Improved Test Statistics for Pure Spatial Autoregressive Models

Francesca Rossi *
The London School of Economics

November 9, 2010

Abstract

This paper develops more accurate tests for lack of spatial correlation than ones based on the usual central limit theorem. We test nullity of the lag parameter in a pure spatial autoregression based on least squares and Gaussian maximum likelihood estimates. In each case, depending on assumptions on the spatial weight matrix, the rate of convergence of the estimate can be slower than $\sqrt{n}$, where n is sample size. Correspondingly, the error in the normal approximation can be larger than the usual parametric order. This provides particularly strong motivation for employing instead refined statistics which entail closer approximations. These are based on (formal) Edgeworth expansions. In Monte Carlo simulations we demonstrate that the new tests (and one based on a bootstrap, which is expected to have similar properties) outperform one based on the usual normal approximation in small and moderate samples. The new tests are also applied in two empirical examples.

JEL classification: C12;C21
Keywords: Spatial independence, Edgeworth expansions, Bootstrap

*Email address: f.rossi@lse.ac.uk

1 Introduction

Spatial autoregressive (SAR) models offer a useful framework for describing observations that are recorded on irregularly-spaced points. In SAR models the notion of irregular spacing is embodied in an $n \times n$ weight matrix, denoted W_n henceforth, which needs to be chosen by the practitioner. In most of the applications, the $i-j$th component of W_n is defined in terms of the inverse of an economic distance between units i and j. An economic distance is a very broad concept which includes the usual geographic distance as a very special case. A common example of distance which does not have a geographic interpretation is the difference in household income. Since SAR models proved to be a simple yet flexible specification, they are extensively used in empirical applications (see e.g. Arbia (2006)).

In this paper we assume that for some scalar $\lambda \in (-1, 1)$ the data follow the pure SAR model, i.e.

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

where Y_n is an $n \times 1$ vector of observations. We denote ϵ_n an $n \times 1$ vector of unobservable random variables that are assumed to be normal and independent and identically distributed (i.i.d.).

A major branch of the spatial literature has focused on testing for spatial independence in SAR models when also a set of exogenous regressors is present, i.e. when the data follow the so called mixed SAR. The asymptotic properties of the estimates and related test statistics for mixed SAR models have been widely considered, see e.g. Anselin (2001), Kelejian and Prucha (2001), Lee (2002), Lee (2004). For sake of clarity, we should mention that often in the spatial econometric literature "spatial independence" is used as a synonym for "lack of spatial correlation", though these concepts are in general identical only under Gaussianity.

However, the problem of testing for the lack of spatial correlation in model (1.1) has received relatively little attention. Although it may appear to be only a particular case of the mixed SAR when none of the exogenous regressors is relevant, making inference on λ in model (1.1) can pose particular difficulties. The rate of convergence of the estimate of λ can be slower than the usual $\sqrt{n}$, depending on the choice of W_n (see, e.g. Lee (2004)). Consequently, the error when the approximation for the distribution is based on the central limit theorem can be larger than the usual $1/\sqrt{n}$ and therefore, the finite sample performance of standard tests based on the normal approximation can be poor.

Procedures based on Gaussian maximum likelihood estimate (MLE) for testing the nullity of λ in model (1.1) have been developed by Cliff and Ord (1972) and have been broadly considered. For a survey we refer to Anselin (1988). In addition, Lagrange multiplier tests have received extensive interest due to their computational simplicity, starting from the early work by Moran (1950) (see e.g. Anselin (2001) for a comprehensive survey). However, little attention has been dedicated to the relatively poor finite sample performances of such tests and the consequent need for small sample corrections when the data follow model (1.1).

An exception can be found in Robinson (2008), where finite sample corrections are derived for a general class of statistics which includes as a special case the Lagrange multiplier test for spatial independence in model (1.1).

In this paper we derive new tests for spatial independence which prove to be more accurate than those based on the usual central limit theorem. More specifically, we are interested in testing the null hypothesis H_0 against the alternative H_1, where those are defined as

$$H_0 : \lambda = 0 \qquad H_1 : \lambda > 0 \ \ (\lambda < 0), \tag{1.2}$$

when λ in model (1.1) is estimated either by ordinary least squares (OLS) or by the MLE.

It is known (Lee (2002)) that the OLS estimate of λ in model (1.1) is inconsistent when $\lambda \neq 0$, but it converges to zero in probability under H_0. Although this case is very limited when the interest is estimation, it becomes crucial when one focuses on testing. However, under H_0, the rate of convergence of the OLS estimate of λ might be slower that the parametric $\sqrt{n}$, depending on assumptions on W_n. On the other hand, the MLE of λ is consistent for every value of λ, although, again, the rate of convergence might be slower than $\sqrt{n}$ (Lee (2004)).

As previously mentioned, when the rate of convergence of the estimate is slower than $\sqrt{n}$, the error when the normal approximation is used can be larger than the usual $1/\sqrt{n}$. Our new tests are based on refined t-statistics, whose cumulative distribution functions (cdf) are closer to the normal than those of the standard statistics and therefore entail better approximations. Alternatively, we show that inference based on standard statistics can be improved by considering more accurate approximations for critical values than ones of the normal cdf.

The corrected testing procedures are derived from formal Edgeworth expansions of the cdf of the standard OLS and MLE t-statistics under H_0. Edgeworth expansion of the cdf is a well known means of improving the accuracy of the normal approximation. Specifically, the first term of the expansion corresponds to the standard normal cdf while later terms are of increasingly smaller order and improve on the approximation when only a small/moderate sample is available. If the rate of convergence of the estimate is slower than the parametric $\sqrt{n}$, as can be the case with (1.1), the inclusion of higher order terms is even more crucial, such terms being larger than those appearing in the expansion when the rate of convergence is $\sqrt{n}$. As common in the literature on higher-order expansions, Gaussianity is assumed in our derivation. For a comprehensive review of derivations of formal Edgeworth expansions, we refer to Hall (1992). Moreover, the work by Taniguchi (1991) is a very useful reference when dealing with expansion of the cdf of implicitly defined estimates such as MLE.

We investigate the finite sample performance of the new tests by means of Monte Carlo simulations. Our results confirm that the inclusion of the second order term of the Edgeworth expansion in the approximation of the cdf of the null statistics leads to substantial improvements. Moreover, the performance of the new tests is compared with one based on a bootstrap. Theoretically, it is established (e.g. Hall (1992)) that a bootstrap procedure may achieve the first Edgeworth correction and hence the results should be comparable. In our simulation study we show that this is indeed the case.

Finally, the new tests are applied in two small empirical examples. Although both examples are intended for illustrative purposes only and do not aim to provide any definite conclusions regarding the empirical issues involved, they do provide further evidence that the refined tests should be preferable to standard ones when the sample is small/moderate.

The paper is organised as follows. In Sections 2 and 3 we develop the refined tests when λ is estimated by OLS and MLE, respectively. The proofs relative to both Sections 2 and 3 are left to appendices. In Section 4 we report and discuss the results of the Monte Carlo simulations, while in Section 5 we apply the new tests in two empirical examples.

2 OLS estimation: Edgeworth-corrected critical values and corrected statistics

We suppose that model (1.1) holds and we are interested in testing H_0 in (1.2). In order to avoid cumbersome notation, we drop reference to the sample size in Y_n, ϵ_n and W_n, i.e. we write $Y = Y_n$, $\epsilon = \epsilon_n$ and $W_n = W$. The OLS estimate of λ in model (1.1) is defined as

$$\hat{\lambda} = \frac{Y'W'Y}{Y'W'WY},$$

where the prime denotes transposition. As previously mentioned, $\hat{\lambda}$ converges in probability to zero under H_0.

We introduce the following assumptions:

Assumption 1 $\lambda = 0$, i.e. H_0 is true.

Assumption 2 The elements of ϵ are i.i.d., normally distributed with mean zero and variance σ^2.

Assumption 3 For all n, $w_{ii} = 0$ $i = 1, ..., n$, where w_{ij} is the $i-j$th element of W.

Assumption 4 W is row-normalized, so that the elements of each row sum to one.

Assumption 5 a Uniformly in i and j, $w_{ij} = O(1/h_n)$, where $\{h_n\}$ is a positive sequence that can be divergent or bounded and such that $h_n/n \to 0$ as $n \to \infty$. Moreover, W is uniformly bounded in row and column sums in absolute value.

5 b The limits

$$\lim_{n\to\infty} \frac{h_n}{n} tr(W'W), \quad \lim_{n\to\infty} \frac{h_n}{n} tr(WW'W), \quad \lim_{n\to\infty} \frac{h_n}{n} tr((W'W)^2),$$
$$\lim_{n\to\infty} \frac{h_n}{n} tr(W^2), \quad \lim_{n\to\infty} \frac{h_n}{n} tr(W^3) \tag{2.1}$$

are non-zero.

Assumptions 3 and 4 provide the main features of the weight matrix (Lee (2002), Lee (2004)). Under Assumptions 3-4 and Assumption 5a, the limits displayed in (2.1) exist and are finite by Lemma 1, reported in Appendix C.1. Thus, the content of Assumption 5b is that such limits are also non-zero.

Let $\Phi(z)$ and $\phi(z)$ be the cdf and the probability density function (pdf) of a standard normal random variable, respectively, while $g^{(i)}$ denotes the ith derivative of the function g. Let ζ be any finite real number.

Theorem 1 Let Assumptions 1-5 hold. The cdf of $\hat{\lambda}$ admits the formal third order Edgeworth expansion

$$\begin{aligned}
\Pr(a\hat{\lambda} \le \zeta) &= \Phi(\zeta) + 2a^{-1}b_1\zeta^2\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) \\
&- (a^{-2}b_2 - 6a^{-2}b_1^2)\zeta^3\phi(\zeta) + 2a^{-2}b_1^2\zeta^4\Phi^{(2)}(\zeta) \\
&- \frac{\kappa_3^c}{3}a^{-1}b_1\zeta^2\Phi^{(4)}(\zeta) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + O\left(\frac{h_n}{n}\right)^{3/2}, \qquad (2.2)
\end{aligned}$$

where

$$a = \frac{tr[W'W]}{(tr[W'W + W^2])^{1/2}}, \quad b_1 = \frac{tr[WW'W]}{tr[W'W + W^2]}, \quad b_2 = \frac{tr[(W'W)^2]}{tr[W'W + W^2]}.$$

Moreover

$$\kappa_3^c \sim \frac{2tr(W^3) + 6tr(W'W^2)}{(tr(W'W + W^2))^{3/2}}$$

and

$$\kappa_4^c \sim \frac{6tr(W^4) + 24tr(W'W^3) + 12tr((WW')^2) + 6tr(W^2W'^2)}{(tr(W'W + W^2))^2},$$

where $\sim$ denotes a rate.

The proof of Theorem 1 is in Appendix A.

Under Assumption 5 we have

$$a^{-1}b_1 \sim \left(\frac{h_n}{n}\right)^{1/2}, \quad a^{-2}b_2 \sim \frac{h_n}{n}, \quad \kappa_3^c \sim \left(\frac{h_n}{n}\right)^{1/2}, \quad \kappa_4^c \sim \frac{h_n}{n},$$

and therefore

$$2a^{-1}b_1\zeta^2\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) \sim \left(\frac{h_n}{n}\right)^{1/2},$$

$$-(a^{-2}b_2 - 6a^{-2}b_1^2)\zeta^3\phi(\zeta) + 2a^{-2}b_1^2\zeta^4\Phi^{(2)}(\zeta) - \frac{\kappa_3^c}{3}a^{-1}b_1\zeta^2\Phi^{(4)}(\zeta) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) \sim \frac{h_n}{n}.$$

Since $a \sim (n/h_n)^{1/2}$ from Assumption 5, obviously when the sequence h_n is divergent the rate of convergence of $Pr(a\hat{\lambda} \leq \zeta)$ to the standard normal cdf is slower than the usual $\sqrt{n}$.

It must be stressed that the expansion in (2.2) is formal and hence the order of the remainder can only be conjectured by the rate of the coefficients.

