Correlation Testing in Time Series, Spatial and Cross-Sectional Data

P.M. Robinson*
The London School of Economics

March 28, 2008

Abstract

We provide a general class of tests for correlation in time series, spatial, spatio-temporal and cross-sectional data. We motivate our focus by reviewing how computational and theoretical difficulties of point estimation mount as one moves from regularly-spaced time series data, through forms of irregular spacing, and to spatial data of various kinds. A broad class of computationally simple tests is justified. These specialize to Lagrange multiplier tests against parametric departures of various kinds. Their forms are illustrated in case of several models for describing correlation in various kinds of data. The initial focus assumes homoscedasticity, but we also robustify the tests to nonparametric heteroscedasticity.

JEL Classifications: C21; C22; C29
Keywords: Correlation; heteroscedasticity; Lagrange multiplier tests.

---

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

1 INTRODUCTION

Irregularly-spaced time series, spatial, and spatio-temporal data, and the possibility of cross-sectional correlation, pose considerable difficulties, with respect to modelling, computations and statistical theory. In general, the possibility has to be recognized that there is correlation across time, or space, or other relevant dimensions. Rules of inference based on the incorrect assumption of independence will generally be invalidated. Unfortunately, even developing models for dependence can be a far more complicated business than in a regularly-spaced time series. Computations can also be more onerous. The development of a satisfactory, useful, asymptotic theory for estimates of both parameters describing the dependence, and parameters of economic interest, such as describing regression effects, can be infeasible. The difficulties arise essentially because of the non-Toeplitz covariance matrix structure that emerges, and the difficulty of separating the regime generating the "location" of observations from that generating the observations themselves, when formulating regularity conditions. Here location can refer to some relevant economic space, not just time or geographical space.

Immense simplification to rules of inference and computations result if there can be assumed to be no dependence. It has been argued (see e.g. Cressie, 1993) that much spatial data can be satisfactorily modelled in terms of the conditional mean, leaving little to be accounted for by disturbance correlation. Likewise, the common assumption of cross-sectional independence may often be reasonable. This favourable circumstance cannot be taken for granted, but it does further motivate carrying out in the first place tests for independence. If the evidence for independence is strong then we may proceed with simple rules of inference on the remaining parameters of interest. If not, we have to look at developing rules that efficiently take account of dependence, or that are robust to dependence. But these tasks are difficult to develop in a very general context. In this paper we focus on testing for independence in such a general context.

This topic has been addressed in a vast time series literature, however little of this permits irregular spacing. It has also been a major, long-standing theme of the spatial literature, with numerous contributions following Moran (1950), Cliff and Ord (1968, 1972), but settings have been fairly specific. It seems useful to discuss a general approach which can be applied in a variety of circumstances, under regularity conditions which may shed light on the suitability of the asymptotic theory in specific situations. In a linear regression setting, a general class of statistics is developed that has a chi-square limit distribution under the null hypothesis of independence of disturbances. Special cases can be interpreted as Lagrange multiplier (LM) statistics directed against specified alternatives where they should have good power, though they may have little power against others. It is thus envisaged that in practice several tests may be employed, based on variety of working parametric models.

The tests are developed in Section 3, along with relevant asymptotic theory, of which proofs are left to appendices. In Section 4 they are discussed in some LM examples. First, however, we provide in the following section further

background and motivation by reviewing how difficulties develop as one moves from equally-spaced time series to irregularly-spaced ones, and to spatial and cross-sectionally-correlated data.

2   IMPLICATIONS OF IRREGULAR SPACING AND SPATIAL DATA

To fix ideas, and avoid distracting complications, we focus entirely on a linear regression setting, where the regression function is correctly specified, and the covariance matrix is parametric. We will also describe our tests for independence in this setting.

2.1   Regression model and Gaussian estimation

We consider the $n \times 1$ vector y_n of scalar observations y_{in}, $i = 1, ..., n$,

$$y_n = (y_{1n}, ..., y_{nn})', \tag{2.1}$$

the prime denoting transposition. The ordering of the y_{in} is arbitrary, though for time series data it would normally be chronological. The triangular-array aspect of the y_{in} allows for such asymptotic regimes as spatial autoregressive (AR) models with row-normalized weight matrices. We suppose that for a given sequence of $n \times q$ matrices X_n, $0 \le q < n$, of rank q, and a $q \times 1$ unknown vector β_0,

$$y_n = X_n \beta_0 + u_n, \tag{2.2}$$

for all sufficiently large n, where

$$u_n = (u_{1n}, ..., u_{nn})' \tag{2.3}$$

is an unobservable vector satisfying

$$E(u_n) = 0, \quad E(u_n u_n') = \sigma_0^2 \Omega_n(\theta_0), \tag{2.4}$$

where σ_0^2 is an unknown positive scalar and $\Omega_n(\theta)$ is a given positive definite $n \times n$ matrix function of a $p \times 1$ vector parameter θ, and with θ_0 being unknown. The case $q = 0$ means that X_n is absent from (2.2), so that $u_n = y_n$ is observable. Lack of correlation in the u_{in} occurs when $\Omega_n(\theta_0)$ is diagonal. This includes the possibility of heteroscedasticity across i, but our main focus is on the implications of non-diagonality.

Most interest may be in β_0, with θ_0 and σ_0^2 representing nuisance parameters, but in any case their estimation is linked. Conventionally, but conveniently, we consider estimates based on a Gaussian pseudo-likelihood. We have used words such as "independent" and "uncorrelated" rather interchangeably, without drawing a distinction. Of course they are identical if Gaussianity holds, but (2.4) only refers to first- and second-order properties. On the other hand,

stronger conditions than (2.4) would be needed in order to develop asymptotic statistical theory, and here the null hypothesis of no correlation will be supplemented by the assumption of independence.

The Gaussian pseudo-log-likelihood for y_n is given by

$$
\begin{aligned}
L_n\left(\beta, \sigma^2, \theta\right) &= -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{n}{2}\log \det \Omega_n(\theta) \\
&\quad - \frac{1}{2\sigma^2}\left(y_n - X_n\beta\right)' \Omega_n(\theta)^{-1}\left(y_n - X_n\beta\right),
\end{aligned}
\tag{2.5}
$$

β, σ^2 and θ denoting any admissible values. As is well known, for given θ $L_n\left(\beta, \sigma^2, \theta\right)$ is maximized with respect to β, σ^2 by

$$
\hat{\beta}_n(\theta) = \left(X_n'\Omega_n(\theta)^{-1}X_n\right)^{-1} X_n'\Omega_n(\theta)^{-1}y_n,
\tag{2.6}
$$

$$
\hat{\sigma}_n^2(\theta) = \frac{1}{n}\left(y_n - X_n\hat{\beta}_n(\theta)\right)' \Omega_n(\theta)^{-1}\left(y_n - X_n\hat{\beta}_n(\theta)\right),
\tag{2.7}
$$

Then

$$
\hat{\theta}_n = \arg\max_{\theta} L_n\left(\hat{\beta}_n(\theta), \hat{\sigma}_n^2(\theta), \theta\right),
\tag{2.8}
$$

the maximization conducted over a suitable compact subset of $\mathbb{R}^q$ that includes θ_0. Equivalently,

$$
\hat{\theta}_n = \arg\min_{\theta} Q_n(\theta),
\tag{2.9}
$$

where

$$
Q_n(\theta) = \log \hat{\sigma}_n^2(\theta) + \frac{1}{n}\log \det \Omega_n(\theta).
\tag{2.10}
$$

2.2 Regularly-spaced time series

For equally-spaced time series, where the $u_i = u_{in}$ are ordered chronologically and stationary, $Q_n(\theta)$ can be typically approximated by simpler quantities which, when minimized, produce estimates of θ_0 with the same limit distribution as $n^{\frac{1}{2}}\left(\hat{\theta}_n - \theta_0\right)$. There are two sources of this favourable outcome. One is that $\Omega_n(\theta)$ is a Toeplitz matrix, and can thus be approximately diagonalized by a unitary transformation, so that $\hat{\sigma}_n^2(\theta)$ can be approximated by an integral or sum, across frequency, of the ratio of the periodogram and the parameterized spectral density. Indeed, in many time series models, such as autoregressive moving average (ARMA) ones, the spectral density can be written down by inspection, whereas the elements of $\Omega_n(\theta)$ cannot, and can be cumbersome. The second simplification arises when the second term on the right of (2.10) is asymptotically negligible. This occurs in "standard parameterizations" of ARMA models, where the innovations variance is free of the parameters describing autocorrelation. In that case the problem (2.9) can be replaced by minimization of $\hat{\sigma}_n^2(\theta)$ or a proxy such as described above. This covers the nonlinear least squares procedures recommended by Box and Jenkins (1970) for ARMA models. The computational simplifications are also reflected in a relatively neat asymptotic statistical theory, exemplified by Hannan (1973), Fox and Taqqu (1986). The

estimates of θ_0 are root-n-consistent and asymptotically normally distributed under conditions that require a one-sided infinite moving average representation for the u_i with innovations that are not necessarily Gaussian or independent and identically distributed, but are homoscedastic martingale differences with moments of order only 2 required to be finite. Moreover, the covariance matrix in the limiting normal distribution is unaffected by non-Gaussianity of u_i.

2.3 Lattice data

Equally-spaced spatial or spatio-temporal data present additional problems. We consider only the case of "increasing-domain" asymptotics, as implicitly assumed in the preceding discussion. Observations are recorded on a rectangular lattice of dimension $d > 1$. Intervals between observations are constant within dimensions, but can vary across dimensions. Here n represents the total number of observations, i.e. $n = \Pi_{j=1}^d n_j$, and asymptotic theory would typically entail $n_j \to \infty$ for all j. Looking again at (2.10), when $u_i = u_{in}$ is stationary a generalization of the Toeplitz property described for the time series case means that again $\sigma_n^2(\theta)$ can be approximated by a weighted periodogram average. However, it is less likely that $\log \det \Omega_n(\theta)$ can be ignored. The problem was first demonstrated by Whittle (1954), occuring in particular when u_i depends on "leads" as well as "lags" in one or more dimensions, as seems plausible in a spatial context, by comparison with the unilateral modelling standard in time series analysis. Whittle (1954) also showed that, quite generally, multilateral models have a "half-plane" kind of unilateral moving average representation, extending the Wold representation of time series, whence the last term in (2.10) can be ignored. However, the half-plane representation typically involves functions of the coefficients in the original multilateral model that cannot be written in closed form. Nor can it necessarily be well approximated by a parsimonious half-plane model, and the curse of dimensionality is a serious potential problem in spatial modelling.

