Finite Sample Corrections for the Lagrange Multiplier Test in Spatial Autoregressive Models

Peter M. Robinson and Francesca Rossi

December 10, 2010

Abstract

Lagrange multiplier tests of spatial uncorrelatedness in a pure spatial autoregressive model have advantages over other forms of testing. They are typically based on the (χ^2) first-order asymptotic approximation to the distribution of the test statistic. In small samples this approximation may be poor. We develop refined tests based on Edgeworth expansion. These are compared by Monte Carlo simulations to ones that are respectively based on a bootstrap, and on the exact finite sample distribution. Generally such tests are found to significantly outperform those based on the χ^2 approximation. We also develop Edgeworth-based tests for uncorrelatedness of disturbances in a regression model, against the alternative of spatial autoregressive disturbances.
JEL classifications: C29
Keywords: Spatial autocorrelation; Lagrange multiplier test; Edgeworth expansion; bootstrap; finite-sample corrections.

1 Introduction

The spatial autoregressive (SAR) model is a parsimonious method of describing spatial dependence, conveniently depending only on economic distances rather than actual locations, which may be unknown or irrelevant. It thus provides a convenient, widely-usable class of alternatives in testing the null hypothesis of spatial uncorrelatedness which, if true, considerably simplifies statistical inference. Lagrange multiplier (LM) testing is especially computationally convenient because it depends on the null model, and thus does not require estimating the spatial autoregressive coefficient. An LM test can be expected to be efficient against local SAR alternatives, and to have an asymptotic null χ^2 distribution under the null hypothesis. However, the χ^2 approximation may not be accurate in modest samples, so a test based on it may be badly sized. Thus we develop tests with improved finite-sample properties.

The SAR model is given by

$$Y_n = \lambda W_n Y_n + \epsilon_n, \tag{1.1}$$

where Y_n is a $n \times 1$ vector of observations, ϵ_n is a $n \times 1$ vector of unobservable, mutually independent, random variables, with zero mean and finite variance, λ is a scalar, and W_n is a given $n \times n$ "weight" matrix. In the sequel, we drop the n subscript, writing $\epsilon = \epsilon_n$, $Y = Y_n$, $W = W_n$, with the same convention for other $n-$dependent quantities. In particular, W has zero diagonal elements, and typically satisfies some normalization restrictions (which aid identification of λ). Generally when $\lambda \neq 0$, (1.1) implies spatial correlation among elements of Y. However, when $\lambda = 0$ these elements are mutually independent. We thus consider testing the null hypothesis

$$H_0 : \lambda = 0 \tag{1.2}$$

in (1.1). Various such tests have already been discussed in the literature (see e.g. Moran (1950), Cliff and Ord (1972), Burridge (1980), Pinkse (2004)).

Superficial inspection of (1.1) suggests a Wald test based on the ordinary least squares (OLS) estimate of λ. However, unlike in the case of stationary time series autoregression, the dependence in general of each element of Y on each element of ϵ (rather than on just the "present and past" ones) produces inconsistency (Lee (2002)). On the other hand when $\lambda = 0$ we have $Y = \epsilon$ and, under regularity conditions, the OLS estimate converges to zero, and thus could be used to test (1.2). But because it converges as $n \to \infty$ to a biased probability limit when $\lambda \neq 0$, there are questions about power, and the classical asymptotic local and non-local power properties of Wald tests will not apply. A second familiar course of action is to consider the maximum likelihood (ML) estimate, assuming also that the elements of ϵ are identically distributed normal variables. Lee (2004) showed that under regularity conditions the ML estimate of λ is consistent and asymptotically normal (for any λ), and thus can be used in Wald testing. The drawback here is computational, as the ML estimate is not defined in closed form, and this affects also likelihood-ratio testing. The other class commonly linked with Wald and likelihood-ratio tests is LM testing, based on a normal likelihood, and as usual this produces a relatively simple closed-form statistic for testing (1.2) (or indeed for testing any value of λ).

With regularity conditions, the LM statistic has a null limiting χ^2 distribution, as $n \to \infty$. Frequently, however, spatial economic data sets are not very large, and there is a question about the accuracy of this asymptotic approximation to the distribution. This is of particular concern in the SAR setting because (as our results indicate), convergence to the χ^2 limit distribution can be slower than the classical parametric rate. Thus we consider tests that are based on the LM statistic but that potentially have better size properties in finite samples. The main contribution of the paper is to develop tests based on the Edgeworth expansion of the distribution function of the LM statistic. This is the focus of the following section. We next provide corresponding tests of (1.2) in linear regression models, with SAR disturbances. In both cases theorem proofs are left to an Appendix. In Section 4 we specify the finite sample corrections of Robinson (2008), so that the finite sample performance of the latter can be compared with that of the Edgeworth-corrected tests. As is well known, a bootstrap can achieve an Edgeworth correction, in Section 5 we compare bootstrap-based tests with the Edgeworth ones in a Monte Carlo study of finite sample. Section 6 compares the Edgeworth approximation with the the exact distribution of the LM statistic.

2 Edgeworth-corrected LM tests for pure SAR

The LM statistic for testing (1.2) against (1.1) with the alternative hypothesis

$$H_1 : \lambda \neq 0, \tag{2.1}$$

is

$$LM = \frac{n^2}{tr(W^2 + WW')} \left(\frac{Y'WY}{Y'Y} \right)^2, \tag{2.2}$$

This statistic was derived by Burridge (1980) who noted that it is equivalent to the test statistic of Cliff and Ord (1972), which in turn is related to a statistic of Moran (1950); see also Anselin(1988, 2001) for extensions to more general models, and Pinkse (2004). As noted by Burridge (1980), (2.2) is also the LM statistic for testing (1.2) against the spatial moving average model

$$Y = \epsilon + \lambda W \epsilon$$

(a corresponding equivalence to that found with time series models).

The derivation of (2.2) is based on a Gaussian likelihood but as is common the same first order limit distribution obtains more generally. Under suitable conditions we have

$$P(LM \leq \eta) = F(\eta) + o(1) \tag{2.3}$$

for any $\eta > 0$, where F denotes the distribution function (df) of a χ_1^2 random variable. Thus (1.2) is rejected in favour of (2.1) if LM exceeds the appropriate percentile of the χ_1^2 distribution. We can likewise test (1.2) against a one-sided alternative, $\lambda > 0$ or $\lambda < 0$, by comparing $\sqrt{LM}$ with the appropriate upper or lower percentiles of the standard normal distribution. However, except in Section 6, we focus throughout on the two-sided tests.

We do not describe sufficient conditions for (2.3), because we wish to consider statistics with better finite-sample properties and we can only justify these under the precise distributional assumption.

Assumption 1 *The elements of ϵ are independent and identically distributed normal random variables with mean zero and unknown variance σ^2.*

We denote by $w_{ij} = w_{ij}$ the (i, j)-th element of W, and introduce

Assumption 2

(i) *For all n, $w_{ii} = 0$, and $\Sigma_{j=1}^n w_{ij} = 1$, $i = 1, ..., n$;*

(ii) *For all* n,

$$\max_j \sum_{i=1}^{n} |w_{ij}| + \max_i \sum_{j=1}^{n} |w_{ij}| \leq K,$$

where K *is a finite generic constant;*

(iii) *Uniformly in* $i, j = 1, ..., n$, $w_{ij} = O(1/h)$, *where* $h = h_n$ *is bounded away from zero for all* n *and* $h/n \to 0$ *as* $n \to \infty$;

(iv) *The limits*

$$\lim_{n\to\infty} \frac{h}{n} tr(W^i),\ i = 2, 3, 4;\ \lim_{n\to\infty} \frac{h}{n} tr\,(W'W) \tag{2.4}$$

are non-zero.

The normalization in (i) is not strictly necessary ((2.2) is invariant to multiplication of W by any scalar), but it (and the restriction $|\lambda| < 1$) plays a role in constructing the likelihood, and it or some other normalization is commonly employed in practice. If W is symmetric with non-negative elements, (i) implies (ii). The sequence h in (iii) and (iv) can be bounded or divergent; a condition governing the behaviour of the w_{ij} is commonly required in asymptotic theory for statistics based on (1.1), and subsequently we discuss its implications in relation to the particular W employed in the simulations there. The limits (2.4) exist and are finite by Lemma 1 of Appendix C, so (iv) just requires them not to vanish.

Throughout, the notation $a \sim b$ means that a/b converges to a positive, finite limit. Moreover, f denotes the χ^2 probability density function (pdf).

Theorem 1 *Under (1.2) and Assumptions 1 and 2, the df of* LM *admits the formal Edgeworth expansion*

$$Pr(LM \leq \eta) = F(\eta) + \frac{\kappa}{4}\eta f(\eta) - \frac{\kappa}{12}\eta^2 f(\eta) + o\left(\frac{h}{n}\right) \tag{2.5}$$

in case h *is divergent, and*

$$Pr(LM \leq \eta) = F(\eta) + \frac{\kappa}{4}\eta f(\eta) - \frac{\kappa}{12}\eta^2 f(\eta) - \frac{2}{n}\eta^2 f(\eta) + o\left(\frac{1}{n}\right) \tag{2.6}$$

when h *is bounded, where*

$$\kappa = \frac{3tr(W' + W)^4}{a^2} \sim \frac{h}{n} \tag{2.7}$$

and $a = tr(W^2 + W'W)$.

The proof of Theorem 1 is in Appendix A. It must be stressed that both expansions in Theorem 1 are formal. It is beyond the scope of the paper to establish validity .

Clearly, (2.5) and (2.6) entail better approximations than (2.3). The leading terms in (2.5) and (2.6) depend on known quantities, so they can be used directly for approximating the df. The two outcomes in Theorem 1 create a dilemma for the practitioner because it cannot be determined from a finite data set whether to treat h as divergent or bounded. However, (2.6) is justified also when h is divergent because the extra term in the expansion, $-2\eta^2 f(\eta)/n$, is $o(h/n)$.