From the expansion (2.2) we can obtain Edgeworth-corrected critical values. We denote w_α and z_α the $\alpha-$quantiles of the null statistic $a\hat{\lambda}$ and the standard normal cdf, respectively. By inversion of (2.2) we can obtain an asymptotic series for w_α, i.e.

$$w_\alpha = z_\alpha + p_1(z_\alpha) + p_2(z_\alpha) + \ldots.., \tag{2.3}$$

where $p_1(z_\alpha)$ and $p_2(z_\alpha)$ are polynomials of orders $(h_n/n)^{1/2}$ and h_n/n, respectively. Both $p_1(z_\alpha)$ and $p_2(z_\alpha)$ can be determined using the identity $\alpha = Pr(a\hat{\lambda} \leq w_\alpha|H_0)$ and the asymptotic expansion given in Theorem 1. Even though the procedure can be extended to higher orders, for algebraic simplicity we focus on the first order Edgeworth correction and therefore only $p_1(z_\alpha)$ has to be determined. For convenience, we report the second order Edgeworth expansion

$$\Pr(a\hat{\lambda} \leq \zeta) = \Phi(\zeta) + 2a^{-1}b_1\zeta^2\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) + O\left(\frac{h_n}{n}\right). \tag{2.4}$$

Let $H_j(x)$ be the $j-$th Hermite polynomial. From (2.4) and the property $(-d/dx)^j\Phi(x) = -H_{j-1}(x)\phi(x)$, we have

$$\begin{aligned}
\alpha &= \Pr(a\hat{\lambda} \leq w_\alpha|H_0) \\
&= \Phi(w_\alpha) - (\frac{\kappa_3^c}{3!}H_2(w_\alpha) - 2a^{-1}w_\alpha^2 b_1)\phi(w_\alpha) + O\left(\frac{h_n}{n}\right).
\end{aligned}$$

Moreover, expanding w_α according to (2.3) and dropping negligible terms, we write

$$\begin{aligned}
\alpha &= \Pr(a\hat{\lambda} \leq w_\alpha|H_0) \\
&= \Phi(z_\alpha) + p_1(z_\alpha)\phi(z_\alpha) - (\frac{\kappa_3^c}{3!}H_2(z_\alpha) - 2a^{-1}z_\alpha^2 b_1)\phi(z_\alpha) + O\left(\frac{h_n}{n}\right) \\
&= \alpha + p_1(z_\alpha)\phi(z_\alpha) - (\frac{\kappa_3^c}{3!}H_2(z_\alpha) - 2a^{-1}z_\alpha^2 b_1)\phi(z_\alpha) + O\left(\frac{h_n}{n}\right), \qquad (2.5)
\end{aligned}$$

where the second equality follows by the Taylor expansion of $\Phi(w_\alpha)$ around z_α.

For the identity displayed in (2.5) to hold, we require

$$p_1(z_\alpha) = \frac{\kappa_3^c}{3!} H_2(z_\alpha) - a^{-1} b_1 z_\alpha^2$$

and hence the expansion for w_α becomes

$$w_\alpha = z_\alpha + \frac{\kappa_3^c}{3!} H_2(z_\alpha) - 2a^{-1} b_1 z_\alpha^2 + O\left(\frac{h_n}{n}\right). \tag{2.6}$$

The size of the test of H_0 in (1.2) obtained with the usual approximation of w_α by z_α, that is

$$Pr(a\hat{\lambda} > z_\alpha | H_0), \tag{2.7}$$

can be compared with the one obtained using the Edgeworth correction as given in (2.6), i.e.

$$Pr(a\hat{\lambda} > z_\alpha + \frac{\kappa_3^c}{3!} H_2(z_\alpha) - 2a^{-1} b_1 z_\alpha^2 | H_0). \tag{2.8}$$

We notice that when z_α is used to approximate w_α, the error has order $(h_n/n)^{1/2}$, while it is reduced to (h_n/n) when the Edgeworth-corrected critical value is used.

Rather than corrected critical values, we can consider the Edgeworth-corrected test statistic. Since

$$\Phi^{(3)}(\zeta) = H_2(\zeta)\phi(\zeta) = (\zeta^2 - 1)\phi(\zeta),$$

the two equivalent representation of (2.4) can be written

$$\begin{aligned} \Pr(a\hat{\lambda} \le \zeta) &= \Phi(\zeta) + (2a^{-1} b_1 \zeta^2 - \frac{\kappa_3^c}{3!}(\zeta^2 - 1))\phi(\zeta) + O\left(\frac{h_n}{n}\right) \\ &= \Phi(\zeta + 2a^{-1} b_1 \zeta^2 - \frac{\kappa_3^c}{3!}(\zeta^2 - 1)) + O\left(\frac{h_n}{n}\right). \end{aligned}$$

When the transformation

$$v(\zeta) = \zeta + 2a^{-1} b_1 \zeta^2 - \frac{\kappa_3^c}{3!}(\zeta^2 - 1) = \zeta + (2a^{-1} b_1 - \frac{\kappa_3^c}{3!})\zeta^2 + \frac{\kappa_3^c}{3!}$$

is monotonic, we can write

$$\Pr(a\hat{\lambda} + (2a^{-1} b_1 - \frac{\kappa_3^c}{3!})(a\hat{\lambda})^2 + \frac{\kappa_3^c}{3!} \le \zeta) = \Phi(\zeta) + O\left(\frac{h_n}{n}\right)$$

and make inference on λ based on the corrected statistic $v(a\hat{\lambda})$. The function $v(\zeta)$ is increasing when $\zeta \ge -1/(2(2a^{-1} b_1 - \kappa_3^c/3!))$ and some numerical work shows that the latter is verified for ζ close to zero. Hence, when the sample size is large enough so that $a\hat{\lambda}$ is close to zero, the transformation $v(a\hat{\lambda})$ is strictly (locally) increasing. However this does not hold in general and therefore a cubic transormation that does not affect the remainder but such that the resulting function is strictly increasing over the whole domain should be considered. A suitable transformation is in Hall (1992) or, in a more general case, Yanagihara et al (2005):

$$g(\zeta) = v(\zeta) + Q(\zeta), \qquad \text{with} \qquad Q(\zeta) = \frac{1}{3}\left(2a^{-1}b_1 - \frac{\kappa_3^c}{3!}\right)^2 \zeta^3.$$

Indeed, it can be shown (Yanagihara et al (2005)) that for a statistic T that admits the general expansion

$$Pr(T \leq x) = \Phi(x) + p_1(x)\phi(x) + O\left(\frac{h_n}{n}\right),$$

where $p_1(x) \sim \sqrt{h_n/n}$, the transformation

$$g(T) = T + p_1(T) + \frac{1}{4}Q(T) \quad \text{with} \quad Q(x) = \int \left(\frac{d}{dx}p_1(x)\right)^2 dx$$

is strictly increasing and does not affect higher order terms, i.e.

$$Pr(g(T) \leq x) = \Phi(x) + O\left(\frac{h_n}{n}\right).$$

It is straightforward to verify that in the present case the function $g(\zeta)$ is strictly increasing for every ζ, its first derivative being $(1 + (2a^{-1}b_1 - (\kappa_3^c/3!)\zeta))^2$.

We can therefore compare the size of the test of H_0 in (1.2) based on such corrected statistic, i.e.

$$Pr(g(a\hat{\lambda}) > z_\alpha | H_0) \tag{2.9}$$

with the standard (2.7). As previously mentioned, the error when the standard statistic is used has order $\sqrt{h_n/n}$, while it is reduced to h_n/n when considering the corrected variant.

3 MLE estimation: Edgeworth-corrected critical values and corrected statistics

In Section 2 we focused on the test of H_0 in (1.2) and derived Edgeworth-corrected critical values and corrected test statistics when λ is estimated by OLS. However, as outlined, $\hat{\lambda}$ is inconsistent when $\lambda \neq 0$. In this section we focus on the test of H_0 in (1.2) using the MLE for λ, denoted $\tilde{\lambda}$ henceforth. Since $\tilde{\lambda}$ is consistent for every value of λ in model (1.1), in principle we could extend the results presented in this section to test the more general hypothesis

$$H_0 : \lambda = \lambda_0$$

against the alternative

$$H_1 : \lambda > \lambda_0 \quad (< \lambda_0),$$

for any fixed λ_0. Although the procedure would be identical, when $\lambda_0 \neq 0$ the algebraic burden would increase dramatically. In addition, $\lambda = 0$ is probably the most interesting value one wishes to test for. Therefore, it seems reasonable to focus only on the test of H_0 as specified in (1.2).

We define $S(\lambda) = I - \lambda W$, where I denotes the $n \times n$ identity matrix. Henceforth, when the dimension of the identity matrix is other than n, we will specify it with a subscript, i.e. $I_{\cdot}$.

The Gaussian log-likelihood function for model (1.1) is given by

$$l(\lambda, \sigma^2) = -\frac{n}{2} ln(2\pi) - \frac{n}{2} ln\sigma^2 + ln|S(\lambda)| - \frac{1}{2\sigma^2} Y' S(\lambda)' S(\lambda) Y. \tag{3.1}$$

It is known that, given λ, the MLE of σ^2 is

$$\tilde{\sigma}^2(\lambda) = \frac{1}{n} Y' S(\lambda)' S(\lambda) Y. \tag{3.2}$$

Hence

$$\tilde{\lambda} = \arg \max_{\lambda \in \Lambda} \; l(\lambda, \tilde{\sigma}^2(\lambda)),$$

where Λ denotes a suitable closed and bounded subset of $\Re$ that includes the true value $\lambda = 0$.

A brief remark regarding Λ is necessary. Under Assumption 4 the Jacobian $|S(\lambda)|$ in (3.1) is positive for every value of λ in the set $(-1, 1)$. Hence, Λ can be chosen as any closed subset of $(-1, 1)$. Moreover, existence of $S^{-1}(\lambda)$ can be easily established when W is row normalized and Λ is any closed subset of $(-1, 1)$, since

$$S^{-1}(\lambda) = \sum_{i=0}^{\infty} (\lambda W)^i. \tag{3.3}$$

For details about the latter result we refer e.g. to Horn and Johnson (1985), pp 296-301.
We have the following result

Theorem 2 Let Assumptions 1-5 hold. The cdf of $\tilde{\lambda}$ admits the formal second order Edgeworth expansion

$$\begin{aligned} Pr(\sqrt{\frac{n}{h_n}} \tilde{a} \tilde{\lambda} \leq \zeta) &= \Phi(\zeta) + \left(2 \left(\frac{h_n}{n} \right)^{3/2} \frac{tr(WW'W)}{\tilde{a}^3} + \left(\frac{h_n}{n} \right)^{3/2} \frac{tr(W^3)}{\tilde{a}^3} \right) \phi(\zeta) \\ &- \frac{\tilde{\kappa}_3^c}{3!} \Phi^{(3)}(\zeta) + o\left(\sqrt{\frac{h_n}{n}} \right), \end{aligned}$$

or equivalently

$$
\begin{aligned}
Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} \leq \zeta) &= = \Phi(\zeta) + \left(2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{\tilde{a}^3} + \left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{\tilde{a}^3}\right)\phi(\zeta) \\
&- \frac{\tilde{\kappa}_3^c}{3!}H_2(\zeta)\phi(\zeta) + o\left(\sqrt{\frac{h_n}{n}}\right), \qquad (3.4)
\end{aligned}
$$

where

$$\tilde{a} = \sqrt{\frac{h_n}{n}}\sqrt{tr(W^2) + tr(W'W)}$$

and

$$
\begin{aligned}
\tilde{\kappa}_3^c &\sim \frac{-4\left(\frac{h_n}{n}\right)^{3/2} tr(W^3) - 6\left(\frac{h_n}{n}\right)^{3/2} tr(WW'W)}{\left(\frac{h_n}{n}\right)^{3/2}(tr(W^2) + tr(W'W))^{3/2}} \\
&= -\left(\frac{h_n}{n}\right)^{3/2}\frac{4tr(W^3) + 6tr(WW'W)}{\tilde{a}^3} \sim \sqrt{\frac{h_n}{n}}.
\end{aligned}
$$

The proof of Theorem 2 is in Appendix B.

Under Assumption 5, we have

$$\left(2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{\tilde{a}^3} + \left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{\tilde{a}^3}\right)\phi(\zeta) - \frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(\zeta) \sim \sqrt{\frac{h_n}{n}}.$$

It should again be stressed that the expansion in (3.4) is formal and hence the order of the remainder can only be conjectured by the rate of the coefficients. Without considering validity issues, the error order $o(\sqrt{h_n/n})$ is the best one can conjecture. As reported in detail in Appendix B, several approximations are used to obtain the expansion (3.4), such as the Taylor expansion of the integrand function in the evaluation of the characteristic function and $E(A/B) \sim E(A)/E(B)$ in the derivation of the cumulants. Therefore, the order $o(\sqrt{h_n/n})$ is the sharpest we can conclude.

