \left(\Omega_{1\dagger}^{-1} \Omega_{10} \right) \right\} + \frac{1}{2n} \log \left| \Omega_1^{-1} \Omega_{1\dagger} \right| \\
&= \frac{1}{2} \log \left(1 + \frac{tr \left(\left(\Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right) \Omega_{10} \right)}{tr \left(\Omega_{1\dagger}^{-1} \Omega_{10} \right)} \right) + \frac{1}{2n} \log \left| I + \left(\Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right) \Omega_{1\dagger} \right|.
\end{aligned}
$$

Denoting by λ_j and ν_j the jth 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

$$
\begin{aligned}
\frac{1}{2} \left| \sum_{j=1}^n \lambda_j \right| / tr \left(\Omega_{1\dagger}^{-1} \Omega_{10} \right) + \frac{1}{2n} \sum_{j=1}^n |\nu_j| &\leq C n^{-1/2} \left\{ \left(\sum_{j=1}^n \lambda_j^2 \right)^{1/2} + \left(\sum_{j=1}^n \nu_j^2 \right)^{1/2} \right\} \\
&\leq C n^{-1/2} \left\| \Omega_1^{-1} - \Omega_{1\dagger}^{-1} \right\|_2 \\
&\leq C \left\| \Omega_1 - \Omega_{1\dagger} \right\|,
\end{aligned}
$$

where C denotes a positive generic constant and we use Assumption 2 and (30). By Assumption 3, for any $\eta > 0$ we can choose ε such that for all sufficiently large n the last displayed expression is bounded by $C\eta$, uniformly on $\left\| \theta_1 - \theta_1^\dagger \right\| < \varepsilon$, as therefore is $\left| \tilde{Q}_1 - \tilde{Q}_{1\dagger} \right|$. In view of (45) the proof of (42) for $i = 1$ is completed by noting that (32) 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 $n \to \infty$,*

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

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

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

30

For $j = 1, ..., p_i$,

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

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

$$n^{1/2} D d_* \to_d \mathcal{N}(0, \Psi), \quad \text{as } n \to \infty. \tag{46}$$

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

$$d_* = \sum_{s=1}^{\infty} a_{ss*} \left(\varepsilon_s^2 - \sigma_{10}^2 \right) + 2 \sum_{s=1}^{\infty} 1(s \geq 2) \sum_{t=1}^{s-1} a_{st*} \varepsilon_s \varepsilon_t = \sum_{s=1}^{\infty} v_s, \tag{47}$$

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

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

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

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

On proving that, for some J sequence,

$$n^{1/2} D d_{*b} \to_p 0, \tag{48}$$

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

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

Introduce a square matrix Z such that $T = ZZ'$. For large enough J, T is positive definite under our conditions (see (50)). For a vector ζ such that $\|\zeta\| = 1$, write

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

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

$$r_* \to_d \mathcal{N}(0,1), \text{ as } n \to \infty, \tag{49}$$

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

$$T \to N_* \text{ as } J \to \infty, \tag{50}$$

the proof of (46) is completed; we omit proof of (50) as it is straightforward given our other proofs.

The details for checking (48) and (49) differ considerably from those of Robinson (2011), mainly because our v_k is quadratic in the ε_i. First, (48) follows on showing that as $J \to \infty$,

$$E \left\| n^{1/2} D d_{*b} \right\|^2 \to 0. \tag{51}$$

From Assumption 7 the v_s are uncorrelated and the left side of (51) is bounded by

$$Cn \sum_{s=J+1}^{\infty} E \left\| Dv_s \right\|^2, \tag{52}$$

where, from (47)

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

The $(p_1 + p_2 + 2) \times 1$ vector Da_{st*} has mth element of form $b_s' R_m b_t / n$, where $b_s = b_s(\theta_{10})$. Now (52) is bounded by

$$
\begin{aligned}
Cn \sum_{s=J+1}^{\infty} \sum_{t=1}^{s} \left\| Da_{st*} \right\|^2 &\leq \frac{C'^{p_1+p_2+2}}{n} \sum_{m=1}^{\infty} \sum_{s=J+1}^{\infty} b_s' R_m \sum_{t=1}^{s} b_t b_t' R_m' b_s \\
&\leq \frac{C'^{p_1+p_2+2}}{n} \sum_{m=1}^{\infty} \left(\sum_{s=J+1}^{\infty} b_s' R_m R_m' b_s \right), \tag{53}
\end{aligned}
$$

since (25) implies $\left\| \sum_{t=1}^{s} b_t b_t' \right\| \leq \left\| \Omega_1 \right\| \leq C$ 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