Theorem 1 can be used to derive Edgeworth-corrected critical values. Let w_α and z_α be the α-quantile of LM and a standard normal variate, respectively. By inverting either expansion, we can expand w_α as an infinite series

$$w_\alpha = z^2_{(1+\alpha)/2} + p_1(z^2_{(1+\alpha)/2}) + ..., \tag{2.8}$$

where $p_1(z^2_{(1+\alpha)/2})$ is a polynomial whose coefficients have order h/n, and that can be determined using the identity $\alpha = Pr(LM \leq w_\alpha)$ and the expansions given in Theorem 1. Specifically, when h is divergent, we have

$$\alpha = Pr(LM \leq w_\alpha) = F(w_\alpha) + \left(\frac{\kappa}{4} w_\alpha - \frac{\kappa}{12} w_\alpha^2\right) f(w_\alpha) + o\left(\frac{h}{n}\right).$$

By substituting (2.8), the leading terms of the LHS are

$$\begin{aligned} & F(z^2_{(1+\alpha)/2}) + p_1(z^2_{(1+\alpha)/2}) f(z^2_{(1+\alpha)/2}) \\ + & \left(\frac{\kappa}{4} z^2_{(1+\alpha)/2} - \frac{\kappa}{12} z^4_{(1+\alpha)/2}\right) f(z^2_{(1+\alpha)/2}) + o\left(\frac{h}{n}\right) \\ = & \alpha + p_1(z^2_{(1+\alpha)/2}) f(z^2_{(1+\alpha)/2}) \\ + & \left(\frac{\kappa}{4} z^2_{(1+\alpha)/2} - \frac{\kappa}{12} z^4_{(1+\alpha)/2}\right) f(z^2_{(1+\alpha)/2}) + o\left(\frac{h}{n}\right). \end{aligned}$$

The latter is $\alpha + o(h/n)$ (rather than $\alpha + O(h/n)$), when we take

$$p_1(x) = -\left(\frac{\kappa}{4} x - \frac{\kappa}{12} x^2\right) \sim \frac{h}{n}. \tag{2.9}$$

Similarly, when h is bounded, we take

$$p_1(x) = -\left(\frac{\kappa}{4} x - \frac{\kappa}{12} x^2 - \frac{2}{n} x^2\right) \sim \frac{1}{n}. \tag{2.10}$$

If w_α were known, the size of a test of H_0 in (1.2) would obviously be $Pr(LM > w_\alpha | H_0) = 1 - \alpha$. We can compare the size of the test of H_0 in (1.2) based on the usual first order approximation, i.e.

$$Pr(LM > z^2_{(\alpha+1)/2} | H_0) \tag{2.11}$$

with

$$Pr(LM > z^2_{(\alpha+1)/2} + p_1(z^2_{(\alpha+1)/2}) | H_0), \tag{2.12}$$

where $p_1(.)$ is defined according to (2.9) if h is divergent and (2.10) if h is bounded.

Thus, the error of the approximation of (2.11) is $O(h/n)$, while that of (2.12) is $o(h/n)$ when the sequence h is divergent, or $o(1/n)$ when it is bounded.

As an alternative to using corrected critical values, we can also apply Theorem 1 to construct a transformation of LM whose distribution better approximates χ^2 than LM itself. Starting from the expansion in (2.5), we consider the cubic transformation

$$g(x) = x + \frac{\kappa}{4}x - \frac{\kappa}{12}x^2 + \frac{1}{4}Q(x), \quad Q(x) = \left(\frac{\kappa}{4}\right)^2 \left(\frac{4}{27}x^3 - \frac{2}{3}x^2 + x\right), \quad (2.13)$$

such that

$$Pr(g(LM) \leq \eta) = F(\eta) + o\left(\frac{h}{n}\right).$$

Similarly, from (2.6), we can write

$$\begin{aligned} g(x) &= x + \frac{\kappa}{4}x - \frac{\kappa}{12}x^2 - \frac{2}{n}x^2 + \frac{1}{4}Q(x), \\ Q(x) &= \left(\frac{\kappa}{4}\right)^2 x + \frac{1}{3}\left(\frac{\kappa}{6} + \frac{4}{n}\right)^2 x^3 - \frac{\kappa}{4}\left(\frac{\kappa}{6} + \frac{4}{n}\right) x^2, \end{aligned} \quad (2.14)$$

such that

$$Pr(g(LM) \leq \eta) = F(\eta) + o\left(\frac{1}{n}\right).$$

The transformations (2.13) and (2.14) were proposed in case of a standard normal limiting distribution by Hall (1992), or, in a slightly more general setting, Yanagihara et al. (2005). In Lemma 4 (reported in Appendix C) we show that such result extends to χ^2 limiting distributions.

Therefore, we can compare

$$Pr(g(LM) > z^2_{(\alpha+1)/2} | H_0), \quad (2.15)$$

where $g(.)$ is defined according to (2.13) or (2.14) depending on h, with (2.11). Again, (2.15) has error $o(h/n)$ compared to the $O(h/n)$ error of (2.11).

3 Improved LM tests in regressions where the disturbances are spatially correlated

We can extend the results obtained in Section 2 to the more general model

$$Y = X\beta + u, \qquad u = \lambda W u + \epsilon, \quad (3.1)$$

where X is an $n \times k$ matrix of nonstochastic regressors and β is a $k \times 1$ vector of unknown parameters. From Burridge (1980), Anselin (1988, 2001), the Lagrange multiplier statistic for testing (1.2) is

$$\tilde{LM} = \frac{n^2}{tr(W'W) + tr(W^2)} \left(\frac{\hat{u}'W\hat{u}}{\hat{u}'\hat{u}}\right)^2 = \frac{n^2}{a}\left(\frac{Y'PWPY}{Y'PY}\right)^2, \quad (3.2)$$

where

$$P = I - X(X'X)^{-1}X'. \tag{3.3}$$

We impose the following condition on X:

Assumption 3 *For all n, each element x_{ij} of X is predetermined and uniformly bounded in absolute value. Moreover, the smallest eigenvalue of $X'X/n$ are bounded away from zero for all sufficiently large n and the limits of at least one component of $X'WX/n$, $X'W^2X/n$ and $X'W'WX/n$ are non zero.*

We have the following results:

Theorem 2 *Under (1.2) and Assumptions 1-3, the df of $\tilde{LM}$ admits the formal Edgeworth expansion*

$$Pr(\tilde{LM} \leq \eta) = F(\eta) + \left(\frac{\kappa}{4}\eta - \frac{\kappa}{12}\eta^2 + 2\omega_1\eta\right) f(\eta) + o\left(\frac{h}{n}\right) \tag{3.4}$$

with

$$\omega_1 = \frac{tr(K_3 - K_2)}{a} - \frac{1}{2}\frac{(tr(K_1))^2}{a} \sim \frac{h}{n} \tag{3.5}$$

if h is divergent, and

$$Pr(\tilde{LM} \leq \eta) = F(\eta) + \left(\frac{\kappa}{4}\eta - \frac{\kappa}{12}\eta^2 + 2\omega_2\eta - \frac{2}{n}\eta^2\right) f(\eta) + o\left(\frac{1}{n}\right) \tag{3.6}$$

with

$$\omega_2 = \frac{tr(K_3 - K_2)}{a} - \frac{1}{2}\frac{(tr(K_1))^2}{a} - \frac{k}{n} \sim \frac{1}{n} \tag{3.7}$$

if h is bounded, where κ is given in (2.7),

$$K_1 = (X'X)^{-1}X'WX, \tag{3.8}$$

$$K_2 = \frac{1}{2}X'(W + W')X(X'X)^{-1}X'(W' + W)X(X'X)^{-1} \tag{3.9}$$

and

$$K_3 = X'(W + W')^2X(X'X)^{-1}. \tag{3.10}$$

The components of $(X'X)^{-1}$ have order $1/n$ by Assumption 3. On the other hand, the components of $X'WX$, $X'(W+W')X$ and $X'(W+W')^2X$ are $O(n)$ by Lemma 2. Assumption 3 imposes that for at least one component of each matrix the latter holds as an exact rate. It follows that $tr(K_1)$, $tr(K_2)$ and $tr(K_3)$ are bounded and non zero. Hence ω_1 and ω_2 have exactly order h/n and $1/n$, respectively.

The proof of Theorem 2 is in Appendix B. Again, both the expansions are formal.

From (3.4) and (3.6), we can obtain Edgeworth-corrected critical values. Much as in Section 2 we can obtain improved critical values. The size based on χ^2 critical value is

$$Pr(\tilde{LM} > z^2_{(\alpha+1)/2}|H_0) \tag{3.11}$$

while the Edgeworth-corrected critical value is

$$Pr(\tilde{LM} > z^2_{(\alpha+1)/2} + \tilde{p}_1(z^2_{(\alpha+1)/2})|H_0), \tag{3.12}$$

where

$$\tilde{p}_1(z^2_{(\alpha+1)/2}) = -\left(\frac{\kappa}{4}z^2_{(\alpha+1)/2} - \frac{\kappa}{12}z^4_{(\alpha+1)/2} + 2\omega_1 z^2_{(\alpha+1)/2}\right)$$

if h is divergent and

$$\tilde{p}_1(z_{(\alpha+1)/2}) = -\left(\frac{\kappa}{4}z^2_{(\alpha+1)/2} - \frac{\kappa}{12}z^4_{(\alpha+1)/2} + 2\omega_2 z^2_{(\alpha+1)/2} - \frac{2}{n}z^4_{(\alpha+1)/2}\right)$$

if h is bounded. As before, (3.11) has error of order h/n, while (3.12) has error $o(h/n)$.

As in Section 2, we can also consider Edgeworth-corrected test statistics. The size of test of H_0 in (1.2) based on the standard Lagrange multiplier statistic, as given in (3.11) is compared with that based on a corrected statistic, i.e.

$$Pr(\tilde{g}(\tilde{LM}) > z^2_{(\alpha+1)/2}|H_0). \tag{3.13}$$

The choice of the function $\tilde{g}$ is motivated by Lemma 4 and in this case is given by

$$\tilde{g}(x) = x + \frac{\kappa}{4}x - \frac{\kappa}{12}x^2 + 2\omega_1 x + \frac{1}{4}Q(x),$$

where

$$Q(x) = \left(\left(\frac{\kappa}{4}\right)^2 + 4\omega_1^2 + \kappa\omega_1\right)x - \frac{1}{2}\left(\frac{2}{3}\kappa\omega_1 + \frac{\kappa^2}{12}\right)x^2 + \frac{1}{3}\left(\frac{\kappa}{6}\right)^2 x^3$$

in case h is divergent and

$$\tilde{g}(x) = x + \frac{\kappa}{4}x - \frac{\kappa}{12}x^2 + 2\omega_2 x - \frac{2}{n}x^2 + \frac{1}{4}Q(x),$$

with

$$Q(x) = \left(\frac{\kappa}{4} + 2\omega_2\right)^2 x - \left(\frac{\kappa}{4} + 2\omega_2\right)\left(\frac{\kappa}{6} + \frac{4}{n}\right)x^2 + \frac{1}{3}\left(\frac{\kappa}{6} + \frac{4}{n}\right)^2 x^3$$

if h is bounded. Similarly to the case of model (1.1), when the standard Lagrange multiplier statistic is used the error of the approximation has order h/n while it is reduced to $o(h/n)$ when the test is based on the Edgeworth-corrected test statistic.