Since under Assumption 5 $\tilde{a}$ is finite and strictly positive for large n, as expected the rate of convergence of $Pr(\sqrt{n/h_n}\tilde{a}\tilde{\lambda} \leq \zeta)$ to the standard normal cdf is slower than the usual $\sqrt{n}$ when the sequence h_n is divergent.

From expansion (3.4), Edgeworth-corrected critical values and the corrected null statistic can be obtained. The derivation is very similar to that reported in Section 2 for the cdf of $a\hat{\lambda}$ and is omitted here. The size of the test of H_0 in (1.2) obtained with the usual standard normal approximation

$$Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > z_\alpha | H_0) \qquad (3.5)$$

can be compared with the one obtained when the Edgeworth-corrected critical value is used, that is

$$Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed}|H_0), \tag{3.6}$$

where

$$\tilde{t}^{Ed} = z_\alpha - \left(2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{\tilde{a}^3} + \left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{\tilde{a}^3}\right) + \frac{\tilde{\kappa}_3^c}{3!}H_2(z_\alpha).$$

As discussed in Section 2, when z_α is used to approximate the true quantile, we have an error of order $\sqrt{h_n/n}$, while the error is decreased to $o(\sqrt{h_n/n})$ when the Edgeworth-corrected critical value is used.

Finally, (3.5) can be compared with the size based on the corrected statistic, i.e.

$$Pr(\tilde{g}(\tilde{a}\tilde{\lambda}) > z_\alpha|H_0), \tag{3.7}$$

where

$$\tilde{g}(x) = x + 2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{\tilde{a}^3} + \left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{\tilde{a}^3} - \frac{\tilde{\kappa}_3^c}{3!}H_2(x) + \tilde{Q}(x),$$

and

$$\tilde{Q}(x) = \left(\frac{\tilde{\kappa}_3^c}{3!}\right)^2\frac{x^3}{3}.$$

As discussed in detail in Section 2, $\tilde{Q}(x)$ is a cubic term so that $\tilde{g}(x)$ is strictly increasing over the whole domain, but that does not affect the order of the remainder.

4 Bootstrap correction and simulation results

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

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

$$W = I_r \otimes B_m, \qquad B_m = \frac{1}{m-1}(l_m l_m' - I_m), \tag{4.1}$$

where r is the number of districts and m is the number of households in each district. We denote l_m 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 $h_n = m - 1$.

W in (4.1) is symmetric and hence

$$a = \frac{\sqrt{tr(W^2)}}{\sqrt{2}}, \qquad b_1 = \frac{tr(W^3)}{2tr(W^2)}, \qquad \kappa_3^c = \frac{2\sqrt{2}tr(W^3)}{(tr(W^2))^{3/2}},$$
$$\tilde{a} = \sqrt{\frac{h_n}{n}}\sqrt{2tr(W^2)}, \qquad \tilde{\kappa}_3^c = -\frac{5tr(W^3)}{\sqrt{2}(tr(W^2))^{3/2}}.$$

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 $\alpha = 95\%$.

A brief remark on W defined in (4.1) is necessary. It is straightforward to verify that Assumptions 3-4 are satisfied for this choice of W, whether h_n is bounded or divergent (that is, whether the number of households in each unit diverges or is bounded as n increases). Assumption 5a holds provided that $r \to \infty$, m being either divergent or bounded. Moreover we can verify that Assumption 5b holds, i.e.

$$\lim_{n\to\infty} \frac{h_n}{n} tr(W^i) \neq 0 \quad \text{for} \quad i = 2, 3, 4,$$

by observing that

$$\frac{h_n}{n} tr((I_r \otimes B_m)^i) = \frac{h_n}{n} tr(I_r) tr(B_m^i) = \frac{h_n}{n} r tr(B_m^i).$$

By standard linear algebra, B_m has one eigenvalue equal to 1 and the other $m-1$ equal to $-1/(m-1)$. Therefore

$$tr(B_m^i) = 1 + (m-1)\left(\frac{-1}{m-1}\right)^i$$

and hence

$$\begin{aligned} \frac{h_n}{n} tr((I_r \otimes B_m)^i) &= \frac{h_n}{n} r tr(B_m^i) = \frac{m-1}{rm} r \left(1 + \frac{(-1)^i}{(m-1)^{i-1}}\right) \\ &= \frac{m-1}{m}\left(1 + \frac{(-1)^i}{(m-1)^{i-1}}\right), \end{aligned}$$

which is non-zero in the limit whether m is divergent or bounded for $i = 2, 4$. When $i = 3$ and m is bounded, we require $m > 2$ (at least for large n) for the latter quantity to be non-zero.

The empirical sizes of the test of (1.2) based on the usual normal approximation are compared with the same quantities obtained with both the Edgeworth-corrected critical values and corrected test statistics. In addition, we consider the simulated sizes based on bootstrap critical values since it is well established that these may achieve the first Edgeworth correction and should then be comparable with the results obtained in Sections 2 and 3 (see e.g. Hall (1992) 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. When λ is estimated by OLS, the bootstrap critical values are obtained by the following algorithm:

Step 1 Given model (1.1), under H_0 we have $Y = \epsilon$, i.e. ϵ is observable.

Step 2 Under Assumption 2 we can use a parametric bootstrap, i.e. we construct B $n-$dimensional vectors whose components are independently generated from the empirical distribution $N(0, \hat{\sigma}^2)$, where $\hat{\sigma}^2 = \epsilon'\epsilon/n = Y'Y/n$. We denote ϵ_j^*, for $j = 1, ...B$, each of these vectors. Hence, we generate B pseudo-samples as $Y_j^* = \epsilon_j^*$ for $j = 1, ...B$. When the distribution of the disturbances is known, the parametric bootstrap proved to be more efficient than the usual procedure based on resampling the residuals with replacement (see e.g Hall (1992)).

Step 3 We obtain B bootstrap OLS null statistics as

$$Z_j = a\frac{Y_j^{*'} W' Y_j^*}{Y_j^{*'} W' W Y_j^*}, \qquad j = 1, ...B.$$

Step 4 The $\alpha-$percentile is computed as the value w_α^* which solves

$$\frac{1}{B}\sum_{j=1}^{B} 1(Z_j \leq w_\alpha^*) = \alpha,$$

where 1(.) denotes the indicator function.

Step 5 The size of the test of (1.2) when the bootstrap critical value is used is then

$$Pr(a\hat{\lambda} > w_\alpha^* | H_0). \tag{4.2}$$

Regarding Step 1, a remark is needed. When interested in testing, the bootstrap procedure when we impose H_0 to obtain the residuals (and then to generate the pseudo-data) gives results at least as good as the same algorithm without imposing H_0 (see Paparoditis and Politis (2005)). Some numerical work actually shows that, in the present case, the procedure where H_0 is imposed outperforms the other.

When λ is estimated by the MLE, the first two steps of the algorithm above are unchanged while Step 3-Step 5 are modified accordingly:

Step 3 We obtain B bootstrap MLE null statistics

$$\tilde{Z}_j = \tilde{a}\tilde{\lambda}_j^*, \quad j = 1, ...B,$$

where

$$\tilde{\lambda}_j^* = \arg\max_{\lambda \in \Lambda} l_j^*(\lambda)$$

and

$$l_j^*(\lambda) = -\frac{n}{2}(ln(2\pi) + 1) - \frac{n}{2}ln(\frac{1}{n}Y_j^{*'}S(\lambda)'S(\lambda)Y_j^*) + ln|S(\lambda)|.$$

Step 4 The $\alpha-$percentile is computed as the value $\tilde{w}_\alpha^*$ which solves

$$\frac{1}{B}\sum_{j=1}^{B} 1(\tilde{Z}_j \leq \tilde{w}_\alpha^*) = \alpha.$$

Step 5 The size of the test of (1.2) when the bootstrap critical value is used is then

$$Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{w}_\alpha^* | H_0). \tag{4.3}$$

In both procedures we set $B = 199$.

Table 1 displays the simulated values corresponding to (2.7), (2.8), (2.9) and (4.2) when h_n (that is, m in (4.1)) is divergent, while in Table 2 we report the simulated values corresponding to the same quantities when h_n is bounded. Moreover, Tables 3 and 4 display the simulated values corresponding to (3.5), (3.6), (3.7) and (4.3) when h_n is either divergent or bounded, respectively. All the values in Tables 1-4 have to be compared with the nominal 5%. For notational convenience, we denote

$$t^{Ed} = z_\alpha + \frac{\kappa_3^c}{3!}H_2(z_\alpha) - 2a^{-1}b_1 z_\alpha^2$$

and

$$\tilde{t}^{Ed} = z_\alpha - \left(2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{\tilde{a}^3} + \left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{\tilde{a}^3}\right) + \frac{\tilde{\kappa}_3^c}{3!}H_2(z_\alpha),$$

as already defined in Section 3.

(Tables 1-4 about here)

For both divergent and bounded h_n and both estimation methods, it is clear that the usual normal approximation does not work well in practice, since the simulated values for the size underestimate the nominal 5% for all sample sizes. On the other hand, the Edgeworth-corrected results seem to perform reasonably well. In particular, when λ is estimated by either OLS or MLE and for divergent h_n, the results obtained with the Edgeworth-corrected critical values exceed the target 0.05 for very small sample sizes, but the convergence to the nominal value appears to be fast. Indeed, such correction performs already quite well for small/moderate sample sizes such as $m = 18$, $r = 14$. When h_n is bounded the results exhibit the same pattern, but with the advantage of a faster rate of convergence to the nominal 5%. The simulated sizes based on the corrected statistics, instead, are very satisfactory also for very small sample sizes, whether h_n is divergent or bounded and λ is estimated by either OLS or MLE.

As expected, the results are similar whether λ is estimated by OLS or MLE. However, for divergent h_n and when considering the Edgeworth-corrected critical values, the results obtained when λ is estimated by MLE slightly outperform those based on the OLS estimate. Other than this case, the values appear to be comparable.

As previously mentioned, Edgeworth-corrected and bootstrap-based tests should have similar properties. From Tables 1-4 we see that this is indeed the case. Specifically, for all cases the bootstrap results outperform ones based on Edgeworth-corrected critical values when the sample is very small, but become comparable as the sample size increases. In case h_n is divergent and λ is estimated by OLS, the bootstrap results seem to outperform those obtained using Edgeworth-corrected critical values for all sample sizes, but the discrepancy is increasingly smaller as n increases. On the other hand, for all sample sizes and both estimation methods the bootstrap results appear to be very similar to ones based on the Edgeworth-corrected statistics, whether h_n is divergent or bounded.

In Figures 1 and 2 we plot the pdf obtained from the Monte Carlo simulation of the non-corrected OLS null statistic $a\hat{\lambda}$ and its corrected version $g(a\hat{\lambda})$, respectively, while in Figures 3 and 4 we plot the non-corrected and corrected pdf for the MLE null statistics, i.e. $\sqrt{n/h_n}\tilde{a}\tilde{\lambda}$ and $\tilde{g}(\sqrt{n/h_n}\tilde{a}\tilde{\lambda})$, respectively. We notice that both non-corrected statistics are skewed to the left but most of this skewness is removed when we consider the corrected versions.

(Figures 1-4 about here)

Finally, we investigate with a Monte Carlo simulation the power of both standard and corrected tests of

$$H_0 : \lambda = 0 \qquad H_1 : \lambda = \bar{\lambda}, \tag{4.4}$$

where $\bar{\lambda}$ is a fixed finite, positive, alternative value. Obviously, the same argument can be carried on with very minor modifications in case of a fixed, negative, alternative value. In Tables 5 and 6 we report the simulated quantities corresponding to $Pr(a\hat{\lambda} > z_\alpha|H_1)$, $Pr(a\hat{\lambda} > t^{Ed}|H_1)$ and $Pr(a\hat{\lambda} > w^*_\alpha|H_1)$, while in Tables 7 and 8 we report the same quantities in case λ is estimated by MLE, i.e. the empirical values of $Pr(\sqrt{n/h_n}\tilde{a}\tilde{\lambda} > z_\alpha|H_1)$, $Pr(\sqrt{n/h_n}\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed}|H_1)$

and $Pr(\sqrt{n/h_n}\tilde{a}\tilde{\lambda} > \tilde{w}^*_\alpha|H_1)$. We choose three different values of $\bar{\lambda}$, specifically $\bar{\lambda} = 0.1, 0.5, 0.8$.