33

$p_1 + 2 \leq m \leq p_1 + p_2 + 1,$

$$|r_{mjk}| \leq \frac{C}{h^{1/2}}, \quad \sum_k |r_{mjk}| \leq Ch^{1/2},$$

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

$$|r_{mjk}| + \sum_k |r_{mjk}| \leq C.$$

The bracketed term in (53) is

$$
\begin{aligned}
\sum_{s=J+1}^{\infty}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} b_{js} r_{mjk} r_{mlk} b_{ls} & \leq C \sum_{s=J+1}^{\infty}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} |c_{js}|\,|r_{mjk}|\,|r_{mlk}|\,|c_{ls}| \\
& \leq C \sum_{s=J+1}^{\infty}\sum_{j=1}^{n}\sum_{l=1}^{n}\left(\max_k |r_{mjk}|\right)|c_{js}|\left(\sum_{k=1}^{n}|r_{mlk}|\right)|c_{ls}| \\
& \leq C\sum_{j=1}^{n}\sum_{s=J+1}^{\infty}|c_{js}|\sum_{l=1}^{n}|c_{ls}| \\
& \leq Cn\max_j \sum_{s=J+1}^{\infty}|c_{js}| \tag{54}
\end{aligned}
$$

from Assumption 7. Also that assumption implies that, for $j = 1, ..., n$, 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 $J = J_n = \max(J_{1n}, ..., J_{nn})$, (54)$\leq Cn\eta_n = o(n)$ as $n \to \infty$. This completes the proof of (51).

The proof of (49) follows (see e.g. Scott (1973)) on checking a Lyapounov type condition

$$\sum_{s=1}^{J} E \left\| n^{1/2} Z^{-1} D v_s \right\|^4 \to 0 \tag{55}$$

and

$$n\sum_{s=1}^{J} \left(E\left(D v_s v_s' D \,|\varepsilon_t, t \leq s-1\right) - E\left(D v_s v_s' D\right)\right) \to_p 0. \tag{56}$$

To check (55) note first that

$$
\begin{aligned}
E\left\| Z^{-1} D v_s \right\|^4 & \leq C\left\| D a_{ss*}\right\|^4 + CE\left\| \sum_{t=1}^{s-1} D a_{st*}\varepsilon_t \right\|^4 \\
& \leq C\sum_{t=1}^{s}\left\| D a_{st*}\right\|^4 + C\left(\sum_{t=1}^{s-1}\left\| D a_{st*}\right\|^2\right)^2 \\
& \leq C\left(\sum_{t=1}^{s}\left\| D a_{st*}\right\|^2\right)^2.
\end{aligned}
$$

Now

$$\sum_{t=1}^{s} \|Da_{st*}\|^2 \leq \frac{C}{n^2}^{p_1+p_2+2} \sum_{m=1}^{s} \left(b'_s R_m \sum_{t=1}^{s} b_t b_t R'_m b_s \right)$$

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

Thus the left side of (55) is bounded by

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

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

on applying both parts of (35) of Assumption 7, to prove (55).

To prove (56), note first that $E\left(v_s v'_s | \varepsilon_t, t \leq s-1 \right)$ is

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

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

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

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

Thus the Euclidean norm of the left side of (56) is bounded by

$$n \left| E\left(\varepsilon_s^3 \right) \right| \left\| \sum_{s=2}^{J} \sum_{t=1}^{s-1} \left(A_{st} + A'_{st} \right) \varepsilon_t \right\|_2 \tag{57}$$

$$+n\sigma_{10}^2 \left\| \sum_{s=2}^{J} \left\{ \left(\sum_{t=1}^{s-1} a_{st*} \varepsilon_t \right) \left(\sum_{t=1}^{s-1} a_{st*} \varepsilon_t \right)' - \sigma_{10}^2 \sum_{t=1}^{s-1} a_{st*} a'_{st*} \right\} \right\|_2 , \tag{58}$$

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 (57) has expectation bounded by