A further difficulty arising with lattice data with dimension $d > 1$ is the "edge effect". Estimates of θ_0 given by (2.9), and by the usual approximations to this, can be seen as functions of sample autocovariances. In the time series case $d = 1$, the lag$-j$ sample autocovariance is the sum of $n-j$ products divided by n. The consequent finite-sample bias causes no problem with asymptotic theory for $\hat{\theta}_n$. However, when $d > 1$ the bias is of greater order, and leads to an asymptotic theory that is not useful. In particular, for $d = 2$ the bias is of order at least $n^{-\frac{1}{2}}$ so that $n^{\frac{1}{2}} \left(\hat{\theta}_n - \theta_0 \right)$ does not converge to a zero-mean random variable. For $d > 3$ the order of the bias is even greater than $n^{-\frac{1}{2}}$. A solution proposed by Guyon (1982) essentially replaces the usual, biased, sample autocovariances by unbiased ones. However, Dahlhaus and Künsch (1987) noted that this sacrifices the desirable positive definite property of the Gaussian pseudo-likelihood, and can lead to possible numerical difficulties and a covariance matrix estimate that is not necessarily non-negative definite. They overcame this drawback by instead employing tapering, but thereby introducing ambiguity due to the choice of

5

taper, and due to an additional tapering parameter if asymptotic efficiency is to be claimed. Robinson and Vidal Sanz (2006) proposed an alternative approach, justifying their estimates of a general class of models for any $d > 1$. However, they also introduced an element of arbitrariness in implementation in order to cope with the edge effect.

2.4   Irregularly-spaced time series

Irregular spacing of data can arise in several ways. Calendar monthly time series data, for example, are not exactly equally-spaced. However, there is evidence that the effects of disregarding this are unlikely to be significant, and in any case this kind of irregular spacing is largely ignored by practitioners. Another phenomenon is a once-and-for-all change in the sampling interval, as when quarterly observation changes to monthly (see, e.g. Sargan and Drettakis, 1974). For a given dynamic model for the monthly observations, a model for the "skip-sampled" quarterly ones can be deduced and the estimation problem addressed in terms of an objective function that combines components from the two regimes.

Observations can be missing from an otherwise regularly-spaced grid in other ways. Periodic sampling, as in case of weekday observations, disturbs the Toeplitz structure of $\Omega_n(\theta)$, but not in a way that severely complicates computation: one can work with a derived model for equally-spaced vector observations (e.g. the five weekday ones). Non-periodic missing can be ignored in case of only a few missing values, but generally $\Omega_n(\theta)$ allows no simplified approximation, and nor can the $\log \det \Omega_n(\theta)$ term in (2.10) be neglected. Nevertheless, for suitable models, the Kalman filter and EM algorithm can be applied to break up the computations into simple steps. However, whether one treats the regime generating the observation times as deterministic or stochastic, it seems difficult to deduce an asymptotic theory based on reasonably primitive and comprehensible conditions, in particular on ones that separate out the conditions on the process from those on the sampling regime. Dunsmuir (1983) developed a central limit theorem that is perhaps as successful as is possible in this respect, though it requires a condition on the information matrix that depends simultaneously on both features. Moreover he did not treat $\hat{\theta}_n$ itself, but rather a one-step Newton approximation commencing from an initial $n^{\frac{1}{2}}$-consistent estimate. This is in order to avoid a consistency proof, a usual preliminary to the central limit theorem for implicitly-defined extremum estimates. Dunsmuir (1983) described the consistency as an open problem. Dunsmuir and Robinson (1981) developed a full asymptotic theory for an alternative estimate employing an equally-spaced "amplitude-modulated" sequence, as introduced by Parzen (1963), but generally this estimate is asymptotically less efficient than $\hat{\theta}_n$.

Some forms of irregular spacing of time series are better viewed in the context of an underlying continuous time process. Spacings would typically be represented as real-valued, possibly generated by a point process. The irregular spacing could be deliberate, in order to avoid loss of identifiability due to aliasing. Again, the Toeplitz structure of $\Omega_n(\theta)$ is lost, and it is generally not possible

to simply approximate either component of (2.10). An exception is when the continuous-time process is generated by a first-order, constant-coefficient, stochastic differential equation driven by white noise. Robinson (1977) deduced a model for the discrete observations, essentially a time-varying first-order autoregression (AR) with heteroscedastic innovations, and consequently approximated (2.10) by a simple form. He established consistency and asymptotic normality of the estimates, but nevertheless in terms of conditions which, to a significant degree, simultaneously restrict the process and the sampling sequence. With more elaborate continuous time models it does not seem possible to deduce a reasonably simple model for the observations, and asymptotic statistical theory would seem difficult to establish under reasonably primitive conditions. See also McDunnough and Wolfson (1979).

2.5   Irregular spacing in spatial data

Irregular spacing is a natural and frequent occurrence with spatial data. In a geographical setting, data are liable to be recorded across heterogeneously-sized administrative regions, while economic distances will not correspond to regular spacing. The difficulties reported above will only be compounded, indeed it even seems hard to substantially extend the model and estimate of Robinson (1977). In general there will not be evident computational simplifications, and while it is possible to write down an asymptotic theory in terms of highly unprimitive conditions, it may be difficult to check them in special cases.

Some of these difficulties can be circumvented by a different approach to modelling which is covered by our set-up, namely spatial AR and related models. Indeed, when there is no geographical aspect, the methods reviewed above are unsuitable. Rules of inference for much microeconomic data routinely take for granted cross-sectional independence, at least at some level, yet there is also an awareness that this can be inappropriate. In some circumstances it is natural to envisage that correlation varies with relevant measures of economic distance, such as differences in household income. Econometricians are familiar with the notion of leads and lags from time series models, and spatial AR models have had considerable appeal; for a recent review of spatial econometrics see Arbia (2006). They rely on specification of an $n \times n$ "weight matrix", which essentially embodies in a simple way notions of irregular spacing. Lee (2004) has developed asymptotic theory for $\hat{\theta}_n$. In general the $\log \det \Omega_n(\theta)$ term in (2.10) cannot be neglected, though for a related model Lee (2002) has shown that this is possible (so least squares works) under suitable conditions on the weight matrix. Under similar conditions, Robinson (2006) has developed asymptotic theory for efficient estimates when the innovations in the spatial AR model are not necessarily normally distributed, both in case of a parametric model for their distribution, and a nonparametric one.

Though asymptotic theory under the null hypothesis of independence is relatively simple with respect to any test statistic, the computational difficulties of point estimation described in the preceding section make LM tests more appealing than Wald or likelihood-ratio ones. These serve to motivate a general class

7

of statistic treated in the following section. It is introduced without reference to LM testing because versions of it lack such an interpretation. Moreover, this will be lost in any case in another statistic also investigated, which nonparametrically robustifies to heteroscedasticity in the u_{in}.

An alternative type of model is motivated by a different form of asymptotics from the "increasing domain" asymptotics usually employed in time series and many spatial settings. This is "fixed domain", or "infill", asymptotics, where the observations are regarded as becoming denser on a bounded region (see e.g. Cressie, 1993, Stein, 1991, Lahiri, 1996). While seemingly more natural in many circumstances, nonstandard results that are not practically useful often emerge, for example estimates may not be consistent, converging to a nondegenerate random variable.

3   A GENERAL CLASS OF TEST STATISTICS

We present a class of test statistics that has a limiting χ^2 distribution under the null hypothesis that the u_{in} in (2.2), (2.3) are independently (and homoscedastically) distributed. For a given $\Omega_n(\theta)$ in (2.4), there is a member of the class that has an LM interpretation, and thus can be expected to have optimal power against local alternatives in directions implied by $\Omega_n(\theta)$. However, such an interpretation is not necessary for the asymptotic validity.

3.1   Testing assuming homoscedasticity

Choose the $p \times 1$ vectors ψ_{ijn}, $i, j = 1, ..., n$, $n \geq 1$, such that $\psi_{iin} = 0$, $\psi_{jin} = \psi_{ijn}$ for all i, j, n. Fixing

$$\Omega_n(0) = I_n, \text{ all sufficiently large } n, \tag{3.1}$$

where I_n is the $n \times n$ identity matrix, define the least squares estimate of β, $\hat{\beta}_n = \hat{\beta}_n(0)$, $\hat{\sigma}_n^2 = \hat{\sigma}_n^2(0)$, and the least squares residuals

$$\hat{u}_n = (\hat{u}_{1n}, ..., \hat{u}_{nn})' = y_n - X_n \hat{\beta}_n. \tag{3.2}$$

Define

$$\hat{a}_n = \sum_{i,j=1}^{n} \psi_{ijn} \hat{u}_{in} \hat{u}_{jn}, \tag{3.3}$$

$$A_n = 2 \sum_{i,j=1}^{n} \psi_{ijn} \psi_{ijn}', \tag{3.4}$$

$$\zeta_n = \hat{\sigma}_n^{-4} \hat{a}_n' A_n^{-1} \hat{a}_n. \tag{3.5}$$

There is no loss of generality in taking $\psi_{ijn} = \psi_{jin}$ because if it were not so we could redefine $\hat{a}_n$ with $\left(\psi_{ijn} + \psi_{jin}\right)/2$ in place of ψ_{ijn}.

8

Denote by x_i the i-th column of X'_n. We allow the x_i to be either deterministically or stochastically generated, but independent of the u_{in}. Likewise, the ψ_{ijn} can also be deterministically or stochastically generated, possibly dependent on the x_i, but again independent of the u_{in}. This is relevant if, say, in a spatial AR model, the weight matrix reflects economic distances between observations measured by the distance between respective stochastically-generated explanatory variables, for example the (i,j)-th element might be proportional to $\|x_i - x_j\| / \left(1 + \|x_i - x_j\|^2\right)$, where the factor of proportionality might vary across rows.

Assumption 1 *For all $n \geq 1$, the u_{in}, $1 \leq i \leq n$, are independent with zero mean, constant variance σ^2, and, for some $\delta > 0$,*

$$\max_{1 \leq i \leq n, \ n \geq 1} E \left|u_{in}\right|^{2+\delta} < \infty. \tag{3.6}$$

Assumption 2 *$\{x_i, \ i \geq 1\}$ is independent of $\{u_{in}, 1 \leq i \leq n, \ n \geq 1\}$, and for some $n \geq q$, X_n has full column rank.*

Define

$$D_n = diag\left\{d_{1n}, ..., d_{pn}\right\}, \tag{3.7}$$

where, with ψ_{ijhn} denoting the h-th element of ψ_{ijn},

$$d_{hn} = \sum_{i,j=1}^{n} \psi_{ijhn}^2, \quad h = 1, ..., p. \tag{3.8}$$

Assumption 3 *$\left\{\psi_{ijn}, \ i, j = 1, .., n, \ n \geq 1\right\}$ is independent of $\{u_{in}, 1 \leq i \leq n, n \geq 1\}$, and, as $n \to \infty$,*

$$\frac{1}{2} D_n^{-\frac{1}{2}} A_n D_n^{-\frac{1}{2}} \to_p R \tag{3.9}$$

for a $p \times p$ positive definite constant matrix R, and

$$d_{hn} \to_p \infty, \quad h = 1, ..., p, \tag{3.10}$$

$$\frac{\max_{1 \leq i \leq n} \sum_{j=1}^{n} \left|\psi_{ijhn}\right|}{d_{hn}^{\frac{1}{2}}} \to_p 0, \quad h = 1, ..., p. \tag{3.11}$$

In time series settings independence in Assumption 1 can be replaced by a martingale difference assumption, but in spatial configurations there may be no natural ordering. The final part of Assumption 3 appears to heavily restrict the ψ_{ijn}, but is satisfied in particular when there is sufficient sparseness, with many elements zero, and it can be checked in LM examples. Asymptotic analysis of

a similar class of statistic was considered by Pinkse (1999, 2004), improving on an earlier treatment of Sen (1976). In some ways his focus was broader, mainly in that his statistic permits investigation also of correlation between two different sets of random variables. Also, he operated in the setting of a more general nonlinear model (see also Kelejian and Prucha, 2001). In this, his regressors are independent of the disturbances, as in Assumption 2 and earlier in the treatment in Robinson (1991) of LM tests in a general class of time series models for regularly-spaced data. As there, we exploit the linear regression structure to enable a treatment under relatively primitive conditions; note also the generality of the last part of Assumption 2, which permits different rates of growth of elements of x_i. Pinkse (1999) did not allow his weights corresponding to ψ_{ijn} to be stochastic, and took $p = 1$. Our allowance for $p > 1$ follows the time series asymptotic treatment of Robinson (1991), and reflects LM statistics against $AR(p)$ and $MA(p)$ time series alternatives (see Godfrey, 1978), and against generalizations of spatial AR models (Anselin, 2001). Notice that no assumptions on the regressors are imposed, except the requirement (in Assumption 2) that X_n has full rank for n large enough.