The pattern of the results when λ is estimated either by OLS or MLE appears quite similar and the values are consistent with the empirical sizes reported in Tables 1-4.

(Tables 5-8 about here)

A further remark about testing of H_0 in (4.4) when λ is estimated by OLS is necessary at this stage. As previously mentioned, $\hat{\lambda}$ is inconsistent when $\lambda \neq 0$. Therefore, in case $plim\hat{\lambda} - \lambda < 0$ as $n \to \infty$, for some finite, strictly positive λ, we might have that, under H_1, $plim\hat{\lambda} = 0$ as $n \to \infty$ (obviously, for λ strictly negative the argument would be modified as: in case $plim(\hat{\lambda} - \lambda) > 0$ as $n \to \infty$ we might have that.....). In this case, the standard test of H_0 in (4.4) would be inconsistent. Clearly, if this were the case, the MLE would be preferable over the OLS estimation, but given the computational simplicity of the latter, it should not be dismissed before some further investigation.

It is quite straightforward to evaluate the sign of the probability limit of $\hat{\lambda} - \lambda$ for any particular choice of W. By Lemma 5 (reported in Appendix C.1), under Assumption 2 and for W given in (4.1), we have that the probability limit of $\hat{\lambda} - \lambda$ is finite and has the same sign of λ. It is worth mentioning that such limit can be computed similarly for any other choices of W. Moreover, for sake of generality, Lemma 5 can be proved without imposing Assumption 2. However, Assumption 2 has been assumed throughout the paper and is retained here for algebraic simplicity.

By Lemma 5, it is straightforward to show that, as $n \to \infty$, $Pr(a\hat{\lambda} > z_\alpha|H_1) \to 1$, $Pr(a\hat{\lambda} > t^{Ed}|H_1) \to 1$ and $Pr(g(a\hat{\lambda}) > z_\alpha|H_1) \to 1$.

5 Empirical examples

In this section we consider two small empirical examples where the corrections developed in Sections 2 and 3 are relevant. It should be stressed that these examples are intended for illustration only and do not aim to be exhaustive analyses of the issues involved. The main purpose of this section is to show how, in some empirical investigation, the conclusion of a test might be different if the small sample correction is taken into account.

5.1 The geography of happiness (Stanca (2009))

The main goal of the empirical work in Stanca (2009) is to investigate the spatial distribution of the effects of both income and unemployment on well-being for a sample of 81 countries. For the purpose of our example we focus on the income effects. Several specifications are considered, the three main ones being

$$P = \lambda WP + X\gamma + \epsilon, \tag{5.1}$$

$$P = \lambda W P + \epsilon, \tag{5.2}$$

$$P = X\gamma + \epsilon, \tag{5.3}$$

where $\epsilon \sim N(0, \sigma^2 I)$. P is the $n-$ dimensional vector of sensitivities of well-being to income in each country, X is a set of exogenous macroeconomic conditions, including GDP per capita, unemployment rate, government size and trade openness. W is the usual matrix of spatial weights and more details about the choice of W will be given below.

The components of P are clearly unobservable. A good proxy for each component of P is given by the estimate of a country-specific, micro-level linear model where well-being (denoted WB, henceforth) is regressed on income (denoted In, henceforth) as well as unemployment status, demographic factors, social conditions, personality traits and environmental characteristics. For notational convenience we denote Z the matrix of all the regressors other than income. Specifically, for each country we estimate the parameters of

$$WB = \beta_1 In + \beta_2 Z + u. \tag{5.4}$$

For each individual in the sample, WB (intended as life satisfaction) is a self reported number from 1 to 10 while income is measured by self reported deciles in the national distribution of income. The vector P contains the estimated values $\hat{\beta}_1$ of each country-specific regression. More details about the choice of variables, data sources, sample size for each country and estimation methods are in Stanca (2009) and a discussion about the specification of the micro analysis is beyond the scope of this example.

The results in Stanca (2009) indicate that by estimating λ in model (5.2) we detect the presence of spatial correlation. However, when the macroeconomic conditions are included among the regressors, such as in specification (5.1), the estimate of λ becomes insignificant, suggesting that the geographical correlation is mainly explained by similar underlying macroeconomics conditions in neighbouring countries. Therefore, either specification (5.2) or (5.3) can be appropriate, as the estimate of λ in model (5.2) should reflect the macroeconomics similarities among countries.

However, we notice that the estimates of the relevant components of γ in (5.3) are strongly significant (1% or 0.5% level), while the estimate for λ in specification (5.2) is barely significant at 5%. Given that specifications (5.2) and (5.3) should both be appropriate, in principle we would expect the estimates of the coefficients of the two specifications to be equally significant (at least roughly). Therefore, it can be useful to investigate whether an Edgeworth-corrected test gives a different result.

We start by considering only a sub-sample of 43 European countries. Since P in specification (5.2) is a vector of estimates and not actual data, some heterogeneity issues might be eliminated by considering only European countries. Indeed, we expect that the micro level analysis to obtain $\hat{\beta}_1$ does not exhibit strong structural differences across a sample of 43 European countries. On the other hand, when considering a broader sample, some systematic differences in the relationship among the country-specific variables might occur. In practice, a model as (5.4) might not be the correct specification for all countries, when such countries are very heterogeneous. In turn, when such differences occur, the reliability of $\hat{\beta}_1$ as proxies for the components of P is not clear. This problem might be reduced by considering only a sub-sample of less heterogeneous countries.

Since the dependent variable in (5.2) is a vector of proxies and not actual data, we acknowledge that the corrections derived in Sections 2 and 3 might not fully hold. In principle, we might be neglecting some relevant term arising from the approximation of the components of P by the estimates $\hat{\beta}_1$ in the Edgeworth expansions of the cdf of the null OLS and MLE statistics. However, at least for illustrative purpose, we think that a preliminary investigation of the effects of the inclusion of the small sample corrections derived in Sections 2 and 3 is worthwhile.

We construct W based on a contiguity criterion, i.e. $w_{ij} = 1$ if country i and country j share a border and $w_{ij} = 0$ otherwise.

(Tables 9-10 about here)

In Tables 9 and 10, we report the outcome of the tests of H_0 in (1.2) when λ is estimated with OLS and MLE, respectively. The actual values of statistics and critical values are reported in brackets. When λ is estimated by OLS, $\hat{\lambda}$ is only (barely) significant at 5%, while it becomes significant at 1% when corrected critical values are used. We notice that in case λ is estimated by MLE, the outcome of the test does not change when corrected critical values are used. This is a result that could be expected, to some extent. From the simulation work, the non-corrected results for the MLE appear to be slightly better than OLS in very small samples.

5.2 Crime rate and social capital (Buonanno et al. (2009))

The last example we present is based on a paper by Buonanno et al. (2009) and deals with crime rates in Italian provinces. In particular, this paper aims to investigate whether social capital, intended as civic norms and associational networks, affects the property crime rate at a provincial level. The 103 Italian provinces are especially suitable for this purpose since Italy displays significant provincial disparities despite being politically, ethically and religiously quite homogeneous. The literature about the influence of social capital on crime rate is broad and a survey is beyond the scope of this example. Similarly, for a discussion about the peculiar contribution of Buonanno et al. (2009), we refer to the paper.

For the purpose of our investigation, we consider the three following models

$$Y = \lambda WY + \epsilon, \tag{5.5}$$

$$Y = \lambda WY + \beta_1 SC + \beta_2 X + \delta D + \epsilon, \tag{5.6}$$

and

$$Y = \beta_1 SC + \beta_2 X + \delta D + \epsilon, \tag{5.7}$$

where $\epsilon \sim N(0, \sigma^2 I)$. Y is the $n-$dimensional vector of crime rates in each province. Each component of Y is obtained by dividing the reported crime rate at provincial level by the corresponding overall report rate at regional level. The dataset contains three sets of observations, regarding car thefts, robberies and general thefts rates. SC is the vector of social capital observations. The paper proposes four different measures of social capital, which are used separately, namely the number of recreational associations, voluntary associations, referenda turnout and blood donation. X is a matrix of exogenous regressors containing deterrence (such as the average length of judicial process and the crime specific clear up rate), demographic and socio-economic variables. In addition, X contains a measure of criminal association at provincial level. Finally, D is a matrix of geographical dummies to control for heterogeneity among the north, centre and south of the country. Our analysis is conducted with and without the inclusion of the geographical dummies and the results do not appear to vary significantly. The data pertain to 2002 or, when an average is considered, to the period 2000-2002.

In Buonanno et al. the parameters in model (5.6) are estimated for each crime type, with different variants of W and measures of social capital. The results of each estimation are reported in the paper. For our discussion, we observe that the estimate of λ in model (5.6) is insignificant in each of the regressions considered (or only barely significant at 10%, in few cases). However, when we estimate λ in model (5.5), we detect spatial correlation, suggesting that the effect of geographical contiguity is mostly taken into account by the regressors included in model (5.6). Hence, both models (5.5) and (5.7) seem to be appropriate and we expect the estimate of λ in model (5.5) to reflect the overall similarities across neighbouring provinces.

To investigate more specifically which are the main determinants of Y, we perform an OLS estimation of the parameters in model (5.7) and we observe that Y is strongly affected by the measure of criminal association (denoted CA, henceforth). Indeed, the estimate of the coefficient of CA is significant at 0.5% level. In turn, we expect that CA displays significant correlation across provinces and to confirm our conjecture we estimate the spatial parameter of the additional model

$$CA = \mu WCA + \epsilon, \tag{5.8}$$

where μ is a scalar parameter. As expected, the estimate of μ is strongly significant (0.5% level).

When the regressors are not included, such as when we consider model (5.5), we would expect to detect a similarly strong spatial correlation in the dependent variable. However, the estimate of λ in model (5.5) is only significant at 5% level.

As discussed for the previous example, we investigate whether we obtain a different outcome of the test of H_0 in (1.2) by including the small sample corrections derived in Sections 2 and 3. We report the results obtained for the robberies rates, W defined by a contiguity criterion and blood donation as a measure of social capital, although similar results can be derived for the other crime rates and alternative measures of social capital.

(Tables 11-12 about here)

The outcomes of the tests of H_0 in (1.2) when λ is estimated by OLS and MLE are reported in Tables 11 and 12, respectively. We notice that when the usual normal approximation is adopted, we are able to reject H_0 only at 5% level, λ being estimated by either OLS or MLE. Instead, when the Edgeworth correction is included, we are able to reject H_0 at 1% level when λ is estimated by OLS and at 0.5% level when λ is estimated by MLE. As is the case in the previous example, these results confirm those of the simulation work, i.e. for small/moderate sample sizes, the results obtained when λ is estimated by MLE slightly outperform those obtained by OLS estimation.

6 Conclusions

With the goal of improving normal-based inference, we have derived refined test statistics for lack of spatial correlation in a pure SAR model based on OLS and MLE estimates of the spatial parameter. The new tests are derived from the formal Edgeworth expansions of the cdf of the standard t-statistics under the null hypothesis.

We have motivated our analysis by observing that in a pure SAR model the rate of convergence of the estimate of the spatial parameter can be slower than the parametric $\sqrt{n}$, depending on assumptions on the weight matrix, entailing an error in the normal approximation which might be larger than the usual $1/\sqrt{n}$.

Monte Carlo simulations confirm that in finite samples the new tests outperform those based on the usual standard normal approximation and are comparable to one based on a bootstrap, which is expected to have similar properties. Moreover, we have applied the refined tests in two empirical examples. In such examples, we have shown that the inclusion of correction terms change the outcome of standard tests for spatial independence.

The framework presented in this paper could be extended along several directions. It should be stressed that our focus on t-tests was motivated by their relative simplicity, but we acknowledge that small sample refinements could be developed also for other statistics, e.g. likelihood ratio. Moreover, a similar, yet algebraically more complicated, derivation could be developed in case the dependent variable in model (1.1) is a vector of residuals. Such an extension might be useful, for instance, when we are interested in testing for spatial independence in the model

$$Y = X\beta + u \qquad \text{with} \quad u = \lambda W u + \epsilon, \qquad \epsilon \sim N(0, \sigma^2 I),$$

where X is a set of regressors, possibly stochastic. Finally, we should mention that the assumption of Gaussianity might be relaxed, at the expense of considerable extra complication in the derivation of the formal Edgeworth expansions.