4 Alternative correction

The results derived in Sections 2 and 3 can be compared with two alternative corrections derived for asymptotically χ^2 statistics in Robinson (2008). In particular, Robinson (2008) proposes both mean-adjusted and mean and variance-adjusted variants of (2.2) and (3.2), which prove to be asymptotically distributed as a χ^2 random variable with one degree of freedom. Such corrected statistics are expected to have better finite sample properties than either (2.2) or (3.2), even though the magnitude of the gain in accuracy is not explicitly shown. In finite sample the corrected statistic based on mean adjustment might have a larger variance than the non-corrected version, resulting in a partial (or total) offset of the gain in accuracy from the mean standardisation. In such case, a mean and variance standardisation should be performed instead.

It should be stressed that such corrected statistics might be convenient in the present case since the ratios $\epsilon' W\epsilon/\epsilon'\epsilon$ and $\epsilon' PWP\epsilon/\epsilon' P\epsilon$ are independent of their own denominator and therefore the expectation of the ratio is equal to the ratio of expectations (Pitman (1937)). If the latter condition failed, a correction based on mean and variance standardisation would not be convenient (or even impossible), since the evaluation of mean and variance would require some approximation.

We suppose that (1.2) and Assumptions 1-3 hold and focus on the simpler case first, i.e. the statistic given in (2.2). Specifically, Robinson (2008) proposes a mean and variance-adjusted null statistic as

$$\left(\frac{2}{Var(LM)}\right)^{1/2}(LM - E(LM)) + 1, \tag{4.1}$$

where $Var(LM)$ denotes the variance of LM. In order to compare the performance of such corrected statistics with that based on the results presented in Section 2, the leading terms of (4.1) have to be derived.

As presented in Robinson (2008),

$$E(LM) = \left(1 + \frac{2}{n}\right)^{-1}, \tag{4.2}$$

while

$$Var(LM) = \frac{n^4}{a^2}\frac{E(\epsilon' W\epsilon)^4}{E(\epsilon'\epsilon)^4} - \left(1 + \frac{2}{n}\right)^{-2}.$$

By some standard formulae on expectations of quadratic forms in normal random variables (see e.g. Ghazal (1996)), we have

$$\begin{aligned} Var(LM) &= \frac{n^2}{a^2}\frac{3a^2 + 3tr((W+W')^4)}{n^4\left(1 + \frac{12}{n} + \frac{44}{n^2} + \frac{48}{n^3}\right)} - \left(1 + \frac{2}{n}\right)^{-2} \\ &= 2 + \frac{3tr((W+W')^4)}{a^2} - \frac{32}{n} + o\left(\frac{1}{n}\right), \end{aligned} \tag{4.3}$$

where the second equality follows by standard Taylor expansion.

Collecting (4.2) and (4.3), (4.1) becomes

$$
\begin{aligned}
& \left(1+\frac{3tr((W+W')^4)}{2a^2}-\frac{16}{n}+o\left(\frac{1}{n}\right)\right)^{-1/2}\left(LM-\left(1+\frac{2}{n}\right)^{-1}\right)+1 \\
= & \left(1-\frac{3}{4}\frac{tr((W+W')^4)}{a^2}+\frac{8}{n}+o\left(\frac{1}{n}\right)\right)\left(LM-1+\frac{2}{n}+o\left(\frac{1}{n}\right)\right)+1,
\end{aligned}
$$

where the second equality follows by Taylor expansion. Hence, when h is divergent, we define

$$
L\bar{M} = LM - \frac{3}{4}\frac{tr((W+W')^4)}{a^2}(LM-1) \tag{4.4}
$$

while

$$
L\bar{M} = LM - \frac{3}{4}\frac{tr((W+W')^4)}{a^2}(LM-1)+\frac{8}{n}LM-\frac{6}{n} \tag{4.5}
$$

when h is bounded.

For both divergent and bounded h, we consider the size of the test of (1.2) based on $L\bar{M}$, i.e.

$$
Pr(L\bar{M} > z^2_{(\alpha+1)/2}|H_0). \tag{4.6}
$$

We expect that when inference is based on $L\bar{M}$ rather than on LM, the error of the approximation is reduced by one order. To this extent, the finite sample performance of $L\bar{M}$ should be similar to that of $g(LM)$, with g defined in (2.13) or (2.14).

Finally, we consider the mean-adjusted null statistic corresponding to (3.2). Since the algebraic burden is larger relative to the previous case, the derivation of the mean and variance-adjusted variant is omitted. At the beginning of this section, we stressed that mean and mean and variance adjustments might be algebraically more convenient than Edgeworth corrections. However, the mean and variance standardisation of (3.2) does not entail significant computational advantage and is therefore omitted.

Given (3.2), Robinson (2008) proposes the mean-adjusted null statistic

$$
\frac{L\tilde{M}}{E(L\tilde{M})}. \tag{4.7}
$$

Using standard formulae, we specify the results of Robinson (2008) as

$$
\begin{aligned}
E(L\tilde{M}) &= \frac{n^2}{a}\frac{E\left(\frac{1}{2}\epsilon'P(W+W')P\epsilon\right)^2}{E(\epsilon'P\epsilon)^2} \\
&= 1+\frac{(tr(K_1))^2}{a}+\frac{tr(K_2-K_3)}{a}-\frac{2(1-k)}{n}+O\left(\frac{1}{n^2}\right),
\end{aligned}
$$

where K_1, K_2 and K_3 are defined according to (3.8), (3.9) and (3.10), respectively. The second equality follows by a standard Taylor expansion of the denominator. Hence, (4.7) becomes

$$\tilde{LM}\left(1-\frac{(tr(K_1))^2}{a}-\frac{tr(K_2-K_3)}{a}\right)+o\left(\frac{h}{n}\right)$$

in case h is divergent, and

$$\tilde{LM}\left(1-\frac{(tr(K_1))^2}{a}-\frac{tr(K_2-K_3)}{a}+\frac{2(1-k)}{n}\right)+o\left(\frac{1}{n}\right)$$

if h is bounded. We define

$$\tilde{\tilde{LM}}=\tilde{LM}\left(1-\frac{(tr(K_1))^2}{a}-\frac{tr(K_2-K_3)}{a}\right) \tag{4.8}$$

in case h_n is divergent, and

$$\tilde{\tilde{LM}}=\tilde{LM}\left(1-\frac{(tr(K_1))^2}{a}-\frac{tr(K_2-K_3)}{a}+\frac{2(1-k)}{n}\right) \tag{4.9}$$

when h_n is bounded.

We consider the size of the test of (1.2) based on $\tilde{\tilde{LM}}$, i.e.

$$Pr(\tilde{\tilde{LM}}>z^2_{(\alpha+1)/2}|H_0). \tag{4.10}$$

As previously mentioned, the finite sample variance of the mean-adjusted statistic can be larger than that of the non corrected one. From (4.8) and (4.9), it is straightforward to notice that this might be the case, depending on the choice of W. By means of some Monte Carlo simulations we can assess whether the mean standardisation correction is worthwhile for any particular choice of W and its performance is therefore comparable with that based on Edgeworth corrections.

5 Bootstrap correction and simulation results

In this section we report some Monte Carlo simulations to investigate the finite sample performance of the refined tests derived in Sections 2, 3 and 4. Tables are reported at the end of the paper.

In this simulation work we adopt the Case (1991) specification for W, i.e.

$$W=I_r\otimes B_m, \quad B_m=\frac{1}{m-1}(ll'-I_m), \tag{5.1}$$

where r is the number of districts and m is the number of households in each district. We denote l an $m-$ dimensional column of ones. With this specification, two households are neighbours if they belong to the same district and each neighbour is given the same weight. Therefore, $n = mr$ and $h = m - 1$. W in (5.1) is symmetric and hence

$$a = 2tr(W^2), \qquad \kappa = \frac{12tr(W^4)}{(tr(W^2))^2}$$

In each of the 1000 replications the disturbance terms are generated from a normal distribution with mean zero and unit variance, i.e. according to Assumption 2 with $\sigma^2 = 1$. We set $\alpha = 95\%$. Moreover, we construct X as an $n \times 3$ matrix (that is, we set $k = 3$) whose first column is a column of ones, while each component of the remaining two columns are generated independently from a uniform distribution with support $[0, 1]$ and kept fixed at each replication.

For both models (1.1) and (3.1), the empirical sizes of the test of H_0 in (1.2) based on the usual normal approximation are compared with the same quantities obtained with both the Edgeworth-corrected critical values and Edgeworth-corrected test statistics. Such values are compared also with the empirical size based on the corrected statistics derived according to the procedure described in Section 4.

In addition, we consider the simulated sizes based on bootstrap critical values since it is well established that these achieve the first Edgeworth correction and should then be similar to the results obtained in Sections 2 and 3 (e.g. Hall (1992), Efron and Tibshirani (1993), DiCiccio and Romano (1995) or DiCiccio and Efron (1996)).

Before discussing and comparing the simulation results, we outline how the bootstrap critical values have been obtained. It must be stressed that we focus on the implementation of the bootstrap procedure, without addressing validity issues. Let Y^* be a vector of independent observations from the $N(0, Y'Y/n)$ distribution. Let

$$LM^* = \frac{n^2}{a}\left(\frac{Y^{*\prime}WY^*}{Y^{*\prime}Y^*}\right)^2.$$

Generating B pseudo-samples Y^*, w^*_α is defined such that the proportion of LM^* that does not exceed w^*_α is α. The bootstrap test rejects H_0 when $LM > w^*_\alpha$. Hence, the size of the test of (1.2) based on bootstrap is

$$Pr(LM > w^*_\alpha | H_0). \tag{5.2}$$

Regarding the procedure to obtain w^*_α, a remark is needed. When interested in testing, the bootstrap procedure when we impose H_0 to generate Y^* gives results at least as good as the same algorithm without imposing H_0 (Paparoditis and Politis (2005)). Some numerical work actually shows that imposing H_0 is convenient in the present case.

When dealing with (3.2), we modify the previous algorithm accordingly, i.e. we define

$$\tilde{LM}^* = \frac{n^2}{a}\left(\frac{u^{*\prime}PWPu^*}{u^{*\prime}Pu^*}\right)^2,$$

where u^* is a vector of independent observations from the $N(0, Y'PY/n)$ distribution. In this case, we denote $\tilde{w}^*_\alpha$ the bootstrap $\alpha-$quantile. The size of the test of (1.2) based on the bootstrap procedure is then

$$Pr(\tilde{LM} > \tilde{w}^*_\alpha | H_0). \tag{5.3}$$

In both procedures we set $B = 199$.

Tables 1 and 2 display the simulated values corresponding to (2.11), (2.12), (2.15), (4.6) and (5.2) when h (that is, m in (5.1)) is divergent and bounded, respectively. Moreover, Tables 3 and 4 display the simulated values corresponding to (3.11), (3.12), (3.13), (4.10) and (5.3) when h is either divergent or bounded, respectively. All the values in Tables 1-4 have to be compared with the nominal 5%. For notational convenience, in the Tables we denote by "chi square", "Edgeworth", "transformation", "mean-variance correction" and "bootstrap" the simulated values corresponding to (2.11)/(3.11),(2.12)/(3.12), (2.15)/(3.13), (4.6)/(4.10) and (5.2)/(5.3), respectively.