Theorem A *Let (2.2) hold for all sufficiently large n, and Assumptions 1-3. Then as $n \to \infty$, $\zeta_n \to_d \chi_p^2$.*

The proof is in Appendix A.

3.2 Finite-sample corrections

The null limit distribution established in Theorem A may not provide a satisfactory approximation in smallish samples. Various modifications which aim to provide a closer approximation are possible. One starts from an Edgeworth expansion, leading to size improvements due to a corrected statistic or a bootstrap. Edgeworth expansions are liable to be somewhat complicated even when Gaussianity is assumed, and in the latter circumstance a simpler, if more *ad hoc*, approach is available. The statistic ζ_n can be expressed as the sum of squares of ratios of quadratic forms in u_n, and it is well known that the moments of such statistics can be derived analytically when:

$$u_{1n}, ..., u_{nn} \text{ are independent } N(0, \sigma^2). \tag{3.12}$$

Due to the n-dependence of the ψ_{ijn}, we could, with no loss of generality, have defined ζ_n in such a way that A_n is a diagonal matrix. This would have led to some simplification in our regularity conditions, but the representation of §3.1 was chosen as providing a natural base for the heteroscedasticity correction in the following sub-section, and involves the ψ_{ijn} which have a simple form in the LM applications of §4. However, moment representations are simplified if A_n is diagonal, and in the present sub-section all our results are based on this convention, so that from (3.7)

$$A_n = 2diag\left\{d_{1n}, ..., d_{pn}\right\}. \tag{3.13}$$

Defining also $P_n = I_n - X_n (X_n' X_n)^{-1} X_n'$, $\Psi_{hn} = (\psi_{ijhn})$ (the $n \times n$ matrix whose (i,j)-th element is indicated) and then $r_{hn} = (u_n' P_n \Psi_{hn} P_n u_n) / (u_n' P_n u_n)$, we may write

$$\zeta_n = \frac{n^2}{2} \sum_{h=1}^{p} \frac{r_{hn}^2}{d_{hn}}. \tag{3.14}$$

Under (3.12), the r_{hn} are independent of $u_n' P_n u_n$ (see e.g. Hannan (1970, p.343), who attributed the property to E.J.G. Pitman). Denoting by E_N the expectation under (3.12) conditional on X_n and $\{\psi_{ijn}, \ i,j = 1, ..., n\}$, we have

$$
\begin{aligned}
E_N(\zeta_n) &= \frac{n^2}{2} \sum_{h=1}^{p} \frac{E_N (u_n' P_n \Psi_{hn} P_n u_n)^2}{E_N (u_n' P_n u_n)^2 d_{hn}} \\
&= \frac{n^2}{2(n-q)(n-q+2)} \\
&\quad \times \sum_{h=1}^{p} \frac{tr^2(P_n \Psi_{hn}) + 2tr\{(P_n \Psi_{hn})^2\}}{d_{hn}}.
\end{aligned} \tag{3.15}
$$

Thus

$$\zeta_n^* = p\zeta_n / E_N(\zeta_n) \tag{3.16}$$

has expectation p, which is also the expectation of a χ_p^2 random variable.

The form of the $E_N(\zeta_n)$ correction factor is worth recording for a couple of simple cases. The first is when X_n is absent, so $q = 0$ and $P_n = I_n$. Then $E_N(\zeta_n) = pn/(n+2)$ and thus

$$\zeta_n^* = \left(1 + \frac{2}{n}\right) \zeta_n. \tag{3.17}$$

In the second case, X_n contains only an intercept, i.e. $q = 1$ and $X_n = (1, ..., 1)'$. Then

$$E_N(\zeta_n) = \left(1 - n^{-2}\right)^{-1} (p + s_n), \tag{3.18}$$

where

$$s_n = \sum_{h=1}^{p} \frac{3 \left(\sum_{i,j=1}^{n} \psi_{ijhn}\right)^2 / (2n^2) - \left(\sum_{i,j,k=1}^{n} \psi_{ijhn}\psi_{jkhn}\right) / n}{d_{hn}}, \tag{3.19}$$

leading to

$$\zeta_n^* = (1 - n^{-2}) \left(1 + \frac{s_n}{p}\right)^{-1} \zeta_n. \tag{3.20}$$

Cliff and Ord (1968) have developed a corrrected statistic similar to ζ_n^* in case of a spatial autoregressive model (cf. Section 4.3 below) while Ljung and Box (1978) have done the same in a regularly-spaced time series setting (cf. Section 4.1 below).

11

Returning to the case of general X_n, while ζ_n^* loses its exact mean-p property when (3.12) does not hold, we confirm that in these circumstances it retains the desirable asymptotic properties of ζ_n.

Theorem B *Let (2.2) hold for all sufficiently large n, and Assumptions 1-3. Then as $n \to \infty$, $\zeta_n^* \to_d \chi_p^2$.*

The proof is in Appendix B.

There can be some cost to ζ_n^* in variance inflation; defining V_N as the variance under (3.12) conditional on X_n and $\{\psi_{ijn}, i, j = 1, ..., n\}$, it is clear that $V_N(\zeta_n^*) > V_N(\zeta_n)$ in case (3.17), while in case (3.20) the inequality holds if $s_n > p/(n^2 - 1)$ - the first component of the summand in s_n is always nonnegative, while the second may or may not be. However it is possible to construct a statistic which has both the mean and the variance $(2p)$ of a χ_p^2 random variable, namely

$$\zeta_n^{**} = \{2p/V_N(\zeta_n)\}^{\frac{1}{2}} \{\zeta_n - E_N(\zeta_n)\} + p. \tag{3.21}$$

To calculate this, independence of the r_{hn} from $u_n' P_n u_n$ gives

$$E_N(\zeta_n^2) = \frac{n^4}{4} \sum_{g,h=1}^p \frac{E_N\left\{(u_n' P_n \Psi_{gn} P_n u_n)^2 (u_n' P_n \Psi_{hn} P_n u_n)^2\right\}}{E_N (u_n' P_n u_n)^4 d_{gn} d_{hn}}, \tag{3.22}$$

where, using also (3.15), it is straightforward but tedious to compute

$$
\begin{aligned}
V_N(\zeta_n) \;=\; & \frac{n^4}{4}(n-q)^{-1}\left\{(n-q)^3 + 12(n-q)^2 - 44(n-q) + 48\right\}^{-1} \\
& \times \sum_{g,h=1}^p \Big[8\left\{tr^2\left(P_n \Psi_{gn} P_n \Psi_{hn}\right)\right\} \\
& + 8tr\left(P_n \Psi_{gn}\right) tr\left(P_n \Psi_{gn} P_n \Psi_{hn}\right) tr\left(P_n \Psi_{hn}\right) \\
& + 32tr\left(P_n \Psi_{gn}\right) tr\left\{P_n \Psi_{gn} \left(P_n \Psi_{hn}\right)^2\right\} \\
& + 32tr\left\{\left(P_n \Psi_{gn}\right)^2 \left(P_n \Psi_{hn}\right)^2\right\} \\
& + 16tr\left\{\left(P_n \Psi_{gn} P_n \Psi_{hn}\right)^2\right\} \Big] / d_{gn} d_{hn}. \tag{3.23}
\end{aligned}
$$

For example, in the case X_n is absent from (2.2)

$$V_N(\zeta_n) = \left\{1 + \frac{12}{n} + \frac{44}{n^2} + \frac{48}{n^3}\right\}^{-1}$$

$$\times \left[2p + 16 \sum_{g,h=1}^p \left(2tr\left(\Psi_{gn}^2 \Psi_{hn}^2\right) + tr\left\{\left(\Psi_{gn} \Psi_{hn}\right)^2\right\}\right) / d_{gn} d_{hn}\right]. \tag{3.24}$$

As with ζ_n^*, we record asymptotic justification of ζ_n^{**} in the more general setting of §3.1.

Theorem C *Let (2.2) hold for all sufficiently large n, and Assumptions 1-3. Then as $n \to \infty$, $\zeta_n^{**} \to_d \chi_p^2$.*

The proof is in Appendix C.

3.3 Testing with robustness to heteroscedasticity

While Assumption 1 does not assume identity of distribution, and limits constancy of moments to the mean and variance, homoscedasticity seems an unreasonable assumption in many kinds of spatial data, where, for example, observations are based on aggregation over administrative regions that differ considerably in size. Versions of ζ_n designed to test for correlation may be significant due to unanticipated heteroscedasticity. In fact, formally, certain versions of ζ_n can be interpreted as LM tests of (conditional or unconditional) heteroscedasticity, not just correlation, though we do not stress this aspect because asymptotically Gauss-Markov efficient weighted least squares estimation of β_0, treating either parametric or nonparametric heteroscedasticity, is entirely feasible. Instead we robustify ζ_n to heteroscedasticity.