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

35

where

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

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

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

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

$$\sum_{r=t+1}^{J} \sum_{l=1}^{n} \sum_{m=1}^{n} |c_{lt}| |r_{klm}| |c_{mr}| \leq C \sum_{l=1}^{n} \sum_{m=1}^{n} |c_{lt}| |r_{klm}| \leq C h^{1/2} \sum_{l=1}^{n} |c_{lt}|,$$

while on the other,

$$\sum_{r=t+1}^{J} \sum_{l=1}^{n} \sum_{m=1}^{n} |c_{lt}| |r_{klm}| |c_{mr}| \leq C \sum_{l=1}^{n} |c_{lt}| \sum_{m=1}^{n} |r_{klm}| \leq C h^{-1/2} \sum_{l=1}^{n} |c_{lt}| \leq C h^{-1/2}.$$

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 $(57) = O_p\left(n^{-1/2}\right)$.

To deal with (58), note that

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

$$= \sum_{t=1}^{J-1} \sum_{s=t+1}^{J} a_{st*} a_{st*}' \left(\varepsilon_t^2 - \sigma_{10}^2\right) + . \sum_{t=1}^{J-1} \sum_{u=1}^{t-1} \left\{ \sum_{s=t+1}^{J} \left(a_{st*} a_{su*}' + a_{su*} a_{st*}'\right) \right\} \varepsilon_t \varepsilon_u. \quad (59)$$

Now

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

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

$$\leq \frac{C}{n^2} \cdot \sum_{t=1}^{J-1} \sum_{u=1}^{t-1} \sum_{s=t+1}^{J} \sum_{r=t+1}^{J} b'_s R_j b_t b'_s R_k b_u b'_r R_j b_t b'_r R_k b_u$$

$$\leq \frac{C}{n^2} \cdot \sum_{t=1}^{J} \sum_{s=1}^{J} \sum_{r=1}^{J} \sum_{u=1}^{J} \sum_j \sum_k \sum_l \sum_m |c_{ls}| |r_{jlm}| |c_{mt}| \sum_l \sum_m |c_{ls}| |r_{klm}| |c_{mu}|$$

$$\times \sum_l \sum_m |c_{lr}| |r_{jlm}| |c_{mt}| \sum_l \sum_m |c_{lr}| |r_{jlm}| |c_{mu}|$$

$$\leq \frac{C}{n^2} \cdot \sum_{t=1}^{J} \sum_{s=1}^{J} \sum_{r=1}^{J} \sum_j \sum_k \left(\sum_l \sum_m |c_{ls}| |r_{jlm}| |c_{mt}| \right) \left(\sum_l \sum_m |c_{ls}| |r_{klm}| \left| \sum_{u=1}^{J} c_{mu} \right| \right)$$

$$\times \left(\sum_l \sum_m |c_{lr}| |r_{jlm}| |c_{mt}| \right) \left(\sum_l \sum_m |c_{lr}| |c_{mu}| h^{-1/2} \right)$$

$$\leq \frac{Ch^{-1/2}}{n^2} \cdot \sum_{t=1}^{J} \sum_{s=1}^{J} \sum_j \left(\sum_l \sum_m |c_{ls}| |r_{jlm}| |c_{mt}| \right) \sum_l |c_{ls}| h^{1/2} \left(\sum_l \sum_m \sum_{r=1}^{J} |c_{lr}| |r_{jlm}| |c_{mt}| \right)$$

$$\leq \frac{C}{n^2} \cdot \sum_j \sum_l \sum_m \left(\sum_{s=1}^{J} |c_{ls}| \right) |r_{jlm}| \left(\sum_{t=1}^{J} |c_{mt}| \right) \sum_m \left(\sum_l |r_{jml}| \right) |c_{mt}|$$

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

for $\leq m \leq p_1$ 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 (59),

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

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

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

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

37

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 (46)

Next consider

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

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

so

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

$$+ \frac{1}{2n} tr \left\{ \Omega_i^{-1} \Omega_{ijk} - \Omega_i^{-1} \Omega_{ij} \Omega_i^{-1} \Omega_{ik} \right\},$$

$$j, k = 1, .., p_i,$$

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

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

It is then readily seen that

$$E_1 \frac{\partial^2 L_{10}}{\partial \tau_1 \partial \tau_1'} = M_{10}, \quad E_1 \frac{\partial^2 L_{2*}}{\partial \tau_2 \partial \tau_2'} = M_{2*}. \tag{60}$$

Now denote

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

We have

$$n^{1/2} N_*^{-1/2} M_* \left(\widehat{\tau} - \tau_* \right) = n^{1/2} N_*^{-1/2} D^{-1} \left(DM_* D \right) D^{-1} \left(\widehat{\tau} - \tau_* \right) \to_d \mathcal{N} \left(0, I_{p+2} \right),$$

38

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

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

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 (60). Because of (46) the result readily follows.