A Appendix A: Proof of Theorem 1

The OLS estimate of λ is defined as

$$\hat{\lambda} - \lambda = \frac{Y'W'\epsilon}{Y'W'WY}$$

and therefore, under Assumption 1,

$$\hat{\lambda} = \frac{\epsilon'W'\epsilon}{\epsilon'W'W\epsilon}.$$

The cdf of $\hat{\lambda}$ under Assumption 1 can be written in terms of a quadratic form in ϵ, i.e.

$$Pr(\hat{\lambda} \leq x) = Pr(f \leq 0),$$

where

$$\begin{aligned} f &= \frac{1}{2}\epsilon'(C + C')\epsilon, \\ C &= W' - xW'W \end{aligned} \tag{A.1}$$

and x is any real number.

Under Assumption 2, the characteristic function of f can be derived as

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

where $det(A)$ denotes the determinant of a generic square matrix A, η_j are the eigenvalues of $(C + C')$ and $i = \sqrt{-1}$. From (A.2) the cumulant generating function of f is

$$\begin{aligned} \psi(t) &= -\frac{1}{2}\sum_{j=1}^{n} ln(1 - it\sigma^2\eta_j) = \frac{1}{2}\sum_{j=1}^{n}\sum_{s=1}^{\infty}\frac{(it\sigma^2\eta_j)^s}{s} \\ &= \frac{1}{2}\sum_{s=1}^{\infty}\frac{(it\sigma^2)^s}{s}\sum_{j=1}^{n}\eta_j^s = \frac{1}{2}\sum_{s=1}^{\infty}\frac{(it\sigma^2)^s}{s} tr((C + C')^s). \end{aligned} \tag{A.2}$$

From (A.2) the s-th cumulant can be derived as

$$\kappa_1 = \sigma^2 tr(C), \tag{A.3}$$

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

$$\kappa_s = \frac{\sigma^{2s} s!}{2} \frac{tr((C+C')^s)}{s}, s > 2. \tag{A.5}$$

Let

$$f^c = \frac{f - \kappa_1}{\kappa_2^{1/2}},$$

i.e. the centred and scaled version of f. The cumulant generating function of f^c can be written as

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

where

$$\kappa_s^c = \frac{\kappa_s}{\kappa_2^{s/2}}, \tag{A.6}$$

while the characteristic function of f^c is

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

Thus, 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

$$\begin{aligned} Pr(\hat{\lambda} \le x) &= Pr(f \le 0) = Pr(f^c \kappa_2^{1/2} + \kappa_1 \le 0) = Pr(f^c \le -\kappa_1^c) \\ &= \Phi(-\kappa_1^c) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(-\kappa_1^c) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(-\kappa_1') + ... \end{aligned} \tag{A.7}$$

From (A.3), (A.4) and (A.6) we have

$$\kappa_1^c = \frac{\sigma^2 tr(C)}{\sigma^2(\frac{1}{2}tr[(C+C')^2])^{1/2}},$$

where C is defined according to (A.1). The numerator of κ_1^c is

$$\sigma^2 tr(W) - \sigma^2 x tr(W'W) = -\sigma^2 x tr(W'W),$$

while the denominator of κ_1^c is σ^2 times

$$(\frac{1}{2}tr(C+C')^2)^{1/2} = (tr(W^2) + tr(WW') - 4x tr(WW'W) + 2x^2 tr[(W'W)^2])^{1/2}.$$

Thus

$$\begin{aligned}
\kappa_1^c &= \frac{-xtr(W'W)}{(tr(W^2)+tr(WW')-4xtr(WW'W)+2x^2tr[(W'W)^2])^{1/2}} \\
&= \frac{-xtr(W'W)}{(tr(W^2)+tr(WW'))^{1/2}(1-\frac{4xtr(WW'W)+2x^2tr[(W'W)^2]}{(tr(W^2)+tr(WW'))})^{1/2}}.
\end{aligned}$$

We choose $x=a^{-1}\zeta$, where

$$a=\frac{tr(W'W)}{(tr(W'W+W^2))^{1/2}}\sim\left(\frac{n}{h_n}\right)^{1/2}. \tag{A.8}$$

Moreover, we write

$$b_1=\frac{tr[WW'W]}{tr[W'W+W^2]} \tag{A.9}$$

and

$$b_2=\frac{tr[(W'W)^2]}{tr[W'W+W^2]}. \tag{A.10}$$

Now,

$$\begin{aligned}
\kappa_1^c &= \frac{-xtr[W'W]}{(tr[W'W+W^2])^{1/2}(1-4xb_1+2x^2b_2)^{1/2}}=\frac{-\zeta}{(1-4xb_1+2x^2b_2)^{1/2}} \\
&= -\zeta\left(1+2a^{-1}\zeta b_1-a^{-2}\zeta^2b_2+6a^{-2}\zeta^2b_1^2+O\left(\frac{h_n}{n}\right)^{3/2}\right) \\
&= -\zeta-2a^{-1}\zeta^2b_1+a^{-2}\zeta^3b_2-6a^{-2}b_1^2\zeta^3+O\left(\frac{h_n}{n}\right)^{3/2},
\end{aligned}$$

where the third equality follows by performing a standard Taylor expansion of the term $(1-4xb_1+2x^2b_2)^{-1/2}$, i.e.

$$(1-4xb_1+2x^2b_2)^{-1/2}=1+2xb_1-x^2b_2+6x^2b_1^2+O\left(\frac{h_n}{n}\right)^{3/2}.$$

From (A.8), (A.9), (A.10) and under Assumption 5 we have

$$2a^{-1}\zeta^2b_1\sim\left(\frac{h_n}{n}\right)^{1/2}, \qquad a^{-2}\zeta^3b_2\sim\frac{h_n}{n}, \qquad 6a^{-2}b_1^2\zeta^3\sim\frac{h_n}{n}.$$

Moreover, by Taylor expansion we obtain

$$\begin{aligned}
\Phi(-\kappa_1^c) &= \Phi\left(\zeta+2a^{-1}\zeta^2b_1-a^{-2}\zeta^3b_2+6a^{-2}\zeta^3b_1^2+O\left(\frac{h_n}{n}\right)^{3/2}\right) \\
&= \Phi(\zeta)+(2a^{-1}\zeta^2b_1-a^{-2}\zeta^3b_2+6a^{-2}\zeta^3b_1^2)\phi(\zeta) \\
&+ 2a^{-2}\zeta^4b_1^2\Phi^{(2)}(\zeta)+O\left(\frac{h_n}{n}\right)^{3/2}
\end{aligned} \tag{A.11}$$

and

$$\Phi^{(3)}(-\kappa_1^c) = \Phi^{(3)}(\zeta) + 2a^{-1}\zeta^2 b_1 \Phi^{(4)}(\zeta) + O\left(\frac{h_n}{n}\right). \tag{A.12}$$

Collecting (A.7), (A.11) and (A.12), the third order Edgeworth expansion of the cdf of $a\hat{\lambda}$ under Assumptions 1-5, becomes

$$\begin{aligned}
\Pr(a\hat{\lambda} \le \zeta) &= \Phi(\zeta) + 2a^{-1}b_1\zeta^2\phi(\zeta) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(\zeta) \\
&- (a^{-2}b_2 - 6a^{-2}b_1^2)\zeta^3\phi(\zeta) + 2a^{-2}b_1^2\zeta^4\Phi^{(2)}(\zeta) \\
&- \frac{\kappa_3^c}{3}a^{-1}b_1\zeta^2\Phi^{(4)}(\zeta) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(\zeta) + O\left(\frac{h_n}{n}\right)^{3/2},
\end{aligned}$$

where, from (A.4), (A.5) and (A.6),

$$\kappa_3^c = \frac{\sigma^6 tr[(C+C')^3]}{\sigma^6(\frac{1}{2}tr[(C+C')^2])^{3/2}} \sim \frac{2tr(W^3) + 6tr(W'W^2)}{(tr(W'W+W^2))^{3/2}}$$

and

$$\kappa_4^c = \frac{3\sigma^8 tr[(C+C')^4]}{\sigma^8(\frac{1}{2}tr[(C+C')^2])^2} \sim \frac{6tr(W^4) + 24tr(W'W^3) + 12tr((WW')^2) + 6tr(W^2W'^2)}{(tr(W'W+W^2))^2}.$$

B Appendix B: Proof of Theorem 2

We first introduce some notation that will be used throughout the proof. We write

$$l(\lambda) = l(\lambda, \tilde{\sigma}^2(\lambda)),$$

where $l(\lambda, \sigma^2)$ and $\tilde{\sigma}^2(\lambda)$ are defined in (3.1) and (3.2), respectively. Moreover,

$$Z^{(1)}(\lambda) = \sqrt{\frac{h_n}{n}}\frac{\partial l(\lambda)}{\partial \lambda}, \quad Z^{(2)}(\lambda) = \sqrt{\frac{h_n}{n}}\left(\frac{\partial^2 l(\lambda)}{\partial \lambda^2} - E\left(\frac{\partial^2 l(\lambda)}{\partial \lambda^2}\right)\right),$$

$$J(\lambda) = \frac{h_n}{n}\frac{\partial^3 l(\lambda)}{\partial \lambda^3}, \quad K(\lambda) = -\frac{h_n}{n}E\left(\frac{\partial^2 l(\lambda)}{\partial \lambda^2}\right), \quad \frac{\partial l(0)}{\partial \lambda} = \frac{\partial l(\lambda)}{\partial \lambda}|_{\lambda=0}.$$

Finally, we denote $O_e(.)$ an exact rate (in probability).

By (3.1) we have

$$\frac{\partial l(\lambda)}{\partial \lambda} = n\frac{(Y'WY - \lambda Y'W'WY)}{Y'S(\lambda)'S(\lambda)Y} - tr(S^{-1}(\lambda)W) \tag{B.1}$$

and

$$\frac{\partial^2 l(\lambda)}{\partial \lambda^2} = -n\frac{Y'W'WY}{Y'S(\lambda)'S(\lambda)Y} + 2n\frac{(\lambda Y'W'WY - Y'WY)^2}{(Y'S(\lambda)'S(\lambda)Y)^2} - tr(S^{-1}(\lambda)WS^{-1}(\lambda)W). \tag{B.2}$$

Therefore, under Assumption 1,

$$Z^{(1)}(0) = \sqrt{h_n}n\frac{\epsilon'W\epsilon}{\epsilon'\epsilon} \tag{B.3}$$

and

$$
\begin{aligned}
Z^{(2)}(0) &= \sqrt{\frac{h_n}{n}}\{-n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon} + 2n\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2 - tr(W^2) \\
&+ nE\left(\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right) - 2nE\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2 + tr(W^2)\} \\
&= \sqrt{\frac{h_n}{n}}\{-n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon} + 2n\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2 - tr(W^2) \\
&+ n\frac{E(\epsilon' W' W\epsilon)}{E(\epsilon'\epsilon)} - 2n\frac{E\left(\frac{1}{2}\epsilon'(W+W')\epsilon\right)^2}{E(\epsilon'\epsilon)^2} + tr(W^2)\} \\
&= \sqrt{\frac{h_n}{n}}\{-n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon} + 2n\frac{(\epsilon' W\epsilon)^2}{(\epsilon'\epsilon)^2} + tr(W'W) \\
&- \frac{1}{n}tr((W+W')^2)(1+\frac{2}{n})^{-1}\}, \qquad \text{(B.4)}
\end{aligned}
$$

since

$$E(\epsilon' W' W\epsilon) = \sigma^2 tr(W'W), \qquad \text{(B.5)}$$

$$E((\epsilon'(W+W')\epsilon)^2) = 2\sigma^4 tr((W+W')^2) \qquad \text{(B.6)}$$

and

$$E((\epsilon'\epsilon)^2) = \sigma^4(n^2+2n). \qquad \text{(B.7)}$$

The second equality in (B.4) follows because both the ratios

$$\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} = \frac{\frac{1}{2}\epsilon'(W+W')\epsilon}{\epsilon'\epsilon} \quad \text{and} \quad \frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}$$

are independent of their own denominator and therefore the expectation of the ratio is equal to the ratio of the expectations (see e.g. Jones (1987), who attributed the result to E.J.G. Pitman). Similarly, we have

$$J(0) = \frac{h_n}{n}\left(\frac{-6n\epsilon' W\epsilon\epsilon' W' W\epsilon}{(\epsilon'\epsilon)^2} + \frac{8n(\epsilon' W\epsilon)^3}{(\epsilon'\epsilon)^3} - 2tr(W^3)\right) \qquad \text{(B.8)}$$

and, using (B.5),(B.6), (B.7),

$$
\begin{aligned}
K(0) &= -\frac{h_n}{n}\left(-n\frac{E(\epsilon' W' W\epsilon)}{E(\epsilon'\epsilon)} + 2n\frac{E\left(\epsilon'\frac{1}{2}(W+W')\epsilon\right)^2}{E(\epsilon'\epsilon)^2}\right) + \frac{h_n}{n}tr(W^2) \\
&= \frac{h_n}{n}tr(W^2) + \frac{h_n}{n}tr(W'W) - \frac{h_n}{n^2}tr((W+W')^2)\left(1+\frac{2}{n}\right)^{-1}. \qquad \text{(B.9)}
\end{aligned}
$$

By Lemma 2, Lemma 3 and Lemma 4 (reported in Appendix C.1) $Z^{(1)}(0) = O_e(1)$, $Z^{(2)}(0) = O_p(1)$ and $J(0) = O_p(1)$, respectively. In addition, under Assumption 5, $K(0)$ is finite and positive for large n.