Define

$$\hat{B}_n = 2 \sum_{i,j=1}^n \psi_{ijn} \psi'_{ijn} \hat{u}_{in}^2 \hat{u}_{jn}^2, \tag{3.25}$$

$$\xi_n = \hat{a}'_n \hat{B}_n^{-1} \hat{a}_n. \tag{3.26}$$

This weighting by squared raw residuals is in the spirit of heteroscedasticity-consistent variance estimation first introduced by Eicker (1963), and much employed since by econometricians. We modify two of our previous assumptions accordingly. Define

Assumption 1* *Assumption 1 holds, with $\delta = 2$ in (3.6) but without the requirement that the variance of u_{in}, now denoted σ_i^2, be constant over i; $\min_{i \geq 1} \sigma_i^2 > 0$.*

Assumption 3* *Assumption 3 holds with (3.9) replaced by*

$$\frac{1}{2} D_n^{-\frac{1}{2}} B_n D_n^{-\frac{1}{2}} \to_p S, \quad \text{as } n \to \infty, \tag{3.27}$$

for some positive definite constant matrix S, where

$$B_n = 2 \sum_{i,j=1}^n \psi_{ijn} \psi'_{ijn} \sigma_i^2 \sigma_j^2. \tag{3.28}$$

The fourth moment condition on u_{in} seems unavoidable, indeed some care is needed in the proof to avoid something stronger. Notice it implies, via Hölder's inequality, that

$$\max_{i \geq 1} \sigma_i^2 < \infty. \tag{3.29}$$

13

Theorem D *Let (2.2) hold for all sufficiently large n, and Assumptions 1^*, 2 and 3^*. Then as $n \to \infty$, $\xi_n \to_d \chi_p^2$.*

The proof is in Appendix D

4 LOCAL POWER AND LAGRANGE MULTIPLIER-MOTIVATED SPECIAL CASES

Considering the limit distribution of ζ_n under local departures from uncorrelatedness of the u_{in} will motivate a focus on the LM special case. In the parametric setting of §2 the null hypothesis of zero correlation is

$$H_0 : \theta_0 = 0. \tag{4.1}$$

The alternative hypothesis postulates departures of the form

$$H_{1n} : \theta_0 = \theta_{(n)} = D_n^{-\frac{1}{2}}\nu, \tag{4.2}$$

where ν is a $p \times 1$ fixed vector. Defining by $A^{\frac{1}{2}}$ the usual positive definite matrix such that $A^{\frac{1}{2}}A^{\frac{1}{2}} = A$, for positive definite A, we correspondingly replace Assumption 1 by

Assumption 4 *For all n,*

$$u_n = \Omega_n^{\frac{1}{2}}(\theta_n)\varepsilon_n, \tag{4.3}$$

where $\varepsilon_n = (\varepsilon_{1n}, ..., \varepsilon_{nn})'$, in which the ε_{in} are independent with zero mean, constant variance σ^2 and, for some $\delta > 0$

$$\max_{1 \leq i \leq n, n \geq 1} E\,|\varepsilon_{in}|^{2+\delta} < \infty. \tag{4.4}$$

The following regularity condition on $\Omega_n(\theta)$ is introduced in order to derive local distributional properties. Denote by $\bar{\lambda}(A)$ the largest eigenvalue of the non-negative definite matrix A.

Assumption 5 *$\Omega_n(\theta)$ satisfies (3.1), and for all sufficiently large n, its (i,j)th element $\omega_{ijn}(\theta)$ is boundedly differentiable in a neighbourhood of $\theta = 0$, uniformly in $i, j = 1, ..., n$. For all sufficiently large n, and all $\eta > 0$, there exists $\tau > 0$ such that*

$$\sup_{\|\theta\|<\tau} \bar{\lambda}\left(\Omega_n^{-1}(\theta)\right) < \eta^{-1}, \tag{4.5}$$

$$\sup_{\|\theta\|<\tau} \bar{\lambda}\left(\left\{\Omega_n^{\frac{1}{2}}(\theta) - I_n\right\}^2\right) < \eta, \tag{4.6}$$

14

and for $h = 1, ..., p$

$$\sup_{\|\theta\| < \tau} \bar{\lambda}\left(\{\Phi_{hn}(\theta) - \Phi_{hn}\}^2\right) < \eta, \tag{4.7}$$

where $\Phi_{hn}(\theta)$ is the $n \times n$ matrix with (i,j)th element $\phi_{ijhn}(\theta) = (\partial/\partial\theta_h)\omega_{ijn}(\theta)$, and $\Phi_{hn} = \Phi_{hn}(0)$. Denoting $\phi_{ijhn} = \phi_{ijhn}(0)$, $\phi_{ijn} = (\phi_{ij1n}, ..., \phi_{ijpn})'$ and

$$C_n = \sum_{i,j=1}^{n} \psi_{ijn}\phi'_{ijn}, \tag{4.8}$$

then as $n \to \infty$

$$\frac{1}{2}D_n^{-\frac{1}{2}}C_n D_n^{-\frac{1}{2}} \to_p T, \tag{4.9}$$

for a $p \times p$ constant non-null matrix T.

Part (4.5) is a local non-singularity assumption on $\Omega_n(\theta)$, while (4.6) and (4.7) are essentially continuity assumptions. Denote by $\chi'^2_p(\mu)$ a non-central χ^2 variate with p degrees of freedom and non-centrality parameter μ.

Theorem E *Let (2.2) hold for all sufficiently large n, and Assumptions 2-5. Then under (4.2), as $n \to \infty$ $\zeta_n \to_d \chi'^2_p\left(\nu'T'R^{-1}T\nu/\sigma^4\right)$. Furthermore, if*

$$\psi_{ijn} = \phi_{ijn}, \quad 1 \le i,j \le n, \quad n \ge 1, \tag{4.10}$$

the non-centrality parameter achieves its maximum.

The proof is in Appendix E.

If T has full rank, the Theorem indicates that ζ_n has local power with respect to all departures of form (4.2) from H. Given the well-known optimality properties of LM statistics, it is unsurprising that, as (4.10) indicates, the LM statistic is locally optimal within the ζ_n class.

In the following sub-sections we develop the LM statistic in case of a number of models, covering several kinds of data.

4.1 Missing data in time series

Here $\{y_t\}$ are the consecutive, un-missed observations from a regularly-spaced time series. Correspondingly $u_n = (u(t_1), ..., u(t_n))'$, where the t_i are integers, $t_1 < t_2 < ... < t_n$, and $u(t)$ is stationary with zero mean and lag-j autocovariance $\gamma(j; \theta_0)$, where $\gamma(j; \theta)$ is a known function of j, θ. Thus $\psi_{ijn} = (\partial/\partial\theta)\gamma(t_i - t_j; 0)$. The ψ_{ijn} are thus functions of $\{t_i\}$, which may be deterministically or stochastically generated, as Assumptions 3 and 3* permit.

One special case not previously considered is a missing-data version of the test of Robinson (1991) against long memory or antipersistent alternatives. Here

15

$p = 1$ and $(1 - L)^{\theta_0} u_i = \varepsilon_i$, where L is the lag operator, the ε_i are independent and heteroscedastic, and $|\theta_0| < \frac{1}{2}$. Then $\psi_{ijn} = |t_i - t_j|^{-1}$, for $i \neq j$. Part (3.10) of Assumption 3 is satisfied if $d_{1n} = \Sigma_{i,j,i \neq j}^n |t_i - t_j|^{-2} \to_p \infty$. In case there is no missing, or with periodic or roughly periodic missing, d_{1n} increases at rate n, but a slower rate with missing is possible, permitting observations to "peter out". We have $\Sigma_{j=1,j\neq i}^n |t_i - t_j|^{-1} \leq \Sigma_{i=1}^{t_n} i^{-1} \sim \log t_n$, uniformly in i, so that (3.11) is implied by

$$(\log t_n)^2 / d_{1n} \to_p 0. \tag{4.11}$$

Another leading alternative is the $AR(p)$ hypothesis already considered by Robinson (1986), who obtained a missing-data version of the Box and Pierce (1970) statistic. We have $\psi_{ijkn} = 1\left(|t_i - t_j| = k\right), d_{kn} = \Sigma_{i,j=1,i\neq j}^n 1\left(|t_i - t_j| = k\right),$ $k = 1, ..., p.$. Then if (3.10) holds so does (3.11), as the numerator of its left-hand side is 2. Differential rates of increase permitted by our conditions are possible for the d_{kn}; for example, starting from a periodic sampling framework in which two consecutive observations are followed by five missed ones, replace the first of the latter by an observed value at points t_i at intervals of order $i^{3/2}$, so d_{1n} increases at rate n whereas d_{2n} increases at rate $n^{2/3}$.

4.2 d-dimensional lattice

Introduce the d-dimensional lattice $\mathcal{L}_d = \{I : I = (i_1, ..., i_d), \ i_j = 0, \pm 1, ..., \ j = 1, ..., d\}$, for $d > 1$. We observe Y_I, for $I \in N = \{I : i_j = 1, ..., n_j, \ j = 1, ..., d\}$, and take $n = \Pi_{j=1}^d n_j$. Corresponding to (2.2) we have $Y_I = \beta_0' x_I + u_I, \ I \in N$. Identifying the i-th element of u_n with U_I (possibly with lexicographic ordering), correspondingly denote the i-th element of $\hat{u}_n$ by $\hat{U}_{In}$. Suppose U_I is stationary with autocovariance $Cov\left(U_I, U_{I+J}\right) = \gamma(J; \theta_0)$ for $J \in \mathcal{L}_d$, where $\gamma(J; \theta)$ is boundedly differentiable in θ but θ_0 is unknown. Denote $\Psi_I = (\partial/\partial\theta)\gamma\left(I; 0\right)$. Thus ζ_n and ξ_n are given by (3.5) and (3.16) with

$$\hat{\sigma}_n^2 = n^{-1} \sum_{I \in N} \hat{U}_{In}^2, \qquad \hat{a}_n = \sum_{I, J \in N} \Psi_{I-J} \hat{U}_{In} \hat{U}_{Jn}, \tag{4.12}$$

$$A_n = 2 \sum_{I, J \in N} \Psi_{I-J} \Psi_{I-J}', \quad \hat{B}_n = 2 \sum_{I, J \in N} \Psi_{I-J} \Psi_{I-J}' \hat{U}_{In}^2 \hat{U}_{Jn}^2. \tag{4.13}$$

One example tests against long memory, taking $d = 1$ and $\Pi_{j=1}^d (1 - L_j)^{\theta_0} U_I = \varepsilon_I$ where L_j is the lag-operator in the j-th dimension only, the ε_I are independent and homoscedastic, and $|\theta_0| < \frac{1}{2}$. Then $\Psi_I = \Sigma_{j=1}^d i_j^{-1}$.

Tests against AR alternatives are also available. Let P be a set of p distinct I indices, such that $I \neq \{0, ..., 0\}$, and consider the model

$$U_I - \sum_{J \in P} \Theta_{0J} \prod_{\substack{i=1 \\ (j_1, ..., j_d) = J}}^{d} L^{j_i} U_{I-J} = \varepsilon_I, \tag{4.14}$$

16

with ε_I as before. Now θ_0 consists of scalars Θ_{0J} (which must satisfy stationarity conditions if $\theta_0 \neq 0$). A typical element of Ψ_I is

$$\Psi_{IJ} = \frac{\partial}{\partial \Theta_{0J}} \gamma(I;0) = 1(I = J), \quad J \in P. \qquad (4.15)$$

However there is a restriction on P which affects multilateral modelling that is motivated by a lack of natural ordering in one or more of the dimensions; in spatio-temporal data there is a natural ordering in the time dimension, but typically not in the others. There is thus a temptation to include J in (4.15) that contain some negative indices, as well as J with all non-negative ones. This can present identification problems as recently reviewed by Robinson and Vidal Sanz (2006). We encounter a corresponding problem. From (4.15) and the symmetry property $\gamma(I;\theta) = \gamma(-I;\theta)$ it is clear that for $K = -J$

$$\Psi_{IK} = \left(\frac{\partial}{\partial \Theta_{0K}}\right) \gamma(I;0) = -1(-I = K) = -1(I = J). \qquad (4.16)$$

Thus A_n (and $\hat{B}_n$) will not be invertible. We might also think of including such mirror-image J but constraining their coefficients to be equal. This avoids the identifiability condition but it is easily seen to produce the same statistic as if we included only one of them. Altogether, taking, say, all J to have non-negative elements, we get

$$\zeta_n = \sum_{J \in P} \left(n^{-1} \sum_{I, I+J \in N} \hat{U}_I \hat{U}_{I+J} \right)^2 / \hat{\sigma}_n^4, \qquad (4.17)$$

a natural extension of the Box and Pierce (1970) statistic for time series. With respect to potential "edge effect", the discrepancy between N and the numbers of summands over I has no asymptotic effect under the null hypothesis because there is no bias, due to $E(U_I U_{I+J}) = 0$, $J \neq \{0,...,0\}$.