(Tables 1-4 about here)

From Tables 1 and 2 we notice that the approximation entailed by the first order asymptotic theory does not work well in practice. Indeed, the nominal 5% is underestimated for all sample sizes and whether h is divergent or bounded, although in the latter case the convergence to the nominal value appears to be faster, as expected. On the other hand, all the corrections we consider improve upon the approximation. In particular, when h is divergent (Table 1) the corrections based on the Edgeworth corrected test statistic and bootstrap critical values appear to outperform the others, at least for the sample sizes considered here. The same considerations hold in case h is bounded (Table 2), although the discrepancy among the performance of the different corrections is less glaring. This results were expected, since, as previously mentioned, the rate of convergence of the cdf of LM to the χ^2 cdf is faster in this case.

From Tables 3 and 4 we see that the usual test based on first order asymptotic theory performs even worse than in the previous case. However, the corrections give very satisfactory results. In particular, when h is divergent, both the test based on Edgeworth-corrected critical values and Edgeworth-corrected statistics appear to perform very well, giving results that are comparable to the bootstrap-based procedure. The simulated values corresponding to (4.6) are closer to the nominal 5% than ones of the standard test for all sample sizes, but not as satisfactory as the Edgeworth-based results. This might be due to the variance inflation discussed in Section 4. Again, when the sequence h is bounded, the pattern of the results appears to be very similar.

6 The exact distribution

In Sections 2 and 3 we developed refined procedures for testing (1.2) based on Lagrange Multiplier statistics, as given in (2.2) and (3.2), respectively. It must be mentioned that, since λ is a scalar parameter, we could have focused on the square root of the statistics in (2.2) and (3.2) and test H_0 against a one-sided alternative. We chose to develop the corrected procedure based on (2.2) and (3.2), and compare its performance in finite samples with that derived in Robinson (2008), because in several circumstances we might not have any preliminary evidence about the sign of λ and therefore the standard two-sided Lagrange Multiplier test might be preferred instead. However, it should be stressed that in case a test against a one-sided alternative is justified, suitable Edgeworth-corrections can be derived by a relatively straightforward modification of the proofs of either Theorems 1 or 2.

In this section we investigate numerically the properties of the distribution under H_0 of the square root of both (2.2) and (3.2), denoted by T and $\tilde{T}$ respectively, by means of Imhof's procedure and compare the results with those obtained using Edgeworth correction terms. The numerical evaluation of the df of T and $\tilde{T}$ and the corresponding quantiles, despite the obvious limitations of numerical algorithms, provides some information about the true distribution of the statistics and, to some extent, confirms the accuracy of Edgeworth corrections.

Since the numerical procedure is implemented using W given in (5.1), we describe the algorithm for a symmetric weight matrix, although it can be easily generalised to any choice of W. Moreover, we will describe the numerical procedure for evaluating the df of T, but the same argument with minor, obvious, modifications holds for $\tilde{T}$.

As discussed in the proof of Theorem 1, we can write $Pr(T \leq \zeta) = Pr(\epsilon' C \epsilon \leq 0)$, where $C = W - \zeta a^{1/2}/nI$ (that is (A.2) with $x = a^{1/2}\zeta$).

When the df can be written in terms of a quadratic form in normal random variables, as is the case in the last displayed expression, a procedure to evaluate it by numerical inversion of the characteristic function has been developed by Imhof (1961) and then improved and extended to different contexts by several authors. For the purpose of our implementation, we rely on the work by Imhof (1961), Davies (1973), Davies (1980), Ansley et al. (1992) and on the survey of Lu and King (2002).

Let s be the number of the distinct eigenvalues of $\sigma^2 C$, which are denoted by μ_j for $j = 1, .., s$, while n_j for $j = 1, ..., s$ is their order of algebraic multiplicity. Staring from the inversion formula of Gil-Pelaez (1951), Imhof (1961) suggests to evaluate the df of $\epsilon' C \epsilon$ as

$$Pr(\epsilon' C \epsilon \leq 0) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{sin\theta(u)}{u\gamma(u)} du, \tag{6.1}$$

where

$$\theta(u) = \sum_{j=0}^{s} \Big(\frac{n_j}{2} tg^{-1}(2u\mu_j) \Big) \quad \text{and} \quad \gamma(u) = \prod_{j=1}^{s} (1 + 4u^2\mu_j^2)^{n_j/4}.$$

The integral on the RHS of (6.1) cannot be evaluated using standard analytical methods because of the oscillatory nature of the integrand function and numerical procedures should be employed instead.

As suggested in Lu and King (2002), we rely on the discretisation rule provided by Davies (1973), which is based on a trapezoidal approximation for the integral on the RHS of (6.1), i.e.

$$Pr(\epsilon' C\epsilon \leq 0) = \frac{1}{2} - \sum_{m=0}^{M} \frac{sin\theta((m+\frac{1}{2})\Delta)}{\pi(m+\frac{1}{2})\gamma((m+\frac{1}{2})\Delta)}, \tag{6.2}$$

where Δ is the step interval and M is related to the truncation point, denoted by U henceforth, by the relationship $U = (M + 1/2)\Delta$. Both Δ and U need to be determined numerically.

We denote by $MGF(t)$ the moment generating function of $\epsilon' C\epsilon$. In order to evaluate Δ, we solve numerically the equation

$$MGF(t) - tMGF^{(1)}(t) - ln(E_I) = 0, \tag{6.3}$$

where $MGF^{(1)}(t) = dMGF(t)/dt$ and E_I is the maximum allowable integration error. It can be shown (see e.g. Ansley et al.(1992)) that the last displayed equation has always two solution $t_1 > 0$ and $t_2 < 0$, both satisfying the constraint $(1 - 2t_i\mu_j) > 0, \ \ \forall j = 1, ...s$, and $i = 1, 2$. For $i = 1, 2$, we define

$$\Delta_i = sign(t_i) \frac{2\pi}{MGF^{(1)}(t)|t = t_i}.$$

We choose Δ appearing in the RHS of (6.2) as the minimum value of Δ_i, for $i = 1, 2$.

U is derived as the numerical solution of

$$ln\xi(U) - lnE_T = 0, \tag{6.4}$$

where E_T is the maximum allowable truncation error and

$$\xi(U) = \frac{2}{\pi n} \prod_{j=1}^{s} |\mu_j|^{-n_j/2} (2U)^{-n/2}.$$

Once both Δ and U are obtained, the df of $\epsilon' C\epsilon$ using (6.2) can be evaluated. As suggested in Davies(1973), we set tolerance $E = 10^{-6}$ and we choose $E_I = 0.1E$ and $E_T = 0.9E$.

In order to calculate the α-quantile of the cdf of T, we need to find ζ so that

$$Pr(T \leq \zeta) = \alpha,$$

where the LHS of the last displayed expression can be obtained, as a function of ζ, by the algorithm described above. However, in the present case, the numerical solution to calculate ζ is particularly troublesome since the approximated df of T is almost flat as ζ varies.

Although Imhof's framework to obtain the cdf and its quantiles is useful to some extent, it obviously relies heavily on several numerical solutions of highly non-linear equations, such as (6.3) and (6.4). Hence it cannot be preferred to analytical procedures that improve upon the approximation given by the central limit theorem, such as those based on Edgeworth expansions or on mean and variance standardization. However, despite being not fully reliable, quantiles obtained with Imhof's procedure can be compared with Edgeworth-corrected ones, to provide further evidence that the latter are closer to the true values than those of the normal df.

Edgeworth-corrected quantiles of the df of T can be obtained from some intermediate results displayed in Theorem 1 and a procedure similar to that described in Section 2. Specifically, in Appendix A we derive the Edgeworth expansion for the df of T as

$$Pr\left(T \leq \zeta\right) \quad = \quad \Phi(\zeta) - \frac{\bar{\kappa}}{3!}H_2(\zeta)\phi(\zeta) + o\left(\sqrt{\frac{h}{n}}\right),$$

where $\bar{\kappa} = tr(W' + W)^3/a^{3/2}$ and H_2 is the second Hermite polynomial. From the last displayed expression we can derive a corresponding expansion for the $\alpha-$quantile by a straightforward modification of the argument presented in Section 2. We denote the true $\alpha-$quantile of the df of T by q_α and we write

$$q_\alpha = z_\alpha + \frac{\bar{\kappa}}{3!}H_2(\zeta) + o\left(\sqrt{\frac{h}{n}}\right),$$

whether h is either divergent or bounded.

(Tables 5 and 6 about here)

As expected, from Tables 5 and 6 we notice that for all sample sizes and for h being either divergent or bounded, the Edgeworth-corrected quantiles for $\alpha = 0.95, 0.975, 0.99$ are closer to those obtained by Imhof's procedure than ones of the standard normal df. Indeed, the standard normal quantiles are significantly lower than Imhof's ones for all sample sizes. To some extent, this confirms that tests based on Edgeworth-corrected critical values should be more reliable than those based on the standard normal approximation.

Imhof's algorithm has also been implemented to obtain the df of $\tilde{T}$. Unfortunately, in this case, the numerical procedure does not work well and it appears to be too sensitive to both the choice of the initial values for the numerical solution of non-linear equations and the choice of X. This give strong motivation to the practitioner to rely on the analytical corrections based on Edgeworth expansions, rather than on numerical procedures to evaluate the exact df.

A Appendix A: Proof of Theorem 1

Under (1.2), (2.2) becomes $LM = n^2(\epsilon' W\epsilon)^2/a$, so we start by deriving the formal Edgeworth expansion of the cdf of

$$n\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}. \tag{A.1}$$

The development is standard. The df of (A.1) can be written in terms of a quadratic form in ϵ, i.e.

$$Pr(n\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} \leq x) = Pr(\epsilon' C\epsilon \leq 0),$$

where

$$C = \frac{1}{2}(W + W') - \frac{x}{n}I \tag{A.2}$$

and x is any real number.