By the Mean Value Theorem we have

$$0 = \frac{h_n}{n}\frac{\partial l(\tilde{\lambda})}{\partial\lambda} = \frac{h_n}{n}\frac{\partial l(0)}{\partial\lambda} + \frac{h_n}{n}\frac{\partial^2 l(0)}{\partial\lambda^2}\tilde{\lambda} + \frac{1}{2}\frac{h_n}{n}\frac{\partial^3 l(0)}{\partial\lambda^3}\tilde{\lambda}^2 + \frac{h_n}{6n}\frac{\partial^4 l(\bar{\lambda})}{\partial\lambda^2}\tilde{\lambda}^3,$$

where $\bar{\lambda}$ is an intermediate point between $\tilde{\lambda}$ and 0. Therefore,

$$0 = \sqrt{\frac{h_n}{n}} Z^{(1)}(0) + \sqrt{\frac{h_n}{n}} Z^{(2)}(0)\tilde{\lambda} - K(0)\tilde{\lambda} + \frac{1}{2} J(0)\tilde{\lambda}^2 + \frac{h_n}{6n} \frac{\partial^4 l(\bar{\lambda})}{\partial \lambda^4} \tilde{\lambda}^3$$

and rearranging we obtain

$$\sqrt{\frac{n}{h_n}} \tilde{\lambda} = \frac{Z^{(1)}(0)}{K(0)} + \frac{Z^{(2)}(0)}{K(0)} \tilde{\lambda} + \frac{1}{2} \sqrt{\frac{n}{h_n}} \frac{J(0)}{K(0)} \tilde{\lambda}^2 + \frac{1}{6} \sqrt{\frac{h_n}{n}} \frac{\partial^4 l(\bar{\lambda})}{\partial \lambda^4} \tilde{\lambda}^3. \tag{B.10}$$

The first term of the RHS of (B.10) is $O_e(1)$, the second and the third are $O_p(\sqrt{h_n/n})$, since it is known that $\tilde{\lambda} = O_e(\sqrt{h_n/n})$ (see Lee (2004)) while $Z^{(2)}(0)$ and $J(0)$ are $O_p(1)$, by Lemma 3 and Lemma 4, respectively. The last term is $o_p(\sqrt{h_n/n})$ since $\bar{\lambda} \xrightarrow{p} 0$ and $\partial^4 l(0)/\partial \lambda^4 \sim tr(W^4) \sim (n/h_n)$. Hence

$$\sqrt{\frac{n}{h_n}} \tilde{\lambda} = \frac{Z^1(0)}{K(0)} + \sqrt{\frac{h_n}{n}} \frac{Z^{(2)}(0) Z^{(1)}(0)}{K(0)^2} + \frac{1}{2} \sqrt{\frac{h_n}{n}} \frac{J(0)(Z^{(1)}(0))^2}{K(0)^3} + o_p\left(\sqrt{\frac{h_n}{n}}\right),$$

where the last displayed expression has been obtained by substituting into (B.10) the approximation for $\tilde{\lambda}$ implicit in (B.10), i.e.

$$\tilde{\lambda} \sim \sqrt{\frac{h_n}{n}} \frac{Z^{(1)}(0)}{K(0)}.$$

Let x be any finite real number. We have

$$\begin{aligned}
& Pr(\sqrt{\frac{n}{h_n}} \tilde{\lambda} \leq x) \\
= \; & Pr(\frac{Z^1(0)}{K(0)} + \sqrt{\frac{h_n}{n}} \frac{Z^{(2)}(0) Z^{(1)}(0)}{K(0)^2} + \frac{1}{2} \sqrt{\frac{h_n}{n}} \frac{J(0)(Z^{(1)}(0))^2}{K(0)^3} + o_p\left(\sqrt{\frac{h_n}{n}}\right) \leq x) \\
= \; & Pr(\frac{1}{K(0)} \sqrt{\frac{h_n}{n}} \frac{\epsilon' W \epsilon}{\frac{1}{n}\epsilon'\epsilon} + \sqrt{\frac{h_n}{n}} \frac{Z^{(2)}(0) Z^{(1)}(0)}{K(0)^2} + \frac{1}{2} \sqrt{\frac{h_n}{n}} \frac{J(0)(Z^{(1)}(0))^2}{K(0)^3} + o_p\left(\sqrt{\frac{h_n}{n}}\right) \leq x),
\end{aligned}$$

where the last equality is obtained by substituting (B.3) and multiplying both numerator and denominator of the first term by $1/n$. We write

$$\begin{aligned}
\tilde{f} &= \sqrt{\frac{h_n}{n}} \epsilon' W \epsilon - x \frac{K(0)}{n} \epsilon'\epsilon + \sqrt{\frac{h_n}{n}} \frac{Z^{(2)}(0) Z^{(1)}(0)}{K(0)} \frac{1}{n} \epsilon'\epsilon \\
&+ \frac{1}{2} \sqrt{\frac{h_n}{n}} \frac{J(0)(Z^{(1)}(0))^2}{K(0)^2} \frac{1}{n} \epsilon'\epsilon + o_p\left(\sqrt{\frac{h_n}{n}}\right) \\
&= \frac{1}{2} \epsilon'(\tilde{C} + \tilde{C}')\epsilon + \sqrt{\frac{h_n}{n}} \frac{Z^{(2)}(0) Z^{(1)}(0)}{K(0)} \frac{1}{n} \epsilon'\epsilon \\
&+ \frac{1}{2} \sqrt{\frac{h_n}{n}} \frac{J(0)(Z^{(1)}(0))^2}{K(0)^2} \frac{1}{n} \epsilon'\epsilon + o_p\left(\sqrt{\frac{h_n}{n}}\right),
\end{aligned} \tag{B.11}$$

where

$$\tilde{C} = \sqrt{\frac{h_n}{n}} W - x\frac{K(0)}{n} I. \tag{B.12}$$

Therefore, by standard algebraical manipulation,

$$Pr(\sqrt{\frac{n}{h_n}}\tilde{\lambda} \leq x) = Pr(\tilde{f} \leq 0).$$

Under Assumption 5 and by a slight modification of the argument in Lemma 2 the first term of the RHS of (B.11) is $O_e(1)$. The second and the third terms are both $O_p(\sqrt{h_n/n})$ by Lemma 3 and Lemma 4, respectively, and since $K(0)$ is finite and positive in the limit.

Under Assumption 2 the characteristic function of $\tilde{f}$ can be written as

$$\begin{aligned} E(e^{it\tilde{f}}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\Re^n} e^{it\tilde{f}} e^{-\frac{\epsilon'\epsilon}{2\sigma^2}} d\epsilon \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\Re^n} e^{\frac{1}{2}it\epsilon'(\tilde{C}+\tilde{C}')\epsilon} \{1 + it\sqrt{\frac{h_n}{n}} \frac{Z^{(2)}(0)Z^{(1)}(0)}{K(0)} \frac{1}{n}\epsilon'\epsilon \\ &+ \frac{1}{2}it\sqrt{\frac{h_n}{n}} \frac{J(0)(Z^{(1)}(0))^2}{K(0)^2} \frac{1}{n}\epsilon'\epsilon + o_p\left(\sqrt{\frac{h_n}{n}}\right)\} \times e^{-\frac{\epsilon'\epsilon}{2\sigma^2}} d\epsilon. \end{aligned}$$

We denote $\tilde{\eta}_j, j = 1....n$, the eigenvalues of $(\tilde{C} + \tilde{C}')$. Next,

$$\begin{aligned} E(e^{it\tilde{f}}) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{\Re^n} e^{-\frac{1}{2\sigma^2}\epsilon'(I - it\sigma^2(\tilde{C}+\tilde{C}'))\epsilon} d\epsilon \\ &+ it\sqrt{\frac{h_n}{n}} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{K(0)} \int_{\Re^n} e^{-\frac{1}{2\sigma^2}\epsilon'(I - it\sigma^2(\tilde{C}+\tilde{C}'))\epsilon} \frac{Z^{(1)}(0)Z^{(2)}(0)\epsilon'\epsilon}{n} d\epsilon \\ &+ \frac{1}{2}it\sqrt{\frac{h_n}{n}} \frac{1}{\sqrt{2\pi\sigma^2}} \frac{1}{K(0)^2} \int_{\Re^n} e^{-\frac{1}{2\sigma^2}\epsilon'(I - it\sigma^2(\tilde{C}+\tilde{C}'))\epsilon} \frac{(Z^{(1)}(0))^2 J(0)\epsilon'\epsilon}{n} d\epsilon + o\left(\sqrt{\frac{h_n}{n}}\right) \\ &= \prod_{j=1}^{n}(1 - it\sigma^2\tilde{\eta}_j)^{-1/2}(1 + it\sqrt{\frac{h_n}{n}} \frac{1}{K(0)} E(\frac{Z^{(1)}(0)Z^{(2)}(0)\epsilon'\epsilon}{n}) \\ &+ \frac{1}{2}it\sqrt{\frac{h_n}{n}} \frac{1}{K(0)^2} E(\frac{(Z^{(1)}(0))^2 J(0)\epsilon'\epsilon}{n})) + o\left(\sqrt{\frac{h_n}{n}}\right), \end{aligned} \tag{B.13}$$

where the expectations in the last displayed expression are intended with respect to $\epsilon \sim N(0, \Sigma)$, with

$$\Sigma = \frac{\sigma^2}{(I - it\sigma^2(\tilde{C} + \tilde{C}'))} = \sigma^2(\sum_{s=0}^{\infty}(it\sigma^2(\tilde{C} + \tilde{C}'))^s). \tag{B.14}$$

A brief digression on the existence of the moments appearing in (B.13) is necessary. The algebraical details of each term are in Appendix C.2, but for convenience the general form of the expectations of interest are reported here, after substituting the expressions for $Z^{(1)}(0)$, $Z^{(2)}(0)$ and $J(0)$:

$$E\left(\frac{(\epsilon' W\epsilon)^p}{(\epsilon'\epsilon)^{p-1}}\right), p = 2, 3, 5 \tag{B.15}$$

and

$$E\left(\frac{(\epsilon' W\epsilon)^p \epsilon' W' W\epsilon}{(\epsilon'\epsilon)^{p}}\right), p = 1, p = 3. \tag{B.16}$$

Regarding existence of (B.15), we have

$$E\left(\frac{(\epsilon' W\epsilon)^p}{(\epsilon'\epsilon)^{p-1}}\right) = E\left(\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^p \epsilon'\epsilon\right) = E\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^p E(\epsilon'\epsilon) = \frac{E(\epsilon' W\epsilon)^p}{E(\epsilon'\epsilon)^p} E(\epsilon'\epsilon),$$

since $\epsilon' W\epsilon/\epsilon'\epsilon$ is independent of $\epsilon'\epsilon$. When $\epsilon \sim N(0, \sigma^2)$ the three expectations in the latter expression can be easily computed by standard statistical formulae. However, the expectations appearing in (B.13) are intended with respect to $\epsilon \sim N(0, \Sigma)$, with Σ given in (B.14). Since the derivation is intended to be formal, validity issues of the series representation given in (B.14) are not taken into consideration and, at least for establishing existence, we rely on the approximation $\Sigma \sim \sigma^2$. Regarding existence of (B.16), by the CauchySchwarz inequality we have

$$\left|E\left(\frac{(\epsilon' W\epsilon)^p \epsilon' W' W\epsilon}{(\epsilon'\epsilon)^{p}}\right)\right| \leq \left\{E\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^{2p} E(\epsilon' W' W\epsilon)^2\right\}^{1/2}.$$

By the same argument presented above, the first factor on the RHS is finite and by Gaussianity the second is finite. The latter argument holds when W is symmetric, but it can be immediately extended to non symmetric W by replacing W with $(W+W')/2$.