Tests can also be based on more parsimonious models that have the property of isometry. For example take $\gamma(I;\theta) = \theta^{\|I\|}$, for scalar $\theta \in (-1,1)$. Then

$$\zeta_n = \left(\sum_{I \in N} U_I \sum_{J: \|J\| = 1} U_{I+J} \right)^2 / \left(\sum_{I, J \in N} 1(\|I - J\| = 1) \right). \qquad (4.18)$$

It is straightforward to extend the above statistics to allow for missing observations, in the manner of the previous sub-section.

4.3 Spatial autoregressive models

Spatial AR models are especially convenient when there is irregular spacing that cannot be handled in the framework of missing values in an otherwise regular

time series or lattice, or when the space is economic rather than geographic. Consider the model

$$\left(I_n - \sum_{k=1}^{p} \theta_{0k} W_{kn} \right) u_n = \varepsilon_n, \tag{4.19}$$

where ε_n is a vector of independent, homoscedastic variables, and the W_{kn} are $n \times n$ weight matrices, possibly stochastically generated and possibly X_n-dependent. The most familiar version of (4.19) has $p = 1$. Anselin (2001) discussed LM tests for spatial independence against a related model where instead of combining (2.2) with (4.19), one incorporates spatially lagged y's in (2.2). The null model is the same in both cases.

Testing for spatial independence in (4.19) and related models, both in the linear regression setting (2.1) and more general ones, has been widely considered (see e.g. Baltagi and Dong Li, 2001, Kelejian and Prucha, 2001), and our purpose here is not to present new tests but to discuss conditions on the W_{kn} for asymptotic validity, and consider the connection with infill asymptotics. The identifiability problem in (4.19) is similar to that discussed by Anselin (2001) in his model. We have

$$\psi_{ijkn} = 2W_{ijkn}, \tag{4.20}$$

where W_{ijkn} is the (i, j)-th element of W_{kn}. Then

$$A_n = \left(4 \sum_{i,j=1}^{n} W_{ijkn} W_{ij\ell n} \right), \tag{4.21}$$

indicating the (k, ℓ)-th element. It is obvious that positive definiteness of R in (3.9) requires in particular that all the W_{kn} must differ. Anselin (2001) assumed that

$$\sum_{j=1}^{n} W_{ijkn} W_{ij\ell n} = 0 \tag{4.22}$$

for all i and for $k \neq \ell$, which implies that A_n is diagonal; a special case is where the n observations are sub-divided into subsets such that W_{kn} has zero elements corresponding to the non-kth subsets, so $\Sigma_{k=1}^{p} W_{kn}$ is block diagonal. Indeed (4.22) requires existence of some negative weights unless $W_{ijkn} = 0$ or $W_{ij\ell n} = 0$ for each i, j and each $k \neq \ell$.

The preceding discussion applies also to $\hat{B}_n$ in ξ_n. Kelejian and Robinson (2004) considered heteroscedasticity in a spatial AR context but applied it to ε_n and adopted a different approach to the problem.

With respect to both ζ_n and ξ_n, conditions (3.10) and (3.11) become respectively

$$d_{hn} = \sum_{i,j=1}^{n} W_{ijhn}^2 \to_p \infty, \quad h = 1, ..., p, \tag{4.23}$$

18

$$\frac{\max\limits_{1 \le i \le n} \sum\limits_{j=1}^{n} |W_{ijhn}|}{d_{hn}^{\frac{1}{2}}} \to_p 0, \quad h = 1, ..., p \tag{4.24}$$

(cf. Sen, 1976, Pinkse, 1999, 2004). Given (4.23), a sufficient condition for (4.24) is that W_{hn} have non-negative elements and are row-normalized. If $W_{jkmn} = O_p\left(h_n^{-1}\right)$ uniformly, for some nonnegative sequence h_n (cf. Lee, 2002), then

$$n/h_n \to \infty, \quad m = 1, ..., p \tag{4.25}$$

is necessary for (3.10).

Some formal comparison is possible between our asymptotic discussion, and (4.25) in particular, and infill asymptotics. Consider for simplicity in place of (4.19) the first order spatial MA

$$u_n = (I_n + \theta_0 W_{1n}) \varepsilon_n. \tag{4.26}$$

On the other hand consider a process $u(t)$, $t \in (0, 1]$ such that

$$u(t) = \varepsilon_t + \frac{1}{n} \sum_{s=1}^{n} \alpha\left(t - \frac{s}{n}; \theta_0\right) \varepsilon_s, \quad t \in (0, 1], \tag{4.27}$$

for a function $\alpha(t; \theta)$, $|t| \le 1$, that is boundedly differentiable in θ. (Extension to a process defined on a finite region in d dimensions is immediate.) For example, $\alpha(t; \theta) \equiv \theta$, where there is a close formal similarity with (4.26). Consider sampling $u(t)$ at intervals $1/n$. Thus taking $u_n = (u(1/n), ..., u(1 - 1/n))'$ and applying the LM principle for testing $\theta_0 = 0$ we find that (3.10) is violated. Likewise, since $h_n \sim n^{-1}$, (4.27) contradicts (4.25).

APPENDIX A: PROOF OF THEOREM A

Proof. We write ψ_{ij} for ψ_{ijn} throughout. The limit distribution is independent of the x_i and ψ_{ij}, so it suffices to show that the result holds conditionally on $\{x_i,\ i \geq 1\}$ and $\{\psi_{ij},\ 1 \leq i, j \leq n,\ n \geq 1\}$; correspondingly, all expectations in what follows will thereby be conditional, though we suppress reference to this. Define $a_n = \Sigma_{i,j}\psi_{ij}u_i u_j$, writing $\psi_{ij} = \psi_{ijn}$ and unqualified summation over i covering $i = 1, ..., n$. The result follows from

$$\hat{\sigma}_n^2 \ \to_p \ \sigma^2, \tag{A.1}$$

$$A_n^{-\frac{1}{2}}\left(\hat{a}_n - a_n\right) \ \to_p \ 0, \tag{A.2}$$

and

$$\sigma^{-2}A_n^{-\frac{1}{2}}a_n \ \to_p \ N(0, I_p). \tag{A.3}$$

We omit the proof of (A.1), as it is essentially implied by that of (A.2). To consider this, write $\hat{u}_i = \hat{u}_{in}$, $\hat{v}_i = \hat{u}_i - u_i = \Sigma_j u_j b_{ij}$, where $b_{ij} = x_i'\left(\Sigma_h x_h x_h'\right)^{-1}x_j$. Thus

$$\hat{a}_n - a_n = \sum_{i,j}\psi_{ij}\left(\hat{v}_i u_j + u_i\hat{v}_j + \hat{v}_i\hat{v}_j\right). \tag{A.4}$$

With $\|.\|$ denoting Euclidean norm,

$$\left\|A_n^{-\frac{1}{2}}D_n^{\frac{1}{2}}\right\|^2 = tr\left\{D_n^{\frac{1}{2}}A_n^{-1}D_n^{\frac{1}{2}}\right\} \ \to_p \ \frac{1}{2}tr(R^{-1}), \tag{A.5}$$

so we can prove (A.2) with $A_n^{-\frac{1}{2}}$ replaced by $D_n^{-\frac{1}{2}}$. We consider an arbitrary element, and so to avoid additional subscripting ψ_{ij} for the time being represents a scalar.

We have

$$\sum_{i,j}\psi_{ij}\hat{v}_i u_j = \sum_{i,j}\psi_{ij}u_j^2 b_{ij} + \sum_{i,j}\psi_{ij}u_j\sum_{h \neq j}u_h b_{ih}. \tag{A.6}$$

The modulus of the first term on the right has expectation bounded by

$$C\sum_{i,j}\left|\psi_{ij}b_{ij}\right| \leq C\sum_{i,j}\left|\psi_{ij}\right|(b_{ii} + b_{jj}) \tag{A.7}$$

because by the Cauchy and elementary inequalities $|b_{ij}| \leq b_{ii}^{\frac{1}{2}}b_{jj}^{\frac{1}{2}} \leq (b_{ii}+b_{jj})/2$, C denoting throughout a generic constant. Because $\Sigma_i b_{ii} = q$ and $\psi_{ij} = \psi_{ji}$, this is bounded by $C\max_i \Sigma_j \left|\psi_{ij}\right|$. The second term has mean zero and variance bounded by

$$C\sum_{i,j,\ell,m}\psi_{ij}\psi_{\ell m}b_{im}b_{\ell j} + C\sum_{h,i,j,\ell}\psi_{ij}\psi_{\ell j}b_{ih}b_{\ell h} \tag{A.8}$$

20

$$\leq C \left(\sum_{i,j} |\psi_{ij}| \, b_{ii}^{\frac{1}{2}} b_{jj}^{\frac{1}{2}} \right)^2 + C \sum_{i,j,\ell} |\psi_{ij}| \, |\psi_{\ell j}| \, b_{ii}^{\frac{1}{2}} b_{\ell\ell}^{\frac{1}{2}}$$

$$\leq C \left(\sum_{i,j} |\psi_{ij}| \, (b_{ii} + b_{jj}) \right)^2 + C \sum_{i,j,\ell} |\psi_{ij}| \, |\psi_{\ell j}| \, (b_{ii} + b_{\ell\ell})$$

$$\leq C \left(\max_i \sum_j |\psi_{ij}| \right)^2. \tag{A.9}$$

Next,

$$\sum_{i,j} \psi_{ij} \hat{v}_i \hat{v}_j = \sum_{i,j} \psi_{ij} \left(\sum_\ell u_\ell b_{i\ell} \right) \left(\sum_m u_m b_{jm} \right)$$

$$= \sum_{i,j,\ell} \psi_{ij} b_{i\ell} b_{j\ell} u_\ell^2 + \sum_{i,j,k} \sum \psi_{ij} b_{ik} \sum_{\ell \neq k} b_{j\ell} u_k u_\ell. \tag{A.10}$$