Under Assumption 1, the characteristic function (cf) of $\epsilon' C\epsilon$ can be derived as

$$\begin{aligned} E(e^{it(\epsilon' C\epsilon)}) &= \frac{1}{(2\pi)^{n/2}\sigma^n}\int_{\Re^n} e^{it(\xi' C\xi)}e^{-\frac{\xi'\xi}{2\sigma^2}}d\xi = \frac{1}{(2\pi)^{n/2}\sigma^n}\int_{\Re^n} e^{-\frac{1}{2\sigma^2}\xi'(I-2it\sigma^2 C)\xi}d\xi \\ &= det(I - 2it\sigma^2 C)^{-1/2} = \prod_{j=1}^{n}(1 - 2it\sigma^2\gamma_j)^{-1/2}, \end{aligned} \tag{A.3}$$

where $det(A)$ denotes the determinant of a generic square matrix A, γ_j are the eigenvalues of C and $i = \sqrt{-1}$. From (A.3) the cumulant generating function of $\epsilon' C\epsilon$ is

$$\begin{aligned} \psi(t) &= -\frac{1}{2}\sum_{j=1}^{n} ln(1 - 2it\sigma^2\gamma_j) = \frac{1}{2}\sum_{j=1}^{n}\sum_{s=1}^{\infty}\frac{(2it\sigma^2\gamma_j)^s}{s} \\ &= \frac{1}{2}\sum_{s=1}^{\infty}\frac{(2it\sigma^2)^s}{s}\sum_{j=1}^{n}\gamma_j^s = \frac{1}{2}\sum_{s=1}^{\infty}\frac{(2it\sigma^2)^s}{s}tr(C^s). \end{aligned} \tag{A.4}$$

From (A.4) the s-th cumulant, κ_s, of $\epsilon' C\epsilon$ is given by

$$\kappa_1 = \sigma^2 tr(C), \tag{A.5}$$

$$\kappa_2 = 2\sigma^4 tr(C^2), \tag{A.6}$$

$$\kappa_s = \frac{\sigma^{2s} s! 2^{s-1} tr(C^s)}{s}, s > 2. \tag{A.7}$$

Thus the cumulant generating function of $f^c = (\epsilon' C\epsilon - \kappa_1)/\kappa_2^{1/2}$ is

$$\psi^c(t) = -\frac{1}{2}t^2 + \sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!},$$

where

$$\kappa_s^c = \frac{\kappa_s}{\kappa_2^{s/2}}, \tag{A.8}$$

Hence, the cf of f^c is

$$\begin{aligned}
E(e^{itf^c}) &= e^{-\frac{1}{2}t^2}\exp\{\sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!}\} = \\
&= e^{-\frac{1}{2}t^2}\{1+\sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!}+\frac{1}{2!}(\sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!})^2+\frac{1}{3!}(\sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!})^3+.....\} \\
&= e^{-\frac{1}{2}t^2}\{1+\frac{\kappa_3^c(it)^3}{3!}+\frac{\kappa_4^c(it)^4}{4!}+\frac{\kappa_5^c(it)^5}{5!}+\{\frac{\kappa_6^c}{6!}+\frac{(\kappa_3^c)^2}{(3!)^2}\}(it)^6+.....\}.
\end{aligned}$$

Denote by $\phi(\zeta)$ and $\Phi(\zeta)$ the normal pdf and df, respectively. By the Fourier inversion formula, we can conclude that

$$Pr(f^c \le z) = \int_{-\infty}^{z}\phi(z)dz + \frac{\kappa_3^c}{3!}\int_{-\infty}^{z}H_3(z)\phi(z)dz + \frac{\kappa_4^c}{4!}\int_{-\infty}^{z}H_4(z)\phi(z)dz + ...,$$

where $H_i(z)$ is the $i-th$ Hermite polynomial. Collecting the results derived above, we have

$$\begin{aligned}
Pr\left(n\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} \le x\right) &= Pr(\epsilon' C\epsilon \le 0) = Pr(f^c\kappa_2^{1/2}+\kappa_1 \le 0) = Pr(f^c \le -\kappa_1^c) \\
&= \Phi(-\kappa_1^c) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(-\kappa_1^c) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(-\kappa_1^c) + ..., \qquad \text{(A.9)}
\end{aligned}$$

where $g^{(i)}$ denotes the ith derivative of the function g.

From (A.5), (A.6) and given (A.2), we obtain

$$\kappa_1 = -\sigma^2 x, \quad \kappa_2 = \sigma^4\left(tr(W^2+W'W)+\frac{2}{n}x^2\right)$$

and hence, from (A.8),

$$\kappa_1^c = \frac{-x}{a^{1/2}\left(1+\frac{2}{na}x^2\right)^{1/2}}. \qquad \text{(A.10)}$$

Denote by ζ any finite real number and set

$$x = a^{1/2}\zeta. \qquad \text{(A.11)}$$

Under Assumption 2, $x \sim \sqrt{n/h}$. By Taylor expansion of the denominator of (A.10) we obtain

$$\kappa_1^c = -\zeta\left(1-\frac{1}{n}\zeta^2\right)+o\left(\frac{1}{n}\right).$$

Moreover, by Assumption 2,

$$\kappa_3^c = \frac{8\sigma^6 tr(C^3)}{\kappa_2^{3/2}} \sim \frac{tr(W'+W)^3}{a^{3/2}} \sim \sqrt{\frac{h}{n}}$$

and

$$\kappa_4^c = \frac{48\sigma^8 tr(C^4)}{(\kappa_2)^2} \sim \frac{3tr(W'+W)^4}{a^2} \sim \frac{h}{n}. \qquad \text{(A.12)}$$

By Taylor expansion we have

$$\Phi(-\kappa_1^c) = \Phi(\zeta) + O\left(\frac{1}{n}\right) = \Phi(\zeta) + o\left(\frac{h_n}{n}\right)$$

when h is divergent and

$$\Phi(-\kappa_1^c) \quad = \quad \Phi(\zeta) - \frac{1}{n}\zeta^3\phi(\zeta) + o\left(\frac{1}{n}\right)$$

when h is bounded. Therefore, for x given in (A.11), when h is divergent (A.9) is

$$\begin{aligned} Pr\left(na^{-1/2}\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} \leq \zeta\right) &= \Phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + o\left(\frac{h}{n}\right) \\ &= \Phi(\zeta) - \frac{\kappa_3^c}{3!}H_2(\zeta)\phi(\zeta) - \frac{\kappa_4^c}{4!}H_3(\zeta)\phi(\zeta) + o\left(\frac{h}{n}\right), \end{aligned} \tag{A.13}$$

where the last equality follows by

$$\left(\frac{d}{dx}\right)^j \Phi(x) = -H_{j-1}(x)\phi(x). \tag{A.14}$$

Similarly when h is bounded (A.9) becomes

$$\begin{aligned} Pr\left(na^{-1/2}\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} \leq \zeta\right) &= \Phi(\zeta) - \frac{\zeta^3}{n}\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + o\left(\frac{1}{n}\right) \\ &= \Phi(\zeta) - \frac{\kappa_3^c}{3!}H_2(\zeta)\phi(\zeta) - \left(\frac{\zeta^3}{n} + \frac{\kappa_4^c}{4!}H_3(\zeta)\right)\phi(\zeta) + o\left(\frac{1}{n}\right). \end{aligned} \tag{A.15}$$

For notational simplicity, let $T = na^{-1/2}\epsilon' W\epsilon/\epsilon'\epsilon$, so that $LM = T^2$. Term by term differentiation of (A.13) and (A.15) gives the corresponding expressions for the pdf of T, $f_T(\zeta)$, i.e.

$$f_T(\zeta) = \phi(\zeta) - \frac{\kappa_3^c}{3!}(-\zeta^3 + 3\zeta)\phi(\zeta) - \frac{\kappa_4^c}{4!}(-\zeta^4 + 6\zeta^2 - 3)\phi(\zeta) + o(\frac{h}{n}) \tag{A.16}$$

and

$$\begin{aligned} f_T(\zeta) &= \phi(\zeta) + \frac{1}{n}(\zeta^4 - 3\zeta^2)\phi(\zeta) - \frac{\kappa_3^c}{3!}(-\zeta^3 + 3\zeta)\phi(\zeta) \\ &- \frac{\kappa_4^c}{4!}(-\zeta^4 + 6\zeta^2 - 3)\phi(\zeta) + o(\frac{1}{n}), \end{aligned} \tag{A.17}$$

respectively.

For divergent h, using (A.16), we can derive an approximate expression for the cf of T^2 as

$$\begin{aligned} &\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{itv^2}e^{-\frac{v^2}{2}}(1 - \frac{\kappa_3^c}{3!}(-v^3 + 3v) - \frac{\kappa_4^c}{4!}(-v^4 + 6v^2 - 3))dv \\ = \quad &\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}(1 - \frac{\kappa_3^c}{3!}(-v^3 + 3v) - \frac{\kappa_4^c}{4!}(-v^4 + 6v^2 - 3))dv. \end{aligned} \tag{A.18}$$

We notice that the first term of the last displayed integral is

$$\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}dv = (1-2it)^{-1/2},$$

which is the χ^2 cf. By Gaussian integration, the second and third terms are, respectively,

$$\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}\frac{\kappa_3^c}{3!}v^3 dv = 0$$

and

$$\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}\frac{\kappa_3^c}{3!}3v dv = 0,$$

while

$$\frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}v^4 dv = \frac{3}{(1-2it)^{5/2}}, \quad \frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)}v^2 dv = \frac{1}{(1-2it)^{3/2}}.$$

Collecting the previously displayed results, (A.18) becomes

$$\frac{1}{\sqrt{1-2it}} + \frac{\kappa_4^c}{8}\frac{1}{\sqrt{1-2it}} - \frac{\kappa_4^c}{4}\frac{1}{(1-2it)^{3/2}} + \frac{\kappa_4^c}{8}\frac{1}{(1-2it)^{5/2}}. \tag{A.19}$$

Term by term Fourier inversion of (A.19) gives

$$\begin{aligned} Pr(LM \leq \eta) &= F(\eta) + \frac{\kappa_4^c}{8}F(\eta) - \frac{\kappa_4^c}{4}F_3(\eta) + \frac{\kappa_4^c}{8}F_5(\eta) + o\left(\frac{h}{n}\right) \\ &= F(\eta) + \frac{\kappa_4^c}{4}\eta f(\eta) - \frac{\kappa_4^c}{12}\eta^2 f(\eta) + o\left(\frac{h}{n}\right). \end{aligned} \tag{A.20}$$

The last displayed equality follows from the recursions (see e.g. Harris (1985))

$$\begin{aligned} f_{k+2}(x) &= xk^{-1}f_k(x), \\ F_{k+2}(x) &= F_k(x) - 2xk^{-1}f_k(x), \end{aligned} \tag{A.21}$$

where f_k and F_k denote the χ^2 pdf and cdf with k degrees of freedom, respectively. When no subscript is specified $k = 1$.