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

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

Theorem 3: *Under Assumptions 1-9 and H_1, and with the ε_s asssumed Gaussian for $i = 1$, as $n \to \infty$,*

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

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

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 ε_s asssumed Gaussian for $i = 1$, as $n \to \infty$,*

$$\frac{LR}{\left(n^{-1}\widehat{e}'\,\widehat{M}^{-1}\widehat{N}_i\widehat{M}^{-1}\widehat{e}\right)^{1/2}} \to_d \mathcal{N}(0,1)\,,\ \ i = 1, 2, 3, 4. \tag{61}$$

Proof of Theorem 4: Writing, as in (19), $\widehat{\sigma}_2^2 = \overline{\sigma}_2^2\left(\widehat{\theta}_i\right)$, $i = 1, 2$, we have

$$
\begin{aligned}
LR &= \log\frac{\widehat{\sigma}_2^2}{\sigma_{2*}^2} - \log\frac{\overline{\sigma}_2^2(\widehat{\theta}_1,\widehat{\theta}_2)}{\sigma_{2*}^2} \\[2mm]
&= \log\left(1 + \frac{\widehat{\sigma}_2^2 - \sigma_{2*}^2}{\sigma_{2*}^2}\right) - \log\left(1 + \frac{\overline{\sigma}_2^2(\widehat{\theta}_1,\widehat{\theta}_2) - \sigma_{2*}^2}{\sigma_{2*}^2}\right) \\[2mm]
&= \frac{\widehat{\sigma}_2^2 - \sigma_{2*}^2}{\sigma_{2*}^2} + \frac{\overline{\sigma}_2^2(\widehat{\theta}_1,\widehat{\theta}_2) - \sigma_{2*}^2}{\sigma_{2*}^2} \\[2mm]
&\quad + O_p\left(\left(\widehat{\sigma}_2^2 - \sigma_{2*}^2\right)^2 + \left(\overline{\sigma}_2^2(\widehat{\theta}_1,\widehat{\theta}_2) - \sigma_{2*}^2\right)^2\right). \tag{62}
\end{aligned}
$$

From calculations below and since $\underline{\lim}_{n\to\infty}\sigma_{2*}^2 > 0$, the remainder term in (62) can be neglected. Now

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

which may be written

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

where, by the mean value theorem,

$$u\left(\widehat{\theta}_1,\widehat{\theta}_2\right) - u\left(\theta_{10},\theta_{2*}\right) = \sum_{j=1}^{p_1} c_j\left(\overline{\theta}\right)\left(\widehat{\theta}_{1j} - \theta_{1j0}\right) + \sum_{j=1}^{p_2} d_j\left(\overline{\theta}\right)\left(\widehat{\theta}_{2j} - \theta_{2j*}\right),$$

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

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

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

$$
\begin{aligned}
n^{1/2} e'_* (\widehat{\tau} - \tau_*) &= e'_* M_*^{-1} N_*^{1/2} n^{1/2} N_*^{-1/2} M_* (\widehat{\tau} - \tau_*) \\
&= e'_* D (D M_* D)^{-1} D N_*^{1/2} n^{1/2} N_*^{-1/2} M_* (\widehat{\tau} - \tau_*) \\
&\to_d \mathcal{N} \left(0, \zeta' \Phi^{-1} \Psi \, \Phi^{-1} \zeta \right),
\end{aligned}
$$

where

$$
\zeta = \lim_{n \to \infty} D e_*
$$

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

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

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 normalisations in Assumption 9 and indicating that when h diverges the $(j, p_1 + 1)th$, $j = 1, ..., p_1$, and $(j + p_1 + 1, p + 2)th, j = 1, ..., p_2$, elements of Φ and Ψ 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 normalisations 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

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

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

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

41

when $h \to \infty$. However, the statistic in (61) 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. Journal of Regional Science 26, 1986.

Anselin, L.,1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, The Netherlands.