The first term on the right has modulus with expectation bounded by

$$C \sum_{i,j} |\psi_{ij}| \, (b_{ii} + b_{jj}) \leq C \max_i \sum_j |\psi_{ij}| \tag{A.11}$$

as before. The second term has mean zero and variance bounded by

$$\left| \sum_{h,i,j,k,\ell,m} \psi_{ij} \psi_{hm} b_{ik} b_{j\ell} \, (b_{hk} b_{m\ell} + b_{h\ell} b_{mk}) \right|$$

$$\leq C \left(\sum_{i,j} \psi_{ij} b_{ii}^{\frac{1}{2}} b_{jj}^{\frac{1}{2}} \right)^2 \leq C \left(\max_i \sum_j |\psi_{ij}| \right)^2 \tag{A.12}$$

as before. Then (A.2) follows from Assumption 3. To prove (A.3) we show that for all $p \times 1$ vectors λ, such that $\|\lambda\|^2 = 1$,

$$\sigma^{-2} \lambda' A_n^{-\frac{1}{2}} a_n \rightarrow_d N(0,1) \tag{A.13}$$

conditionally. The left side can be written $\Sigma_i z_{in}$, where

$$z_{in} = 2\sigma^{-2} u_i \lambda' A_n^{-\frac{1}{2}} \sum_{j<i} \psi_{ij} u_j. \tag{A.14}$$

Clearly $\Sigma_i z_{in}$ has mean zero and variance 1, so (A.3) follows from Theorem 2 of Scott (1973) on showing that conditionally

$$\sum_i E\left(z_{in}^2 | u_j, \ j < i \right) \rightarrow_p 1 \tag{A.15}$$

$$\sum_i E\left\{z_{in}^2 1\left(|z_{in}| \geq \varepsilon\right)\right\} \rightarrow_p 0, \quad \forall \varepsilon > 0. \tag{A.16}$$

Since

$$E\left(z_{in}^2 \mid u_j, j < i\right) = 4\sigma^2 \lambda' A_n^{-\frac{1}{2}} \left(\sum_{j<i} \psi_{ij} u_j\right) \left(\sum_{j<i} \psi_{ij} u_j\right)' A_n^{-\frac{1}{2}} \lambda \tag{A.17}$$

and

$$A_n^{-\frac{1}{2}} \sum_i \sum_{j<i} \psi_{ij} \psi_{ij}' A_n^{-\frac{1}{2}} = \frac{1}{4} I_p, \tag{A.18}$$

we deduce from (A.5) that (A.15) will follow if

$$D_n^{-\frac{1}{2}} \sum_i \left\{ \left(\sum_{j<i} \psi_{ij} u_j\right) \left(\sum_{j<i} \psi_{ij} u_j\right)' - \sigma^2 \sum_{j<i} \psi_{ij} \psi_{ij}' \right\} D_n^{-\frac{1}{2}} \rightarrow_p 0. \tag{A.19}$$

Again we consider a typical element, and again identify a scalar ψ_{ij} with this; strictly speaking the differential norming needs to be taken account of, but this is a routine aspect. Thus in place of the expression in braces in (A.19) we consider

$$2\sum_i \sum_{j<i} \psi_{ij}^2 \left(u_j^2 - \sigma^2\right) + \sum_i \sum_{\substack{j,k<i \\ j \neq k}} \psi_{ij} \psi_{ik} u_j u_k. \tag{A.20}$$

By inequalities of Jensen and of von Bahr and Esseen (1965), the modulus of the first term has (conditionally) mean bounded by

$$
\begin{aligned}
C\left\{\sum_i \left|\sum_{j<i} \psi_{ij}^2\right|^{1+\delta/2}\right\}^{2/(2+\delta)} &\leq C\left\{\left(\max_i \sum_j \psi_{ij}^2\right)^{\delta/2} \sum_{i,j} \psi_{ij}^2\right\}^{2/(2+\delta)} \\
&\leq C\left(\max_i \sum_j |\psi_{ij}|\right)^{\delta} \left\{\sum_{i,j} \psi_{ij}^2\right\}^{2/(2+\delta)} \\
&= o_p\left(\sum_{i,j} \psi_{ij}^2\right), \tag{A.21}
\end{aligned}
$$

as desired. The second term in (A.20) has (conditionally) zero mean and vari-

ance bounded by

$$
\begin{aligned}
C \sum_{\substack{h,i,j,k \\ j,k<i,h}} \left| \psi_{ij}\psi_{ik}\psi_{hj}\psi_{hk} \right| &\leq C \sum_{h,i,j,k} \left| \psi_{ij}\psi_{ik} \right| \left(\psi_{hj}^2 + \psi_{hk}^2 \right) \\
&\leq C \left(\max_i \sum_k |\psi_{ik}| \right) \left(\max_j \sum_i |\psi_{ij}| \right) \sum_{h,j} \psi_{hj}^2 \\
&= o_p \left(\left(\sum_{i,j} \psi_{ij}^2 \right)^2 \right)
\end{aligned}
\tag{A.22}
$$

by Assumption 3. This completes the proof of (A.15). To prove (A.16) we check the sufficient Lyapunov condition

$$
\sum_i E \left| z_{in} \right|^{2+\delta} \to_p 0,
\tag{A.23}
$$

where again we condition on the ψ_{ij}. We can instead check the condition with z_{in} replaced by $\left(\Sigma_{i,j}\psi_{ij}^2 \right)^{-\frac{1}{2}} u_i \Sigma_{j<i}\psi_{ij}u_j$, again treating ψ_{ij} as a generic element. By the Marcinkiewicz-Zygmund inequality it suffices to show that

$$
\left(\sum_{i,j} \psi_{ij}^2 \right)^{-1-\delta/2} \sum_i E \left(\sum_j \psi_{ij}^2 u_j^2 \right)^{1+\delta/2} \to_p 0.
\tag{A.24}
$$

Now

$$
\begin{aligned}
\sum_i E \left(\sum_j \psi_{ij}^2 u_j^2 \right)^{1+\delta/2} &= \sum_i E \left(\sum_j \psi_{ij}^2 \left(u_j^2 - \sigma^2 \right) + \sigma^2 \sum_j \psi_{ij}^2 \right)^{1+\delta/2} \\
&\leq C \left\{ \sum_{i,j} \left| \psi_{ij} \right|^{2+\delta} + \left(\sum_{i,j} \psi_{ij}^2 \right)^{1+\delta/2} \right\} \\
&\leq C \left(\sum_{i,j} \psi_{ij}^2 \right)^{1+\delta/2}.
\end{aligned}
\tag{A.25}
$$

Thus the left side of (A.22) is bounded by

$$
C \left(\sum_{i,j} \psi_{ij}^2 \right)^{-1-\delta/2} \sum_i \left(\sum_j \psi_{ij}^2 \right)^{1+\delta/2}
$$

$$
\leq \quad \left\{ \left(\max_i \sum_j \psi_{ij}^2 \right) / \sum_{i,j} \psi_{ij}^2 \right\}^{\delta/2}
$$

$$
\leq \quad \left\{ \left(\max_i \sum_j |\psi_{ij}| \right)^2 / \sum_{i,j} \psi_{ij}^2 \right\}^{\delta/2} \quad \to_p 0. \qquad \text{(A.26)}
$$

$\square$

APPENDIX B: PROOF OF THEOREM B

It suffices to show that (3.15) $\to_p p$, i.e. that, for $h = 1, ..., p$,

$$
tr\left(P_n \Psi_{hn}\right) \quad = \quad o_p\left(d_{hn}^{\frac{1}{2}}\right), \qquad \text{(B.1)}
$$

$$
tr\left\{\left(P_n \Psi_{hn}\right)^2\right\} / d_{hn} \quad \to \quad _p 1. \qquad \text{(B.2)}
$$

Writing $V_n = X_n \left(X_n' X_n\right)^{-\frac{1}{2}}$, the left side of (B.1) is $-tr\left(V_n' \Psi_{hm} V_n\right)$, which is bounded in absolute value by $tr^{\frac{1}{2}}\left(V_n' \Psi_{hn}^2 V_n\right) tr^{\frac{1}{2}}\left(V_n' V_n\right) \leq q\bar{\lambda}^{\frac{1}{2}}\left(\Psi_{hn}^2\right)$. But with $v = (v_1, ..., v_n)'$ denoting a non-null vector,

$$
\bar{\lambda}\left(\Psi_{hn}^2\right) \quad = \quad \max_v v' \Psi_{hn}^2 v / v' v = \sum_{i,j} v_i v_j \sum_k \psi_{ikh} \psi_{kjh} / v' v
$$

$$
\leq \quad \sum_{i,j} v_i^2 \left| \sum_k \psi_{ikh} \psi_{kjh} \right| / v' v \leq \left(\max_k \sum_j |\psi_{kjh}| \right) \left(\max_i \sum_k |\psi_{ikh}| \right)
$$

$$
= \quad o_p\left(d_{hn}\right). \qquad \text{(B.3)}
$$

The left side of (B.2) is

$$
1 - \left\{ 2tr\left(V_n' \Psi_{hn}^2 V_n\right) - tr\left(V_n V_n' \Psi_{hn}\right)^2 \right\} / d_{hn}. \qquad \text{(B.4)}
$$

The first term in braces has already been shown to be $o_p(d_{hn})$. The second term in braces is bounded in absolute value by $qtr\left(V_n' \Psi_{hn}^2 V_n\right) = o_p\left(d_{hn}\right)$, to complete the proof. $\square$

24

APPENDIX C: PROOF OF THEOREM C

An easy extension of the proof of (B.2) shows that $tr\left\{P_n\Psi_{gn}P_n\Psi_{hn}\right\}/d_{gn}^{\frac{1}{2}}d_{hn}^{\frac{1}{2}} \to_p 1$, whence in view of Theorem C it suffices to show that

$$tr\left\{P_n\Psi_{gn}\left(P_n\Psi_{hn}\right)^2\right\} = o_p\left(d_{gn}^{\frac{1}{2}}d_{hn}\right), \tag{C.1}$$

$$tr\left\{P_n\Psi_{gn}P_n\Psi_{hn}P_n\Psi_{in}P_n\Psi_{jn}\right\} = o_p\left(d_{gn}^{\frac{1}{2}}d_{hn}^{\frac{1}{2}}d_{in}^{\frac{1}{2}}d_{jn}^{\frac{1}{2}}\right). \tag{C.2}$$

The left side of (C.2) is bounded in absolute value by

$$\bar{\lambda}\left\{(P_n\Psi_{gn})^2\right\}\bar{\lambda}\left\{(P_n\Psi_{in})^2\right\}tr\left\{(P_n\Psi_{hn})^2\right\}tr\left\{(P_n\Psi_{jn})^2\right\}. \tag{C.3}$$

Since $\bar{\lambda}\left\{(P_n\Psi_{gn})^2\right\} \leq \bar{\lambda}\left\{\Psi_{gn}^2\right\}$, (C.2) follows from (B.2) and (B.3). The left side of (C.1) is bounded in absolute value by

$$\bar{\lambda}\left\{(P_n\Psi_{gn})^2\right\}^{\frac{1}{2}}tr\left\{(P_n\Psi_{hn})^2\right\}, \tag{C.4}$$

whence (C.1) follows as before. □

APPENDIX D: PROOF OF THEOREM D

Proof. In place of (A.1)-(A.3) we need that, as $n \to \infty$,

$$\frac{1}{2}D_n^{-\frac{1}{2}}\hat{B}_nD_n^{-\frac{1}{2}} \to_p S, \tag{D.1}$$

$$B_n^{-\frac{1}{2}}\left(\hat{a}_n - a_n\right) \to_p 0, \tag{D.2}$$

$$B_n^{-\frac{1}{2}}a_n \to_d N(0, I_p). \tag{D.3}$$

To prove (D.1) we define

$$B_n = 2\sum_{i,j}\psi_{ij}\psi'_{ij}\sigma_i^2\sigma_j^2 \tag{D.4}$$

and prove

$$D_n^{-\frac{1}{2}}\left(\tilde{B}_n - B_n\right)D_n^{-\frac{1}{2}} \to_p 0, \tag{D.5}$$

$$D_n^{-\frac{1}{2}}\left(\hat{B}_n - \tilde{B}_n\right)D_n^{-\frac{1}{2}} \to_p 0. \tag{D.6}$$

In both cases, as in part of the proof of Theorem 1, it clearly suffices to give the proof as if $p = 1$, and show that $\tilde{B}_n - B_n$ and $\hat{B}_n - \tilde{B}_n$ are both $o_p\left(\Sigma_{i,j}\psi_{ij}^2\right)$.