For notational simplicity, we write

$$Q = Q_1 + Q_2 + o\left(\sqrt{\frac{h_n}{n}}\right),$$

where

$$Q_1 = it\sqrt{\frac{h_n}{n}} \frac{1}{K(0)} E\left(\frac{Z^{(1)}(0) Z^{(2)}(0)\epsilon'\epsilon}{n}\right)$$

and

$$Q_2 = \frac{1}{2} it\sqrt{\frac{h_n}{n}} \frac{1}{K(0)^2} E\left(\frac{(Z^{(1)}(0))^2 J(0)\epsilon'\epsilon}{n}\right).$$

From (B.13) the cumulant generating function for $\tilde{f}$, denoted $\tilde{\psi}(t)$, can be written as

$$\begin{aligned} \tilde{\psi}(t) &= -\frac{1}{2}\sum_{j=1}^{n} ln(1 - it\sigma^2\tilde{\eta}_j) + ln(1+Q) \\ &= \frac{1}{2}\sum_{s=1}^{\infty}\frac{(it\sigma^2)^s}{s} tr((\tilde{C}+\tilde{C}')^s) + \sum_{s=1}^{\infty}\frac{(-1)^{s+1}}{s} Q^s. \end{aligned} \tag{B.17}$$

Let $\tilde{\kappa}_s$ be the sth cumulant of f. Since we are interested in the second order Edgeworth correction, we need to evaluate $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ up to order $O(\sqrt{h_n/n})$. The contributions of the first term of the RHS of (B.17) to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ are given by

$$\sigma^2 tr(\tilde{C}) = -\sigma^2 xK(0), \tag{B.18}$$

$$\frac{\sigma^4}{2} tr((\tilde{C}+\tilde{C}')^2)) = \frac{h_n}{n}\sigma^4(tr(W^2)+tr(W'W)) + O\left(\frac{1}{n}\right) \tag{B.19}$$

and

$$\sigma^6 tr(\tilde{C}+\tilde{C}')^3 = \sigma^6\left(\frac{h_n}{n}\right)^{3/2}\left(2tr(W^3)+6tr(W^2W')\right) + o\left(\sqrt{\frac{h_n}{n}}\right), \tag{B.20}$$

respectively. The contribution to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ of the second term of the RHS of (B.17) are evaluated in Appendix C.2. Collecting (B.18), (B.19), (B.20) and the results in Appendix C.2 we have

$$\begin{aligned}\tilde{\kappa}_1 &= -\sigma^2 xK(0) - 2\sigma^2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{K(0)} \\ &- \sigma^2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{K(0)} + o\left(\sqrt{\frac{h_n}{n}}\right),\end{aligned} \tag{B.21}$$

$$\tilde{\kappa}_2 = \sigma^4\frac{h_n}{n}(tr(W^2)+tr(W'W)) + o\left(\sqrt{\frac{h_n}{n}}\right) \tag{B.22}$$

and

$$\tilde{\kappa}_3 = -4\sigma^6\left(\frac{h_n}{n}\right)^{3/2} tr(W^3) - 6\sigma^6\left(\frac{h_n}{n}\right)^{3/2} tr(WW'W) + o\left(\sqrt{\frac{h_n}{n}}\right). \tag{B.23}$$

We centres and scale the statistic $\tilde{f}$, as

$$\tilde{f}^c = \frac{\tilde{f}-\tilde{\kappa}_1}{\tilde{\kappa}_2^{1/2}}.$$

The cumulant generating function of $\tilde{f}^c$ can be written as:

$$\tilde{\psi}^c(t) = -\frac{1}{2}t^2 + \sum_{s=3}^{\infty}\frac{\tilde{\kappa}_s^c(it)^s}{s!}, \tag{B.24}$$

where $\tilde{\kappa}_s^c = \tilde{\kappa}_s/\tilde{\kappa}_2^{s/2}$. From (B.24), the characteristic function of $\tilde{f}^c$ is

$$\begin{aligned}E(e^{it\tilde{f}^c}) &= e^{-\frac{1}{2}t^2}\exp\{\sum_{s=3}^{\infty}\frac{\tilde{\kappa}_s^c(it)^s}{s!}\} = \\ &= e^{-\frac{1}{2}t^2}\{1+\sum_{s=3}^{\infty}\frac{\tilde{\kappa}_s^c(it)^s}{s!} + \frac{1}{2!}(\sum_{s=3}^{\infty}\frac{\tilde{\kappa}_s^c(it)^s}{s!})^2 + \frac{1}{3!}(\sum_{s=3}^{\infty}\frac{\tilde{\kappa}_s^c(it)^s}{s!})^3 + ...\} \\ &= e^{-\frac{1}{2}t^2}\{1+\frac{\tilde{\kappa}_3^c(it)^3}{3!} + \frac{\tilde{\kappa}_4^c(it)^4}{4!} + \frac{\tilde{\kappa}_5^c(it)^5}{5!} + \{\frac{\tilde{\kappa}_6^c}{6!} + \frac{(\tilde{\kappa}_3^c)^2}{(3!)^2}\}(it)^6 + ...\}.\end{aligned}$$

Thus, by the Fourier inversion formula, we can conclude that

$$Pr(\tilde{f}^c \leq z) = \int_{-\infty}^{z} \phi(z)dz + \frac{\tilde{\kappa}_3^c}{3!}\int_{-\infty}^{z} H_3(z)\phi(z)dz + \frac{\tilde{\kappa}_4^c}{4!}\int_{-\infty}^{z} H_4(z)\phi(z)dz + \ldots.$$

Collecting the results derived above, we have

$$\begin{aligned} Pr(\sqrt{\frac{n}{h_n}}\tilde{\lambda} \leq x) &= Pr(\tilde{f} \leq 0) = Pr(\tilde{f}^c\tilde{\kappa}_2^{1/2} + \tilde{\kappa}_1 \leq 0) = Pr(\tilde{f}^c \leq -\tilde{\kappa}_1^c) \\ &= \Phi(-\tilde{\kappa}_1^c) - \frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(-\tilde{\kappa}_1^c) + \frac{\tilde{\kappa}_4^c}{4!}\Phi^{(4)}(-\tilde{\kappa}_1^c) + \ldots \end{aligned} \quad \text{(B.25)}$$

Now, from (B.21) and (B.22)

$$\begin{aligned} \tilde{\kappa}_1^c &= \frac{-\sigma^2 xK(0) - 2\sigma^2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{K(0)} - \sigma^2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{K(0)}}{\sigma^2(\frac{h_n}{n}(tr(W^2)+tr(W'W)))^{1/2}} + o\left(\sqrt{\frac{h_n}{n}}\right) \\ &= \frac{-x\frac{h_n}{n}(tr(W^2)+tr(W'W)) - 2\sqrt{\frac{h_n}{n}}\frac{tr(WW'W)}{(tr(W^2)+tr(W'W))} - \sqrt{\frac{h_n}{n}}\frac{tr(W^3)}{(tr(W^2)+tr(W'W))}}{(\frac{h_n}{n}(tr(W^2)+tr(W'W)))^{1/2}} \\ &+ o\left(\sqrt{\frac{h_n}{n}}\right), \end{aligned}$$

where the second equality has been obtained by substituting

$$K(0) = \frac{h_n}{n}(tr(W^2)+tr(W'W)) + O\left(\frac{1}{n}\right),$$

according to (B.9). We choose $x = \tilde{a}^{-1}\zeta$, where

$$\tilde{a} = \sqrt{\frac{h_n}{n}}\sqrt{tr(W^2)+tr(W'W)}.$$

Therefore

$$\tilde{\kappa}_1^c = -\zeta - 2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{\tilde{a}^3} - \left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{\tilde{a}^3} + o\left(\sqrt{\frac{h_n}{n}}\right)$$

and, from (B.22) and (B.23),

$$\begin{aligned} \tilde{\kappa}_3^c &\sim \frac{-4\left(\frac{h_n}{n}\right)^{3/2}tr(W^3) - 6\left(\frac{h_n}{n}\right)^{3/2}tr(WW'W)}{\left(\frac{h_n}{n}\right)^{3/2}(tr(W^2)+tr(W'W))^{3/2}} \\ &= -\left(\frac{h_n}{n}\right)^{3/2}\frac{4tr(W^3)+6tr(WW'W)}{\tilde{a}^3} \sim \sqrt{\frac{h_n}{n}}. \end{aligned}$$

By Taylor expansion of the function $\Phi(-\tilde{\kappa}_1^c)$ in (B.25), we conclude

$$
\begin{aligned}
Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} \leq \zeta) &= \Phi(\zeta) + \left(2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{\tilde{a}^3} + \left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{\tilde{a}^3}\right)\phi(\zeta) \\
&- \frac{\tilde{\kappa}_3^c}{3!}\Phi^{(3)}(\zeta) + o\left(\sqrt{\frac{h_n}{n}}\right).
\end{aligned}
$$

C Appendix C

C.1 Some useful Lemmas

In this Appendix we will present and prove some of the auxiliary results used in the proofs of Theorem 1 and Theorem 2. As already stressed, the expansions in Theorem 1 and Theorem 2 are formal, so we do not deal with convergence issues in some of the results that follow. Moreover, it must be mentioned that for notational simplicity, we prove Lemmas 2, 3 and 4 for a symmetric W. When W is not symmetric the same results hold simply by substituting $(W+W')/2$ instead of W where necessary.

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

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

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

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

uniformly in i. Therefore:

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

Lemma 2 Under Assumptions 2-5

$$Z^{(1)}(0) = \sqrt{h_n}n\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} = O_e(1).$$

Proof We have

$$E\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2 = \frac{E(\epsilon' W\epsilon)^2}{E(\epsilon'\epsilon)^2} = \frac{2tr(W^2)}{n^2+2n} \sim \frac{1}{nh_n},$$

under Assumptions 2-5. Hence, by Markov's inequality

$$\sqrt{h_n}n\frac{\epsilon' W\epsilon}{\epsilon'\epsilon} = O_e(1).$$

Lemma 3 Under Assumptions 2-5

$$Z^{(2)}(0) = O_p(1),$$

where $Z^{(2)}(.)$ is defined according to (B.4).

Proof By rearranging terms in the first two lines of (B.4) we have

$$\begin{aligned} Z^{(2)}(0) &= -\sqrt{\frac{h_n}{n}}\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon} - nE\left(\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right)\right) \\ &+ \sqrt{\frac{h_n}{n}}\left(2n\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2 - 2nE\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2\right). \end{aligned}$$

By the C_r moment inequality

$$\begin{aligned} E(Z^{(2)}(0))^2 &\leq 2\frac{h_n}{n}E\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon} - nE\left(\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right)\right)^2 \\ &+ 2\frac{h_n}{n}E\left(2n\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2 - 2nE\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2\right)^2. \end{aligned} \tag{C.1}$$

Now,

$$\begin{aligned} & E\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon} - nE\left(\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right)\right)^2 = E\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon} - tr(W'W)\right)^2 \\ = \; & n^2\frac{E(\epsilon' W' W\epsilon)^2}{E(\epsilon'\epsilon)^2} + (tr(W'W))^2 - 2ntr(W'W)\frac{E(\epsilon' W' W\epsilon)}{E(\epsilon'\epsilon)} \\ = \; & ((tr(W'W))^2 + 2tr((W'W)^2))\left(1+\frac{2}{n}\right)^{-1} + (tr(W'W))^2 - 2(tr(W'W))^2 \\ = \; & ((tr(W'W))^2 + 2tr((W'W)^2))\left(1-\frac{2}{n}+O\left(\frac{1}{n^2}\right)\right) - (tr(W'W))^2 \\ = \; & 2tr((W'W)^2)\left(1-\frac{2}{n}+O\left(\frac{1}{n^2}\right)\right) - (tr(W'W))^2\left(\frac{2}{n}+O\left(\frac{1}{n^2}\right)\right) \end{aligned} \tag{C.2}$$

and hence

$$E\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon} - nE\left(\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right)\right)^2 \sim 2tr((W'W)^2) \sim \frac{n}{h_n}, \tag{C.3}$$

under Assumption 5. In case the sequence h_n is bounded, the latter result would be modified as

$$E\left(n\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon} - nE\left(\frac{\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right)\right)^2 \sim 2tr((W'W)^2) - \frac{2}{n}(tr(W'W))^2 \sim n.$$

It is worth stressing that, despite we are not attempting to provide an exact rate, we could not use the inequality

$$E\left(X - E(X)\right)^2 \leq E(X^2)$$

instead of (C.2), as it would neglect relevant terms. Moreover,

$$\begin{aligned} & 4n^2 E\left(\left(\frac{\epsilon' W \epsilon}{\epsilon'\epsilon}\right)^2 - E\left(\frac{\epsilon' W \epsilon}{\epsilon'\epsilon}\right)^2\right)^2 \leq 4n^2 E\left(\frac{\epsilon' W \epsilon}{\epsilon'\epsilon}\right)^4 \\ = \quad & 4n^2 \frac{E(\epsilon' W \epsilon)^4}{E(\epsilon'\epsilon)^4} \sim 4n^2 \frac{12(tr(W^2))^2 + 48tr(W^4)}{n^4} \sim \frac{1}{h_n^2}. \qquad \text{(C.4)} \end{aligned}$$

Collecting (C.1), (C.3), (C.4) and by Markov's inequality, we conclude $Z^{(2)}(0) = O_p(1)$.