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

Burridge, P., 2012. Improving the J test in the SARAR model by likelihood-based estimation. Spatial Economic Analysis 7, 75-107

Burridge, P., Fingleton, B., 2010. Bootstrap inference in spatial econometrics: the J test. Spatial Economic Analysis 5, 93-119.

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

Cox. D.R., 1961. Tests of separate families of hypotheses. 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. Journal of the Royal Statistical Society B 24, 406-424.

Davis, M.W., 1987. Production of conditional simulation via the LU triangular decomposition of the covariance matrix. Mathematical Geology 2, 91-98.

De Oliveira, V., Kedem, B., Short, D.A., 1997. Bayesian prediction of transformed Gaussian fields. Journal of the American Statistical Association 92, 1422-1433.

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

Dunsmuir, W., 1983. A central limit theorem for estimation in Gaussian stationary time series observed at unequally spaced times. Stochastic Processes and their

Applications 14, 279-295.

Fuentes, M., 2007. Approximate likelihood for large irregularly spaced spatial data. Journal of the American Statistical Association 102, 321-331.

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

Han, X., Lee, L.-F.,, 2013. Model selection using J-test for the spatial autoregressive model. Regional Science and Urban Economics, forthcoming.

Haining, R. P., 1978. The moving average model of spatial interaction. Transactions of the Institute of British Geographers 3, 202-225.

Handcock, M.S., Wallis, J.R., 1994. An approach to statistical spatial-temporal modeling of meteorological fiellds (with discussion). Journal of the American Statistical Association 89, 368-390.

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

Jin, F., Lee, L.-F. 2013. Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. Regional Science and Urban Economics, forthcoming.

Jones, R.H., Vecchia, A.V., 1993. Fitting continuous ARMA models to unequally spaced spatial data. Journal of the American Statistical Association 88, 947-954.

Kelejian, H. H., 2008. A spatial J-test for model specification against a single or a set of non-nested alternatives. Letters in Spatial and Resource Sciences 1, 3-11.

Kelejian, H. H., Piras, G., 2011. An extension of Kelejian's J-test for non-nested spatial models. Regional Science and Urban Economics 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., 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. Journal of the Royal Statistical Society B 71, 191-217.

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

Robinson, P.M., 1977. Estimation of a time series model from unequally spaced data. Stochastic Processes and Their Applications 6, 9-24.

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

Robinson, P.M., 2011. Asymptotic theory for nonparametric regression with spatial data. Journal of Econometrics 165, 5-19.

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

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. Journal of the Royal Statistical Society B 66, 275-296.

Vecchia, A.V., 1988. Estimation and model identification for continuous spatial processes. Journal of the Royal Statistical Society B 50, 297-312.

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

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

TABLE I											
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			$SMA/MESS$		
$\hat{N}_i \backslash n$	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
$\varepsilon_j \sim (\Gamma(2,1) - 2)/\sqrt{2},\ Kurtosis\ Excess = 3$											
H_1/H_2	SAR/SMA			$SAR/MESS$			SMA/SAR			$SMA/MESS$		
$\hat{N}_i \backslash n$	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
$\varepsilon_j \sim t_5/\sqrt{5/3},\ Kurtosis\ Excess = 6$											
H_1/H_2	SAR/SMA			$SAR/MESS$			SMA/SAR			$SMA/MESS$		
$\hat{N}_i \backslash n$	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 II

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

| $H_2 \backslash H_1$ | 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

TABLE III

Power: % rejections under H_2

$H_1 : MESS$ vs $H_2 : SMA$

Tests using Gaussian u_j and $\hat{N}_4$

| | $NN,\ k = 5$ | | | | | ANN | |
| | ANN | | SNN | | | $\phi_1 = 0.5$ | |
$\phi_1 \backslash n$	500	1000	500	1000	$k \backslash n$	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

		Size				Power			
TABLE IV									
Size and power using Gaussian u_j and $\hat{N}_4$									
Tests based on residuals of simple linear regression									
$H_2 \backslash H_1$	n	Size % Rejections under H_1				Power % Rejections under H_2			
		SAR	SMA	$MESS$	EXP	SAR	SMA	$MESS$	EXP
SAR	100		1.80	1.35	1.80		6.60	4.20	80.00
	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
SMA	100	1.15		0.90	1.20	3.35		1.60	35.25
	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
$MESS$	100	3.45	2.50		1.40	5.70	4.15		8.95
	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 V

$H_1 : SAR$ 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		
$n \backslash 100\alpha\%$	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