Similarly, for bounded h, from (A.17) we obtain

$$\frac{1}{\sqrt{1-2it}} + \frac{\kappa_4^c}{8}\frac{1}{\sqrt{1-2it}} - \frac{\kappa_4^c}{4}\frac{1}{(1-2it)^{3/2}} + \frac{\kappa_4^c}{8}\frac{1}{(1-2it)^{5/2}} + \frac{1}{n}\frac{3}{(1-2it)^{5/2}} - \frac{3}{n}\frac{1}{(1-2it)^{3/2}}$$

and thus, term by term Fourier inversion gives

$$\begin{aligned} Pr(LM \leq \eta) &= F(\eta) + \frac{\kappa_4^c}{8}F(\eta) - \frac{\kappa_4^c}{4}F_3(\eta) + \frac{\kappa_4^c}{8}F_5(\eta) + \frac{3}{n}(-F_3(\eta) + F_5(\eta)) + o\left(\frac{1}{n}\right) \\ &= F(\eta) + \frac{\kappa_4^c}{4}\eta f(\eta) - \frac{\kappa_4^c}{12}\eta^2 f(\eta) - \frac{2}{n}\eta^2 f(\eta) + o\left(\frac{1}{n}\right). \end{aligned} \tag{A.22}$$

The claim in Theorem 1 follows from (A.20) and (A.22) by letting $\kappa = 3tr(W' + W)^4/a^2$, which is the leading term of κ_4^c, as given in (A.12).

B Appendix B: Proof of Theorem 2

Parts of the proof of Theorem 2 are similar to Theorem 1 and are omitted. We derive the third order Edgeworth expansion of the df of

$$n\frac{\epsilon' PWP\epsilon}{\epsilon' P\epsilon}, \tag{B.1}$$

where P is defined according to (3.3). The cdf of (B.1) can be written in terms of a quadratic form in ϵ, i.e.

$$Pr(\frac{\epsilon' PWP\epsilon}{\frac{1}{n}\epsilon' P\epsilon} \leq z) = Pr(\epsilon'\tilde{C}\epsilon \leq 0),$$

where

$$\tilde{C} = \frac{1}{2}P(W+W')P - \frac{1}{n}Pz \tag{B.2}$$

and z is any real number.

The same argument presented in Appendix A for the evaluation of both characteristic and cumulative generating functions holds here with $\tilde{C}$ instead of C. Therefore we have

$$\tilde{\kappa}_1 = \sigma^2 tr(\tilde{C}), \quad \tilde{\kappa}_2 = 2\sigma^4 tr(\tilde{C}^2)$$

and

$$\tilde{\kappa}_s = \frac{\sigma^{2s}s!}{2s} tr((2\tilde{C})^s), s > 2.$$

From (B.2) we obtain

$$\tilde{\kappa}_1 = \sigma^2 tr(PW) - \sigma^2\frac{1}{n}tr(P)z = -\sigma^2(tr((X'X)^{-1}X'WX) - \frac{n-k}{n}z).$$

Also, by straightforward algebra,

$$\begin{aligned}
\tilde{\kappa}_2 &= \sigma^4(tr(WPWP) + tr(W'PWP) + 2\frac{n-k}{n}z^2 - \frac{4}{n}tr(PW)z) \\
&= \sigma^4(tr((W+W')PWP) + 2\frac{n-k}{n^2}z^2 - \frac{4}{n}tr(PW)z) \\
&= \sigma^4(tr(W^2) + tr(W'W) + \frac{1}{2}tr(X'(W+W')X(X'X)^{-1}X'(W'+W)X(X'X)^{-1}) \\
&- tr(X'(W+W')^2X(X'X)^{-1}) + 2\frac{n-k}{n^2}z^2 + \frac{4}{n}tr((X'X)^{-1}X'WX)z).
\end{aligned}$$

By (3.8), (3.9), and (3.10), we write

$$\tilde{\kappa}_1 = -\sigma^2 tr(K_1) - \sigma^2 z + \sigma^2\frac{k}{n}z \tag{B.3}$$

and

$$\tilde{\kappa}_2 = \sigma^4(a + tr(K_2 - K_3) + 2\frac{n-k}{n^2}z^2 + \frac{4}{n}tr(K_1)z). \tag{B.4}$$

Similarly to the proof of Theorem 1, we define $\tilde{f}^c = (\epsilon'\tilde{C}\epsilon - \tilde{\kappa}_1)/\tilde{\kappa}_2^{1/2}$ and derive the centred cumulants as $\tilde{\kappa}_s^c = \tilde{\kappa}_s/\tilde{\kappa}_2^{s/2}$. From (B.3) and (B.4) we have

$$\tilde{\kappa}_1^c = \frac{-\sigma^2 tr(K_1) - \sigma^2 z + \sigma^2\frac{k}{n}z}{\sigma^2 a^{1/2}\left(1 + \frac{tr(K_2)}{a} - \frac{tr(K_3)}{a} + \frac{2}{a}\frac{n-k}{n^2}z^2 + \frac{4}{n}\frac{tr(K_1)z}{a}\right)^{1/2}}. \tag{B.5}$$

We choose $z = a^{1/2}\zeta$. Under Assumptions 2 and 3 we have $a \sim n/h$ and

$$z \sim \sqrt{\frac{n}{h}}, \qquad \frac{tr(K_1)}{a} \sim \frac{h}{n}, \qquad \frac{tr(K_2)}{a} \sim \frac{h}{n}, \quad \frac{tr(K_3)}{a} \sim \frac{h}{n}.$$

Hence, substituting the expression for z in (B.5) and performing a standard Taylor expansion of the denominator we obtain

$$\begin{aligned} \tilde{\kappa}_1^c &= -\left(\zeta + \frac{tr(K_1)}{a^{1/2}}\right)\left(1 + \frac{tr(K_3 - K_2)}{2a} + o\left(\frac{h}{n}\right)\right) \\ &= -\zeta - \frac{tr(K_1)}{a^{1/2}} - \frac{tr(K_3 - K_2)}{2a}\zeta + o\left(\frac{h}{n}\right) \end{aligned}$$

in case h is divergent, and

$$\begin{aligned} \tilde{\kappa}_1^c &= -\left(\zeta + \frac{tr(K_1)}{a^{1/2}} - \frac{k}{n}\zeta\right)\left(1 + \frac{tr(K_3 - K_2)}{2a} - \frac{1}{n}\zeta^2 + o\left(\frac{1}{n}\right)\right) \\ &= -\zeta - \frac{tr(K_1)}{a^{1/2}} + \frac{k}{n}\zeta - \frac{tr(K_3 - K_2)}{2a}\zeta + \frac{1}{n}\zeta^3 + o\left(\frac{1}{n}\right) \end{aligned}$$

if h is bounded.

Moreover,

$$\tilde{\kappa}_3^c = \frac{8\sigma^6 tr(\tilde{C}^3)}{\kappa_2^{3/2}} \sim \frac{tr(((W+W')P)^3)}{a^{3/2}} \sim \sqrt{\frac{h}{n}}$$

and

$$\tilde{\kappa}_4^c = \frac{48\sigma^8 tr(\tilde{C}^4)}{\tilde{\kappa}_2^2} \sim \frac{3tr(((W+W')P)^4)}{a^2} \sim \frac{h}{n}.$$

Therefore, proceeding as in the proof of Theorem 1 with $\tilde{f}^c$, $\tilde{\kappa}_s$ and $\tilde{\kappa}_s^c$ instead of f^c, κ_s and κ_s^c, we have

$$\begin{aligned} Pr(na^{-1/2}\frac{\epsilon' PWP\epsilon}{\epsilon' P\epsilon} \leq z) &= Pr(\epsilon'\tilde{C}\epsilon \leq 0) = Pr(\tilde{f}^c\tilde{\kappa}_2^{1/2} + \tilde{\kappa}_1 \leq 0) = Pr(\tilde{f}^c \leq -\tilde{\kappa}_1^c) \\ &= \Phi(-\tilde{\kappa}_1^c) - \frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(-\tilde{\kappa}_1^c) + \frac{\tilde{\kappa}_4^c}{4!}\Phi^{(4)}(-\tilde{\kappa}_1^c) + ... \qquad \text{(B.6)} \end{aligned}$$

By Taylor expansion we have

$$\Phi(-\tilde{\kappa}_1^c) = \Phi(\zeta) + \frac{tr(K_1)}{a^{1/2}}\phi(\zeta) + \frac{tr(K_3 - K_2)}{a}\zeta\phi(\zeta) + \frac{1}{2}\left(\frac{tr(K_1)}{a^{1/2}}\right)^2\Phi^{(2)}(\zeta) + o\left(\frac{h}{n}\right)$$

when h is divergent and

$$\begin{aligned} \Phi(-\tilde{\kappa}_1^c) &= \Phi(\zeta) + \frac{tr(K_1)}{a^{1/2}}\phi(\zeta) + \frac{tr(K_3 - K_2)}{a}\zeta\phi(\zeta) \\ &- \frac{k}{n}\zeta\phi(\zeta) - \frac{1}{n}\zeta^3\phi(\zeta) + \frac{1}{2}\left(\frac{tr(K_1)}{a^{1/2}}\right)^2\Phi^{(2)}(\zeta) + o\left(\frac{1}{n}\right) \end{aligned}$$

when h is bounded. Therefore, (B.6) becomes

$$\begin{aligned}
Pr(na^{-1/2}\frac{\epsilon' PWP\epsilon}{\epsilon' P\epsilon} \leq \zeta) &= \Phi(\zeta) + \frac{tr(K_1)}{a^{1/2}}\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(\zeta) + \frac{tr(K_3 - K_2)}{a}\zeta\phi(\zeta) \\
&+ \frac{1}{2}\left(\frac{tr(K_1)}{a^{1/2}}\right)^2 \Phi^{(2)}(\zeta) + \frac{\tilde{\kappa}_4^c}{4!}\Phi^{(4)}(\zeta) + o\left(\frac{h}{n}\right) \\
&= \Phi(\zeta) + \frac{tr(K_1)}{a^{1/2}}\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}H_2(\zeta)\phi(\zeta) + \frac{tr(K_3 - K_2)}{a}\zeta\phi(\zeta) \\
&- \frac{1}{2}\left(\frac{tr(K_1)}{a^{1/2}}\right)^2 H_1(\zeta)\phi(\zeta) - \frac{\tilde{\kappa}_4^c}{4!}H_3(\zeta)\phi(\zeta) + o\left(\frac{h}{n}\right),
\end{aligned} \tag{B.7}$$

where the last equality follows by (A.14). Similarly, when h is bounded,

$$\begin{aligned}
Pr(na^{-1/2}\frac{\epsilon' PWP\epsilon}{\epsilon' P\epsilon} \leq \zeta) &= \Phi(\zeta) + \frac{tr(K_1)}{a^{1/2}}\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(\zeta) + \frac{tr(K_3 - K_2)}{a}\zeta\phi(\zeta) \\
&+ \frac{1}{2}\left(\frac{tr(K_1)}{a^{1/2}}\right)^2 \Phi^{(2)}(\zeta) - \frac{k}{n}\zeta\phi(\zeta) - \frac{1}{n}\zeta^3\phi(\zeta) + \frac{\tilde{\kappa}_4^c}{4!}\Phi^{(4)}(\zeta) + o\left(\frac{h}{n}\right) \\
&= \Phi(\zeta) + \frac{tr(K_1)}{a^{1/2}}\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}H_2(\zeta)\phi(\zeta) + \frac{tr(K_3 - K_2)}{a}\zeta\phi(\zeta) \\
&- \frac{1}{2}\left(\frac{tr(K_1)}{a^{1/2}}\right)^2 H_1(\zeta)\phi(\zeta) - \frac{k}{n}\zeta\phi(\zeta) - \frac{1}{n}\zeta^3\phi(\zeta) \\
&- \frac{\tilde{\kappa}_4^c}{4!}H_3(\zeta)\phi(\zeta) + o\left(\frac{1}{n}\right).
\end{aligned}$$

For notational convenience, we write $\tilde{T} = na^{-1/2}\epsilon' PWP\epsilon/\epsilon' P\epsilon$, so that $L\tilde{M} = \tilde{T}^2$. Moreover, we recall that $H_1(\zeta) = \zeta$, $H_2(\zeta) = \zeta^2 - 1$ and $H_3(\zeta) = \zeta^3 - 3\zeta$.