We have

$$\tilde{B}_n - B_n = \sum_{i,j} \psi_{ij}^2 \left\{ 2\sigma_i^2(u_j^2 - \sigma_j^2) + (u_i^2 - \sigma_i^2)(u_j^2 - \sigma_j^2) \right\}. \qquad \text{(D.7)}$$

The contribution from the first term in braces has absolute value with conditonal expectation bounded by

$$C \left\{ \sum_i \left(\sum_j \psi_{ij}^2 \right)^{1+\delta/2} \right\}^{1/(1+\delta/2)} = o_p \left(\sum_{i,j} \psi_{ij}^2 \right) \qquad \text{(D.8)}$$

as in (A.21). The contribution from the second term has, conditionally, mean zero (because $\psi_{ii} \equiv 0$) and variance bounded by

$$C \sum_{i,j} \psi_{ij}^4 \leq C \left(\max_i \sum_j |\psi_{ij}| \right)^2 \sum_{i,j} \psi_{ij}^2, \qquad \text{(D.9)}$$

to complete the proof of (D.5).

With respect to (D.6) routine development indicates that it suffices to show that each of the following expresions is $o_p \left(\Sigma_{i,j} \psi_{ij}^2 \right)$:

$$s_1 = \sum_{i,j} \psi_{ij}^2 u_i^2 u_j \hat{v}_j, \qquad s_2 = \sum_{i,j} \psi_{ij}^2 u_i^2 \hat{v}_j^2, \qquad \text{(D.10)}$$

$$s_3 = \sum_{i,j} \psi_{ij}^2 u_i u_j \hat{v}_i \hat{v}_j, \qquad s_4 = \sum_{i,j} \psi_{ij}^2 u_i \hat{v}_i \hat{v}_j^2 \qquad \text{(D.11)}$$

$$s_5 = \sum_{i,j} \psi_{ij}^2 \hat{v}_i^2 \hat{v}_j^2. \qquad \text{(D.12)}$$

We have

$$s_1 = \sum_{i \neq j} \psi_{ij}^2 u_i^2 u_j b_{ij} + \sum_{i,j} \psi_{ij}^2 u_i^2 u_j \left(\sum_{h \neq j} b_{hj} u_h \right). \qquad \text{(D.13)}$$

From previous calculations, the first term is easily seen to be $O_p \left(\left(\max_i \Sigma_j |\psi_{ij}| \right)^2 \right)$. The second term can be written

$$\sum_{i,j} \psi_{ij}^2 u_i^3 u_j b_{ij} + \sum_{i,j} \psi_{ij}^2 u_i^2 u_j \left(\sum_{h \neq i,j} b_{hj} u_h \right). \qquad \text{(D.14)}$$

The modulus of the first term has (conditional) expectation $O_p \left(\max_i \left(\Sigma_j |\psi_{ij}| \right)^2 \right)$. The second term has mean zero and variance bounded by

$$C \sum_{i,j,\ell,h} \psi_{ij}^2 \psi_{i\ell}^2 b_{hj}^2 + C \sum_{i,j,k,\ell} \psi_{ij}^2 \psi_{k\ell}^2 b_{\ell j}^2 \qquad \text{(D.15)}$$

26

$$
\begin{aligned}
&\leq\ C\sum_{i,j,\ell}\psi_{ij}^2\psi_{i\ell}^2 b_{jj} + \sum_{i,j,k,\ell}\psi_{ij}^2\psi_{k\ell}^2\left(b_{\ell\ell}+b_{jj}\right)\\
&\leq\ C\max_i\sum_\ell\psi_{i\ell}^2\sum_{i,j}\psi_{ij}^2 b_{jj} + \sum_{i,j}\psi_{ij}^2\max_\ell\sum_k\psi_{k\ell}^2\\
&\leq\ C\left(\max_i\sum_j|\psi_{ij}|\right)^4 + C\left(\max_i\sum_j|\psi_{ij}|\right)^2\left(\sum_{i,j}\psi_{ij}^2\right)\\
&=\ o\left(\left(\sum_{i,j}\psi_{ij}^2\right)^2\right).
\end{aligned}
\tag{D.16}
$$

Next,

$$
\begin{aligned}
E\,|s_2| &\leq\ C\sum_{i,j}\psi_{ij}^2\left(E\hat{v}_j^4\right)^{\frac12}\leq\sum_{i,j}\psi_{ij}^2\left\{\sum_h b_{hj}^2 + \left(\sum_h b_{hj}^4\right)^{\frac12}\right\}\\
&\leq\ C\sum_{i,j}\psi_{ij}^2 b_{jj}\leq C\left(\max_i\sum_j|\psi_{ij}|\right)^2.
\end{aligned}
\tag{D.17}
$$

The remaining terms are dealt with similarly. Indeed application of Hölder's and elementary inequalities gives the same bound for $E\,|s_3|$ as $E\,|s_2|$, while

$$
\begin{aligned}
E\,|s_4| &\leq\ C\sum_{i,j}\psi_{ij}^2\left(E\hat{v}_i^4\right)^{\frac14}\left(E\hat{v}_j^4\right)^{\frac12}\\
&\leq\ C\sum_{i,j}\psi_{ij}^2\left(\sum_i b_{hi}^2\right)^{\frac12}\left(\sum_h b_{hj}^2\right)\\
&\leq\ C\sum_{i,j}\psi_{ij}^2 b_{ii}^{\frac12}b_{jj}\\
&\leq\ C\left(\max_i\sum_j|\psi_{ij}|\right)^2,
\end{aligned}
\tag{D.18}
$$

$$
\begin{aligned}
Es_5 &\leq C \sum_{i,j} \psi_{ij}^2 \left(E\hat{v}_i^4 E\hat{v}_j^4 \right)^{\frac{1}{2}} \\
&\leq C \sum_{i,j} \psi_{ij}^2 \left(\sum_h b_{hi}^2 \right) \left(\sum_h b_{hj}^2 \right) \\
&\leq C \sum_{i,j} \psi_{ij}^2 b_{ii} b_{jj} \\
&\leq C \left(\max_i \sum_j |\psi_{ij}| \right)^2, \quad\quad\quad\quad \text{(D.19)}
\end{aligned}
$$

in both cases using $b_{ii} \leq 1$. Thus (D.6) is proved.

The proofs of (D.2) and (D.3) hardly differ from those of (A.2) and (A.3). With respect to (D.2), after replacing B_n by D_n there is no difference due to the uniform bound on relevant moments. The latter is also relevant to (D.3); we only note that in place of (A.19) we need to establish

$$
D_n^{-\frac{1}{2}} \left\{ \sum_i \left(\sum_{j<i} \psi_{ij} u_j \right) \left(\sum_{j<i} \psi_{ij} u_j \right)' - \sum_{j<i} \psi_{ij} \psi_{ij}' \sigma_j^2 \right\} D_n^{-\frac{1}{2}} \to_p 0, \quad \text{(D.20)}
$$

and the details differ only trivially. $\square$

APPENDIX E: PROOF OF THEOREM E

The limiting distribution of ζ_n follows from the properties

$$
\sigma^{-2} A_n^{-\frac{1}{2}} a_n \to_p N\left(R^{-\frac{1}{2}} T\nu, I_p \right) \quad\quad\quad\quad \text{(E.1)}
$$

and (A.1), (A.2). Of course (E.1) implies $\sigma^{-4} a_n' A_n^{-1} a_n \to_d \chi_p'^2 \left(\nu T' R^{-1} T\nu \right)$. We focus only on the proof of (E.1) since given this the proof of (A.2) straightforwardly extends that under H_0 in Appendix A, and again essentially implies that of (A.1) (cf. Robinson, 1994).

The h-th element of $D_n^{-\frac{1}{2}} a_n$ is

$$
d_{hn}^{-\frac{1}{2}} a_n' \Psi_{hn} u_n = d_{hn}^{-\frac{1}{2}} \varepsilon_n' \Omega_n^{\frac{1}{2}} (\theta_{(n)}) \Psi_{hn} \Omega_n^{\frac{1}{2}} (\theta_{(n)}) \varepsilon_n. \quad\quad \text{(E.2)}
$$

Write $\Lambda_{in}(\theta) = (\partial/\partial\theta_i) \Omega_n^{\frac{1}{2}}(\theta)$ and

$$
\begin{aligned}
\Xi_{hin}(\theta) &= \frac{\partial}{\partial\theta_i} \left\{ \Omega_n^{\frac{1}{2}}(\theta) \Psi_{hn} \Omega_n^{\frac{1}{2}}(\theta) \right\} \\
&= \Lambda_{in}(\theta) \Psi_{hn} \Omega_n^{\frac{1}{2}}(\theta) + \Omega_n^{\frac{1}{2}}(\theta) \Psi_{hn} \Lambda_{in}(\theta). \quad\quad \text{(E.3)}
\end{aligned}
$$

By the mean value theorem, (E.2) is

$$d_{hn}^{-\frac{1}{2}} \varepsilon_n' \Psi_{hn} \varepsilon_n + d_{hn}^{-\frac{1}{2}} \sum_{i=1}^{q} \varepsilon_n' \Xi_{hin} \left(\tilde{\theta}_{(n)} \right) \varepsilon_n \nu_i / d_{in}^{\frac{1}{2}}, \tag{E.4}$$

where $\left\| \tilde{\theta}_{(n)} \right\| \leq \left\| \theta_{(n)} \right\|$ and ν_i is the i-th element of ν. The second term on the right of (E.4) has, conditional on Ψ_{hn}, mean

$$\sigma^2 d_{hn}^{-\frac{1}{2}} \sum_{i=1}^{p} tr \left\{ \Xi_{hin}(\tilde{\theta}_{(n)}) \right\} \nu_i / d_{in}^{\frac{1}{2}}, \tag{E.5}$$

and variance bounded by

$$C d_{hn}^{-1} \sum_{i=1}^{p} tr \left\{ \Xi_{hin}(\tilde{\theta}_{(n)}) \right\}^2 / d_{in}. \tag{E.6}$$