As discussed in detail in Appendix A, term by term differentiation of (B.7) and (B.8) gives

$$\begin{aligned}
f_{\tilde{T}}(\zeta) &= \phi(\zeta) - \frac{tr(K_1)}{a^{1/2}}\zeta\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}(-\zeta^3 + 3\zeta)\phi(\zeta) \\
&+ \frac{tr(K_3 - K_2)}{a}(1-\zeta^2)\phi(\zeta) - \frac{1}{2}\frac{(tr(K_1))^2}{a}(1-\zeta^2)\phi(\zeta) \\
&- \frac{\tilde{\kappa}_4^c}{4!}(-\zeta^4 + 6\zeta^2 - 3)\phi(\zeta) + o\left(\frac{h}{n}\right)
\end{aligned} \tag{B.8}$$

and

$$\begin{aligned}
f_{\tilde{T}}(\zeta) &= \phi(\zeta) - \frac{tr(K_1)}{a^{1/2}}\zeta\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}(-\zeta^3 + 3\zeta)\phi(\zeta) \\
&+ \frac{tr(K_3 - K_2)}{a}(1-\zeta^2)\phi(\zeta) - \frac{1}{2}\frac{(tr(K_1))^2}{a}(1-\zeta^2)\phi(\zeta) \\
&- \frac{k}{n}(1-\zeta^2)\phi(\zeta) - \frac{1}{n}(3\zeta^2 - \zeta^4)\phi(\zeta) \\
&- \frac{\tilde{\kappa}_4^c}{4!}(-\zeta^4 + 6\zeta^2 - 3)\phi(\zeta) + o\left(\frac{1}{n}\right),
\end{aligned} \tag{B.9}$$

respectively.

In order to simplify the notation, we write

$$\omega_1 = \frac{tr(K_3 - K_2)}{a} - \frac{1}{2}\frac{(tr(K_1))^2}{a}, \quad \omega_2 = \frac{tr(K_3 - K_2)}{a} - \frac{1}{2}\frac{(tr(K_1))^2}{a} - \frac{k}{n}$$

and

$$\omega_3 = \frac{tr(K_1)}{a^{1/2}} + \frac{\tilde{\kappa}_3^c}{2}.$$

Proceeding as described in the proof of Theorem 1, when h is divergent we approximate the cf of $\tilde{T}$ as

$$\begin{aligned}
& \frac{1}{\sqrt{2\pi}}\int_{\Re} e^{itv^2} e^{-\frac{v^2}{2}} (1 - \omega_3 v + \frac{\tilde{\kappa}_3^c}{3!}v^3 + \omega_1(1 - v^2) - \frac{\tilde{\kappa}_4^c}{4!}(-v^4 + 6v^2 - 3))dv \\
= \quad & \frac{1}{\sqrt{2\pi}}\int_{\Re} e^{-\frac{v^2}{2}(1-2it)} (1 - \omega_3 v + \frac{\tilde{\kappa}_3^c}{3!}v^3 + \omega_1(1 - v^2) - \frac{\tilde{\kappa}_4^c}{4!}(-v^4 + 6v^2 - 3))dv \\
= \quad & \frac{1}{\sqrt{1-2it}}\left(1 + \omega_1 - \frac{\omega_1}{1-2it} + \frac{\tilde{\kappa}_4^c}{8}\frac{1}{(1-2it)^2} - \frac{\tilde{\kappa}_4^c}{4}\frac{1}{1-2it} + \frac{\tilde{\kappa}_4^c}{8}\right). \qquad \text{(B.10)}
\end{aligned}$$

By term by term Fourier inversion of (B.10) and some standard algebraic manipulation, we obtain

$$\begin{aligned}
Pr(L\tilde{M} \le \eta) \quad &= \quad F(\eta) + \left(\frac{\tilde{\kappa}_4^c}{8} + \omega_1\right)F(\eta) - \left(\omega_1 + \frac{\tilde{\kappa}_4^c}{4}\right)F_3(\eta) + \frac{\tilde{\kappa}_4^c}{8}F_5(\eta) + o\left(\frac{h}{n}\right) \\
&= \quad F(\eta) + \left(\frac{\tilde{\kappa}_4^c}{4}\eta - \frac{\tilde{\kappa}_4^c}{12}\eta^2 + 2\omega_1\eta\right)f(\eta) + o\left(\frac{h}{n}\right). \qquad \text{(B.11)}
\end{aligned}$$

Similarly, when h is bounded we have

$$\begin{aligned}
Pr(L\tilde{M} \le \eta) \quad &= \quad F(\eta) + \left(\frac{\tilde{\kappa}_4^c}{8} + \omega_2\right)F(\eta) - \left(\omega_2 + \frac{\tilde{\kappa}_4^c}{4} + \frac{3}{n}\right)F_3(\eta) \\
&+ \quad \left(\frac{\tilde{\kappa}_4^c}{8} + \frac{3}{n}\right)F_5(\eta) + o\left(\frac{1}{n}\right) \\
&= \quad F(\eta) + \left(\frac{\tilde{\kappa}_4^c}{4}\eta - \frac{\tilde{\kappa}_4^c}{12}\eta^2 + 2\omega_2\eta - \frac{2}{n}\eta^2\right)f(\eta) + o\left(\frac{1}{n}\right). \text{(B.12)}
\end{aligned}$$

The claim in Theorem 2 follows from (B.11) and (B.12) by observing that the leading term of $\tilde{\kappa}_4^c$ is $\kappa = 3tr(W' + W)^4/a^2$. Indeed, each term in $(tr(W + W')^4 P)$ other than $tr((W + W')^4) \sim n/h$ is $O(1)$ by Assumption 3 and Lemma 2, and is therefore $o(n/h)$.

C Appendix C

In this Appendix we state and prove some Lemmas that have been used in the proofs of Theorems 1 and 2, as well as in the derivation of some other results in the paper. In particular, Lemma 1 is used in the proof of both Theorems 1 and 2, while Lemmas 2 and 3 are auxiliary results for the proof of Theorem 2. Both Lemma 1 and 2 are similar to results reported in Lee (2004). Lemma 4, instead, is used to construct Edgeworth corrected statistics staring from the corresponding asymptotic expansion when the limiting distribution is that of a χ^2 random variable.

Lemma 1 If $w_{ij} = O(1/h)$, uniformly in i and j,

$$tr(WA) = O\left(\frac{n}{h}\right),$$

where A is an $n \times n$ matrix, uniformly bounded in row and column sums in absolute value.

Proof Let a_{ij} be the $i-j$th element of A. The $i-$th diagonal element of WA has absolute value given by

$$|(WA)_{ii}| \leq \underset{j}{max}|w_{ij}|\sum_{j=1}^{n}|a_{ji}| = O\left(\frac{1}{h}\right),$$

uniformly in i. Therefore:

$$|tr(WA)| \leq \sum_{i=1}^{n}|(WA)_{ii}| \leq n\underset{i}{max}|(WA)_{ii}| = O\left(\frac{n}{h}\right).$$

Lemma 2 Suppose that the elements of the $n \times 1$ vectors R and S are uniformly bounded. If the matrix A is uniformly bounded in either row or column sums in absolute value it follows that $|R'AS| = O(n)$.

Proof Let r_i, s_i and a_{ij} be the $i-$th element of R, $i-$th element of S and $i-j$th element of A, respectively. Moreover, we denote K a generic constant. We suppose that A is uniformly bounded in absolute value in column sums. The same argument (with A' instead of A) would hold if A were uniformly bounded in absolute value in row sums, since $|R'AS| = |S'A'R|$.

By assumption, we have

$$\underset{1\leq i\leq n}{max}|r_i| \leq K \qquad \underset{1\leq i\leq n}{max}|s_i| \leq K \qquad \underset{1\leq i\leq n}{max}\sum_{j=1}^{n}|a_{ij}| \leq K.$$

We have

$$\begin{aligned}|R'AS| &= |\sum_{i=1}^{n}\sum_{j=1}^{n}r_i a_{ij} s_j| \leq \sum_{i=1}^{n}\sum_{j=1}^{n}|r_i||a_{ij}||s_j| \leq \underset{1\leq i,j\leq n}{max}|r_i||s_j|\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}| \\ \leq K &\quad \sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}| \leq K\underset{1\leq i\leq n}{max}\sum_{j=1}^{n}|a_{ij}|\sum_{i=1}^{n}1 = O(n).\end{aligned}$$

Lemma 3 Under Assumptions 3, P is uniformly bounded in row and column sums in absolute value.

Proof We show that $X(X'X)^{-1}X'$ is uniformly bounded in row sums in absolute value. Let x_{ij} be the $i-j$th element of X and x_i' the ith row of X. Moreover, we denote C and c generic large and small constants, respectively.

Under Assumption 3

$$\underset{1\leq i,j\leq n}{max}|x_{ij}| < C \quad 0 < c < \nu_{min}\left(\frac{1}{n}X'X\right),$$

for n large enough. We have

$$\max_{1\leq i\leq n}\sum_{j=1}^{n}|(X(X'X)^{-1}X')_{ij}| = \max_{1\leq<i\leq n}\sum_{j=1}^{n}|(x_i'(X'X)^{-1}x_j)|$$

$$\leq \quad \max_{1\leq i\leq n}\sum_{j=1}^{n}||x_i'||||(X'X)^{-1}||||x_j|| \leq \max_{1\leq i,j\leq n}||x_i'||||(\frac{1}{n}X'X)^{-1}||||x_j||,$$

where $||.||$ denotes the spectral norm. Now,

$$||(\frac{1}{n}X'X)^{-1}|| \quad = \quad \nu_{max}((\frac{1}{n}X'X)^{-1}) = \frac{1}{\nu_{min}(\frac{1}{n}X'X)} \leq \frac{1}{c},$$

where $\nu_{max}()$ denotes the largest eigenvalue. Moreover,

$$\max_{0<i\leq n}||x_i'|| = \max_{0<i\leq n}(x_i'x_i)^{1/2} \leq (kC^2)^{1/2}.$$

It follows that

$$\max_{0<i,j\leq n}||x_i'||||(\frac{1}{n}X'X)^{-1}||||x_j|| \leq \frac{1}{c}kC^2 < \infty.$$

By symmetry, we conclude that $X(X'X)^{-1}X'$ is also bounded in column sums in absolute value. Trivially, the same property holds for $P = I - X(X'X)^{-1}X'$.

Lemma 4 Let ξ be a statistic such that its cdf admits the expansion

$$Pr(\xi \leq \eta) = F(\eta) + \frac{h}{n}s(\eta)f(\eta) + o\left(\frac{h}{n}\right), \tag{C.1}$$

where h can be either divergent or bounded and $s(\eta)$ is a polynomial in η, whose coefficients are finite and non-zero as $n \to \infty$. We define the function $g(.)$ as

$$g(x) = x + \frac{h}{n}s(x) + \left(\frac{h}{n}\right)^2 Q(x), \quad \text{with} \quad Q(x) = \frac{1}{4}\int\left(\frac{d}{dx}s(x)\right)^2 dx. \tag{C.2}$$

We have

$$Pr(g(\xi) \leq \eta) = o\left(\frac{h}{n}\right).$$

Proof It is straightforward to verify that $g(x)$ is strictly increasing, its first derivative being

$$1 + \frac{h}{n}\frac{ds(x)}{dx} + \frac{1}{4}\left(\frac{h}{n}\right)^2\left(\frac{ds(x)}{dx}\right)^2 = \left(1 + \frac{1}{2}\frac{h}{n}\frac{ds(x)}{dx}\right)^2.$$

Since $g(.)$ is monotonic

$$\begin{aligned} Pr(g(\xi) \leq \eta) &= Pr(\xi \leq g^{-1}(\eta)) \\ &= F_1(g^{-1}(\eta)) + \frac{h}{n}s(g^{-1}(\eta))f_1(g^{-1}(\eta)) + o\left(\frac{h}{n}\right). \end{aligned} \tag{C.3}$$

Now, by (C.2) we have

$$\begin{aligned} \eta &= g^{-1}\left(\eta + \frac{h}{n}s(\eta) + \left(\frac{h}{n}\right)^2 Q(\eta)\right) \\ &= g^{-1}(\eta) + \frac{h}{n}\frac{dg^{-1}(x)}{dx}|_{|x=\eta}s(\eta) + o\left(\frac{h}{n}\right), \end{aligned} \tag{C.4}$$

where the second equality follows by a standard Taylor expansion. We define $q = g^{-1}(x)$. Therefore,

$$\frac{dg^{-1}(x)}{dx}|_{x=\eta} = \left(\frac{dg(q)}{dq}\right)^{-1}|_{x=\eta} = 1 + O\left(\frac{h}{n}\right), \tag{C.5}$$

where the last equality follows by total differentiation of the function $g(.)$ and Taylor expansion. Collecting (C.4) and (C.5), we obtain

$$\eta = g^{-1}(\eta) + \frac{h}{n}s(\eta) + o\left(\frac{h}{n}\right)$$

and hence

$$g^{-1}(\eta) = \eta - \frac{h}{n}s(\eta) + o\left(\frac{h}{n}\right). \tag{C.6}$$

Finally, by substitution of (C.6) into (C.3) and using

$$F(g^{-1}(\eta)) = F(\eta) - \frac{h}{n}s(\eta)f(\eta) + o\left(\frac{h}{n}\right),$$

$$f(g^{-1}(\eta)) = f(\eta) + O\left(\frac{h}{n}\right), \qquad s(g^{-1}(\eta)) = s(\eta) + O\left(\frac{h}{n}\right),$$

we obtain

$$Pr(g(\xi) \leq \eta) \quad = \quad F(\eta) - \frac{h}{n}s(\eta)f(\eta) + \frac{h}{n}s(\eta)f(\eta) + o\left(\frac{h}{n}\right) = F(\eta) + o\left(\frac{h}{n}\right).$$

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models.* Kluwer Academic Publishers.

Anselin, L. (2001). Rao's score test in spatial econometrics, *Journal of Statistical Planning and Inference* **97**. 113-39.

Ansley, C.F., R.Kohn and T.S. Shively (1992). Computing $p-$ values for the generalized Durbin-Watson and other invariant test statistics. *J.of Econometrics* **54**, 277-300.

Burridge, P. (1980). On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society, Series B*, **42**, 107-108.

Cliff, A. and J.K. Ord, (1972). Testing for spatial autocorrelation among regression residuals. *Geographical Analysis* **4**, 267-84.

Davies, R.B. (1973). Numerical inversion of a characteristic function. *Biometrika* **60**, 415-417.

Davies, R.B. (1980). The distribution of a linear combination of χ^2 random variables. *Applied Statistics* **29**, 323-333.

DiCiccio, T.J. and J.P. Romano (1995). On bootstrap procedures for second-order accurate confidence limits in parametric models. *Statistica Sinica* **5**, 141-60.

DiCiccio, T.J. and B. Efron (1996). Bootstrap confidence intervals. *Statistical science* **11**, 189-228.

Efron, B. and R.J. Tibshirani (1993). *An introduction to the bootstrap.* London:Chapman and Hall.

Ghazal, G.A. (1996) Recurrence formula for expectation of products of quadratic forms. *Statistics and Probability Letters* **27**, 101-9.

Hall,P.(1992). *The bootstrap and Edgeworth expansion.* Springer-Verlag.

Hall,P.(1992). On the removal of skewness by transformation. *Journal of the Royal Statistical Society. Series B* **54**, 221-228.

Harris,P. (1985). An asymptotic expansion for the null distribution of the efficient score statistic. *Biometrika* **72**, 653-659.

Horn, R.A. and C.R. Johnson (1985). *Matrix Analysis.* New York: Cambridge University Press.

Imhof,J.P.(1961).Computing the distribution of quadratic forms in normal variables. *Biometrika* **48**, 266-83.

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.

Lu, Z.H. and M.L.King (2002). Improving the numerical techniques for computing the accumulated distribution of a quadratic form in normal variables. *Econometric Reviews* **21**, 149-165.

Moran, P.A.P. (1950). A test for the serial dependence of residuals. *Biometrika* **37**, 178-81.

Paparoditis, E. and D.N. Politis (2005).Bootstrap hypothesis testing in regression models. *Statistics & Probability Letters* **74**, 356-365.

Pinkse, J. (2004). Moran-flavoured tests with nuisance parameters: examples. In Anselin, L., Florax, R.G.M., Rey, S. (eds.). *Advances in Spatial Econometrics: Methodology, Tools and Applications.* Springer-Verlag, Berlin, 66-77.

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

Yanagihara, H. and K. Yuan (2005). Four improved statistics for contrasting means by correcting skewness and kurtosis. *British journal of mathematical and statistical psychology* **58**, 209-37.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
chi square	0.0320	0.0360	0.0380	0.0370
Edgeworth	0.0400	0.0390	0.0410	0.0420
transformation	0.0450	0.0480	0.0460	0.0480
mean-variance correction	0.0350	0.0370	0.0410	0.0420
bootstrap	0.0540	0.0460	0.0470	0.0530

Table 1: Empirical sizes of the tests of H_0 in (1.2) for model (1.1) when the sequence h is divergent. The reported values have to be compared with the nominal 0.05

.

	$m = 5$ $r = 8$	$m = 5$ $r = 20$	$m = 5$ $r = 40$	$m = 5$ $r = 80$
chi square	0.0340	0.0360	0.0370	0.0370
Edgeworth	0.0410	0.0420	0.0470	0.0480
transformation	0.0340	0.0450	0.0480	0.0500
mean-variance correction	0.0410	0.0430	0.0460	0.0520
bootstrap	0.0630	0.0520	0.0510	0.0520

Table 2: Empirical sizes of the tests of H_0 in (1.2) for model (1.1) when the sequence h is bounded. The reported values have to be compared with the nominal 0.05.

	$m = 8$ $r = 5$	$m = 12$ $r = 8$	$m = 18$ $r = 11$	$m = 28$ $r = 14$
chi square	0.0230	0.0270	0.0360	0.0270
Edgeworth	0.0440	0.0480	0.0460	0.0470
transformation	0.0550	0.0490	0.0470	0.0490
mean-variance correction	0.0300	0.0340	0.0320	0.0380
bootstrap	0.0450	0.0520	0.0560	0.0510

Table 3: Empirical sizes of the tests of H_0 in (1.2) for model (3.1) when the sequence h is divergent. The reported values have to be compared with the nominal 0.05

.

	$m = 5$ $r = 8$	$m = 5$ $r = 20$	$m = 5$ $r = 40$	$m = 5$ $r = 80$
chi square	0.0250	0.0350	0.0380	0.0360
Edgeworth	0.0310	0.0440	0.0470	0.0520
transformation	0.0330	0.0420	0.0470	0.0520
mean-variance correction	0.0270	0.0440	0.0460	0.0520
bootstrap	0.0430	0.0470	0.0480	0.0480

Table 4: Empirical sizes of the tests of H_0 in (1.2) for model (3.1) when the sequence h is bounded. The reported values have to be compared with the nominal 0.05.

		$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$
$m = 8$	Edgeworth	1.9334	2.4403	3.0715
$r = 5$	Imhof	1.8620	2.3250	2.9000
$m = 12$	Edgeworth	1.8925	2.3722	2.9658
$r = 8$	Imhof	1.8430	2.3100	2.8850
$m = 18$	Edgeworth	1.8668	2.3294	2.8994
$r = 11$	Imhof	1.8310	2.2880	2.8550
$m = 28$	Edgeworth	1.8482	2.2985	2.8514
$r = 14$	Imhof	1.8200	2.2700	2.8250

Table 5: Edgeworth-corrected and Imhof's α-quantiles of the cdf of T in when h is divergent.

		$\alpha = 95\%$	$\alpha = 97.5\%$	$\alpha = 99\%$
$m = 5$	Edgeworth	1.8357	2.2777	2.8191
$r = 8$	Imhof	1.7840	2.1920	2.6800
$m = 5$	Edgeworth	1.7656	2.1609	2.6379
$r = 20$	Imhof	1.7450	2.1280	2.5850
$m = 5$	Edgeworth	1.7303	2.1021	2.5465
$r = 40$	Imhof	1.7200	2.0860	2.5200
$m = 5$	Edgeworth	1.7053	2.0605	2.4819
$r = 80$	Imhof	1.7010	2.0530	2.4730

Table 6: Edgeworth-corrected and Imhof's α-quantiles of the cdf of T when h is bounded.