We examine first (E.5). Since

$$\Phi_{in}(\theta) = \Lambda_{in}(\theta) \Omega_n^{\frac{1}{2}}(\theta) + \Omega_n^{\frac{1}{2}}(\theta) \Lambda_{in}(\theta), \tag{E.7}$$

we have

$$vec \left\{ \Lambda_{in}(\theta) \right\} = \left\{ \Omega_n^{\frac{1}{2}}(\theta) \otimes I_n + I_n \otimes \Omega_n^{\frac{1}{2}}(\theta) \right\}^{-1} vec \Phi_{in}(\theta). \tag{E.8}$$

Thus

$$\Lambda_{in}(0) = \frac{1}{2} \Phi_{in}, \tag{E.9}$$

$$\Xi_{hin}(\theta) = \frac{1}{2} \left\{ \Phi_{in} \Psi_{hn} + \Psi_{hn} \Phi_{in} \right\}, \tag{E.10}$$

and so

$$\begin{aligned} tr \left\{ \Xi_{hin}(\theta) - \Xi_{hin}(0) \right\} &= tr \left\{ (\Lambda_{in}(\theta) - \Lambda_{in}(0)) \Psi_{hn} \Omega_n^{\frac{1}{2}}(\theta) \right\} \\ &\quad + tr \left\{ \Lambda_{in}(0) \Psi_{hn} \left(\Omega_n^{\frac{1}{2}}(\theta) - I_n \right) \right\}. \end{aligned} \tag{E.11}$$

The first term on the right is bounded in absolute value by

$$tr^{\frac{1}{2}} \left\{ \Psi_{hn}^2 \right\} tr^{\frac{1}{2}} \left\{ \left(\Omega_n^{\frac{1}{2}}(\theta) \left(\Lambda_{in}(\theta) - \Lambda_{in}(0) \right) \right)^2 \right\}. \tag{E.12}$$

The second factor in (E.12) is bounded by

$$\bar{\lambda}^{\frac{1}{2}} \left(\Omega_n(\theta) \right) tr^{\frac{1}{2}} \left\{ (\Lambda_{in}(\theta) - \Lambda_{in}(0))^2 \right\}. \tag{E.13}$$

29

From (E.8) and (E.9)

$$\Lambda_{in}(\theta) - \Lambda_{in}(0) = \left[\left\{ \Omega_n^{\frac{1}{2}}(\theta) \otimes I_n + I_n \otimes \Omega_n^{\frac{1}{2}}(\theta) \right\}^{-1} - \frac{1}{2} I_{n^2} \right] vec \Phi_{in}(\theta)$$
$$+ \frac{1}{2} vec \left\{ \Phi_{in}(\theta) - \Phi_{in} \right\}. \tag{E.14}$$

The factor in square brackets is

$$\frac{1}{2} \left\{ \Omega_n^{\frac{1}{2}}(\theta) \otimes I_n + I_n \otimes \Omega_n^{\frac{1}{2}}(\theta) \right\}^{-1} \left[\left\{ \left(I_n - \Omega_n^{\frac{1}{2}}(\theta) \right) \otimes I_n \right\} \times \left\{ I_n \otimes \left(I_n - \Omega_n^{\frac{1}{2}}(\theta) \right) \right\} \right]. \tag{E.15}$$

The largest eigenvalue of the product of (E.15) with its transpose is bounded by a constant times

$$c_n(\theta) = \bar{\lambda}^2 \left(\Omega_n^{-1}(\theta) \right) \bar{\lambda}^2 \left(\left(\Omega_n^{\frac{1}{2}}(\theta) - I_n \right)^2 \right). \tag{E.16}$$

Thus

$$tr \left\{ (\Lambda_{in}(\theta) - \Lambda_{in}(0))^2 \right\} \le C \left[c_n(\theta) tr \left\{ \Phi_{in}(\theta)^2 \right\} + tr \left\{ (\Phi_{in}(\theta) - \Phi_{in})^2 \right\} \right]. \tag{E.17}$$

The second term on the right of (E.11) is likewise bounded in absolute value by

$$tr^{\frac{1}{2}} \left\{ \Psi_{hn}^2 \right\} tr^{\frac{1}{2}} \left\{ \Phi_{in}^2 \right\} \bar{\lambda}^{\frac{1}{2}} \left(\left(\Omega_n^2(\theta) - I_n \right)^2 \right). \tag{E.18}$$

It follows from the above bounds and Assumption 5 that $tr \left\{ \Xi_{hin}(\theta_{(n)}) - \Xi_{hin}(0) \right\}$
$= o_p \left(d_{hn}^{\frac{1}{2}} d_{in}^{\frac{1}{2}} \right)$. On the other hand, from Assumption 5,

$$d_{hn}^{-\frac{1}{2}} tr \left\{ \Xi_{hin}(0) \right\} d_{in}^{-\frac{1}{2}} \to_p \tau_{hi}, \tag{E.19}$$

where τ_{hi} is the (h, i)-th element of T.
Considering now (E.6), we have

$$tr \left\{ \Xi_{hin}(\theta) \right\}^2 2 \le 2 tr \left\{ \Xi_{hin}(0)^2 \right\} + 2 tr \left\{ (\Xi_{hin}(\theta) - \Xi_{hin}(0))^2 \right\}. \tag{E.20}$$

The first term on the right is bounded by

$$2 tr \left\{ (\Psi_{hn} \Phi_{in})^2 \right\} = o_p (d_{hn} d_{in}) \tag{E.21}$$

which is proved like in (C.1), (C.2). The second term is bounded by
$tr^2 \left\{ \Xi_{hin}(\theta) - \Xi_{hin}(0) \right\}$. Thus from calculations above, (E.6) $= o_p(1)$. It follows that the second term in (E.4) converges in probability to the h-th element of $T\nu$. Given the proof of Theorem A, (E.1) readily follows. The final statement of the Theorem is a familiar consequence. $\square$

Acknowledgement

References

Anselin, L., 2001. Rao's score test in spatial econometrics. Journal of Statistical Planning and Inference 97, 113-139.

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

Baltagi, B., Dong Li, 2001. LM tests for functional form and spatial error correlation. International Regional Science Review 24, 194-225.

Box, G.E.P., Jenkins, G.M., 1970. Time Series. Forecasting and Control. (Holden-Day, San Francisco).

Box, G.E.P., Pierce, D.B., 1970. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. Journal of the American Statistical Association 65, 1509-1526.

Cliff, A., Ord, J.K., 1968. The problem of spatial autocorrelation. Joint Discussion Paper, Universty of Bristol: Department of Ecnomics No. 26, Department of Geography, Series A, No. 15.

Cliff, A., Ord, J.K., 1972. Testing for spatial autocorrelation among regression residuals. Geographical Analysis 4, 267-284.

Cressie, N., 1993. Statistics for Spatial Data. (Wiley, New York).

Dahlhaus, R., Künsch, H., 1987. Edge effects and efficient parameter estimation for stationary random fields. Biometrika 74, 877-882.

Dunsmuir, W., 1983. A central limit theorem for estimation in Gaussian stationary time series observed at unequally spaced lines. Stochastic Processes and Their Applications 14, 279-295.

Dunsmuir, W., Robinson, P.M., 1981. Parametric estimates for stationary time series with missing observations. Advances in Applied Probability 13, 129-146.

Eicker, F., 1963. Asymptotic normality and consistency of the least squares estimator for families of linear regressions. Annals of Mathematical Statistics 34, 447-456.

Fox, R., Taqqu, M.S., 1986. Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. Annals of Statistics 14, 517-532,

Godfrey, L.G., 1978. Testing against autoregressive and moving average error models when the regressors include lagged dependent variables. Econometrica 46, 1293-1301.

Guyon, X., 1982. Parameter estimation for a stationary process on a d-dimensional lattice. Biometrika 69, 95-106.

Hannan, E.J., 1970. Multiple Time Series. (John Wiley, New York).

Hannan, E.J., 1973. The asymptotic theory of linear time series models. Journal of Applied Probability 10, 130-145.

Kelejian, H., Prucha, I., 2001. On the asymptotic distribution of the Moran I test statistic with applications. Journal of Econometrics 104, 219-257.

Kelejian, H., Robinson, D.P., 2004. The influence of spatially correlated heteroskedasticity on tests for spatial correlation. In Anselin, L., Florax, R.G.M., Rey, S. (eds.). Advances in Spatial Econometrics: Methodology, Tools and Applications (Springer-Verlag, Berlin), pp.26-97.

Lahiri, S.N., 1996. On inconsistency of estimates based on spatial data under infill asymptotics. Sankhya A 58, 403-417.

Lee, L.F., 2002. Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models. Econometric Theory 18, 252-277.

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

Ljung, G.M., Box, G.E.P., 19T78. On a measure of lack of fit in tme series models. Biometrika 65, 297-303.

McDunnough, P., Wolfson, D.B., 1979. On some sampling schemes for estimating the parameters of a continuous time series. Annals of the Institute of Statistical Mathematics 31, 487-497.

Moran, P.A.P., 1950. A test for the serial dependence of residuals. Biometrika 37, 178-181.

Parzen, E., 1963. On spectral analysis with missing observations and amplitude modulation. Sankhya A 25, 383-392.

Pinkse, J. 1999. Asymptotic properties of Moran and related tests and testing for spatial correlation in probit models. Preprint, University of British Columbia.

Pinkse, J., 2004. Moran-flavoured tests with nuisance parametrics: examples. In Anselin, L., Florax, R.G.M., Rey, S. (eds.). Advances in Spatial Econometrics: Methodology, Tools and Applications (Springer-Verlag, Berlin), pp.67-77.

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., 1986. Testing for serial correlations in regresssion with missing observations. Journal of the Royal Statistical Society B 47, 426-437.

Robinson, P.M., 1991. Testing for strong serial correlation and dynamic conditional heteroskedasticity in multiple regression. Journal of Econometrics 47, 67-84.

Robinson, P.M., 1994. Efficient tests of nonstationary hypotheses. Journal of the American Statistical Association 89, 1420-1437.

Robinson, P.M., 2006. Efficient estimation of the semiparametric spatial autoregressive model. Journal of Econometrics, forthcoming.

Robinson, P.M., Vidal Sanz, J., 2006. Modified Whittle estimation of multilateral models on a lattice. Journal of Multivariate Analysis 97, 1090-1120.

Sargan, J.D., Drettakis, E.G., 1974. Missing data in an autoregressive model. International Economic Review 15, 39-59.

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

Sen, A., 1976. Large sample-size distribution of statistics used in testing for spatial correlation. Geographical Analysis 9, 175-184.

Stein, M.L., 1991. Fixed-domain asymptotics for spatial periodograms. Journal of the American Statistical Association 90, 1277-1288.

von Bahr, B., Esseen, C.-G., 1965. Inequalities for the rth absolute moment of a sum of random variables, $1 \leqq r \leqq 2$. The Annals of Mathematical Statistics 36, 299-303.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 41, 434-449.