Lemma 4 Under Assumptions 1-5

$$J(0) = O_p(1),$$

where $J(0)$ is defined according to (B.8).

Proof By the C_r moment inequality (applied twice), we have

$$\begin{aligned} E(J(0))^2 \quad & \leq \quad 2\frac{h_n^2}{n^2}\left(E\left(\frac{6n\epsilon' W \epsilon\epsilon' W' W \epsilon}{(\epsilon'\epsilon)^2}\right)^2 + E\left(\frac{8n(\epsilon' W \epsilon)^3}{(\epsilon'\epsilon)^3} - 2tr(W^3)\right)^2\right) \\ & \leq \quad 2\frac{h_n^2}{n^2} E\left(\frac{6n\epsilon' W \epsilon\epsilon' W' W \epsilon}{(\epsilon'\epsilon)^2}\right)^2 + 4\frac{h_n^2}{n^2} E\left(\frac{8n(\epsilon' W \epsilon)^3}{(\epsilon'\epsilon)^3}\right)^2 \\ & + \quad 4\frac{h_n^2}{n^2}(2tr(W^3))^2. \qquad \text{(C.5)} \end{aligned}$$

In order to evaluate the rate of the first term in (C.5), we use the approximation $E(A/B) \sim E(A)/E(B)$, without deriving the exact order of the remainder. Some comments about the existence of the expectations in (C.5) are provided in Appendix B. Using also standard results on the expectations of quadratic forms, we have

$$\begin{aligned} & E\left(\frac{6n\epsilon' W \epsilon\epsilon' W' W \epsilon}{(\epsilon'\epsilon)^2}\right)^2 \sim 36n^2 \frac{E(\epsilon' W \epsilon\epsilon' W' W \epsilon)^2}{E(\epsilon'\epsilon)^4} \\ \sim \quad & 36n^2 \frac{2tr(W^2)(tr(W'W))^2}{n^4} \sim \frac{n}{h_n^3}. \qquad \text{(C.6)} \end{aligned}$$

Moreover, by a recursive formula (Ghazal (1996))

$$E(\epsilon' W \epsilon)^n = \sum_{i=0}^{n-1} g_i E(\epsilon' W \epsilon)^{n-1-i}, \qquad \text{(C.7)}$$

where

$$g_i = \begin{pmatrix} n-1 \\ i \end{pmatrix} i! 2^i \sigma^{2i+2} tr((W)^{i+1}),$$

we have

$$E\left(\frac{8n(\epsilon'W\epsilon)^3}{(\epsilon'\epsilon)^3}\right)^2 = \frac{64n^2E(\epsilon'W\epsilon)^6}{E(\epsilon'\epsilon)^6} \sim 64n^2\frac{120(tr(W^2))^3}{n^6} \sim \frac{1}{nh_n^3}. \tag{C.8}$$

Hence, the term

$$4(tr(W^3)^2) \sim \frac{n^2}{h_n^2}$$

in (C.5) dominates both (C.6) and (C.8), whether h_n is divergent or bounded. Therefore

$$E(J(0))^2 = O(\frac{h_n^2}{n^2}\frac{n^2}{h_n^2}) = O(1),$$

implying $J(0) = O_p(1)$.

Lemma 5 Under Assumption 2 and for W given by (4.1), $\underset{n\to\infty}{plim}(\hat{\lambda}-\lambda)$ is finite and has the same sign of λ.

Proof The OLS estimate of λ is

$$\hat{\lambda}-\lambda = \frac{\frac{h_n}{n}Y'W\epsilon}{\frac{h_n}{n}Y'W^2Y} = \frac{\frac{h_n}{n}\epsilon'S^{-1}(\lambda)W\epsilon}{\frac{h_n}{n}\epsilon'S^{-1}(\lambda)W^2S^{-1}(\lambda)\epsilon}, \tag{C.9}$$

since W in (4.1) is symmetric and $Y = S^{-1}(\lambda)\epsilon$.

Regarding the limit of the numerator of the RHS of (C.9), we have

$$\begin{aligned}
&\left(\frac{h_n}{n}\right)^2 E(\epsilon'S^{-1}(\lambda)W\epsilon - \sigma^2 tr(S^{-1}(\lambda)W))^2 \\
= &\left(\frac{h_n}{n}\right)^2 E(\epsilon'S^{-1}(\lambda)W\epsilon)^2 - \sigma^4(tr(S^{-1}(\lambda)W))^2 \\
= &\ 2\sigma^4\left(\frac{h_n}{n}\right)^2 tr((S^{-1}(\lambda)W)^2) \to 0
\end{aligned}$$

as $n\to\infty$, since $tr((S^{-1}(\lambda)W)^2) = O(n/h_n)$ by Lemma 1. Hence

$$\frac{h_n}{n}(\epsilon'S^{-1}(\lambda)W\epsilon - \sigma^2 tr(S^{-1}(\lambda)W)) \to 0$$

in second mean, implying

$$\underset{n\to\infty}{plim}\frac{h_n}{n}\epsilon'S^{-1}(\lambda)W\epsilon = \lim_{n\to\infty}\sigma^2\frac{h_n}{n}tr(S^{-1}(\lambda)W). \tag{C.10}$$

Similarly,

$$\underset{n\to\infty}{plim}\frac{h_n}{n}\epsilon'S^{-1}(\lambda)W^2S^{-1}(\lambda)\epsilon = \lim_{n\to\infty}\sigma^2\frac{h_n}{n}tr((S^{-1}(\lambda)W)^2). \tag{C.11}$$

From (C.10) and (C.11), we have

$$\hat{\lambda}-\lambda \xrightarrow{p} \lim_{n\to\infty}\frac{\frac{h_n}{n}tr(S^{-1}(\lambda)W)}{\frac{h_n}{n}tr((S^{-1}(\lambda)W)^2)}. \tag{C.12}$$

First, we show that the RHS of (C.12) is finite. Lemma 1 implies

$$\frac{h_n}{n} tr(S^{-1}(\lambda)W) = O(1).$$

The denominator in the RHS of (C.12) is non-negative and by (3.3) we have

$$\frac{h_n}{n} tr((S^{-1}(\lambda)W)^2) \sim \frac{h_n}{n} tr(W^2),$$

which is non-zero for W given in (4.1), as shown in section 4. Hence, the RHS of (C.12) is finite and its sign depends on its numerator.

From (4.1) and the series representation in (3.3), we have

$$tr(S^{-1}(\lambda)W) \quad = \quad tr(\sum_{i=0}^{\infty} \lambda^i tr(W^{i+1})) = r\sum_{i=0}^{\infty} \lambda^i tr(B_m^{i+1}).$$

Since B_m has one eigenvalue equal to 1 and the other $(m-1)$ equal to $-1/(m-1)$, we have

$$tr(B_m^{i+1}) = 1 + (m-1)\left(\frac{-1}{m-1}\right)^{i+1}$$

and hence, since $|\lambda| < 1$,

$$\begin{aligned} tr(S^{-1}(\lambda)W) &= r\sum_{i=0}^{\infty} \lambda^i \left(1 - \left(\frac{-1}{m-1}\right)^i\right) \\ &= \frac{r}{1-\lambda} - \frac{r}{1+\frac{\lambda}{m-1}} = \frac{\lambda}{1-\lambda}\frac{rm}{m-1+\lambda}. \end{aligned} \tag{C.13}$$

By substituting $h_n = m - 1$ and $n = mr$ into (C.13), we obtain

$$\frac{h_n}{n} tr(S^{-1}(\lambda)W) = \frac{m-1}{mr}\frac{\lambda}{1-\lambda}\frac{rm}{m-1+\lambda} = \frac{\lambda}{1-\lambda}\frac{m-1}{m-1+\lambda},$$

which has the same sign of λ, whether m is divergent or bounded, provided that $m > 1$.

C.2 Evaluation of cumulants

In this section we evaluate the contribution to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ of the term

$$Q_1 = it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}\frac{1}{n}E(Z^{(1)}(0)Z^{(2)}(0)\epsilon'\epsilon)$$

appearing in (B.13). Since the expansion in Theorem 2 is formal, the approximation $E(A/B) \sim E(A)/E(B)$ is used without proving the exact order of the remainder terms. Substituting (B.3) and (B.4), we write

$$Q_1 = Q_{11} + Q_{12} + Q_{13},$$

where:

$$Q_{11} = -it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}h_n E\left(\frac{\epsilon' W\epsilon\epsilon' W' W\epsilon}{\epsilon'\epsilon}\right),$$

$$Q_{12} = 2it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}h_n E\left(\frac{(\epsilon' W\epsilon)^3}{(\epsilon'\epsilon)^2}\right)$$

and

$$Q_{13} = it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}\frac{h_n}{n}\left(tr(W'W) - \frac{1}{n}tr((W+W')^2)\left(1+\frac{2}{n}\right)^{-1}\right)E(\epsilon' W\epsilon).$$

Contribution from term Q_{11}

Using also some standard results on the expectations of quadratic forms in normal random variables, we have

$$\begin{aligned} Q_{11} &\sim -it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}h_n\frac{E\left(\frac{1}{2}\epsilon'(W+W')\epsilon\epsilon' W'W\epsilon\right)}{E(\epsilon'\epsilon)} \\ &= -it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}h_n\frac{\frac{1}{2}tr((W+W')\Sigma)tr(W'W\Sigma) + tr(\Sigma(W+W')\Sigma W'W)}{tr(\Sigma)}, \end{aligned}$$

where Σ is defined according to (B.14). It is straightforward to show that $tr(\Sigma) \sim \sigma^2 n$.

The contribution from Q_{11} to $\tilde{\kappa}_1$ is then

$$\begin{aligned} & -2\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}\frac{h_n}{n}\sigma^2 tr(WW'W) + o\left(\sqrt{\frac{h_n}{n}}\right) \\ =\; & -2\sigma^2\sqrt{\frac{h_n}{n}}\frac{tr(WW'W)}{tr(W^2)+tr(W'W)} + o\left(\sqrt{\frac{h_n}{n}}\right), \end{aligned} \tag{C.14}$$

since

$$K(0) = \frac{h_n}{n}(trW^2 + tr(W'W)) + O(\frac{1}{n}),$$

according to (B.9).

The contribution to $\tilde{\kappa}_2$ comes from the term

$$\begin{aligned} & -(it)^2\sigma^4(\frac{h_n}{n})^{3/2}\frac{1}{K(0)}[\frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}'))tr(W'W) \\ + \;& tr((\tilde{C}+\tilde{C}')(W+W')W'W) + tr((W+W')(\tilde{C}+\tilde{C}')W'W)], \end{aligned}$$

with $\tilde{C}$ given by (B.12). The contribution to $\tilde{\kappa}_2$ is given by

$$-\sigma^4(\frac{h_n}{n})^2\frac{1}{K(0)}(tr((W+W')^2)tr(W'W) + o_p\left(\sqrt{\frac{h_n}{n}}\right), \tag{C.15}$$

since

$$tr((W+W')(\tilde{C}+\tilde{C}')) = (tr(W+W')^2)\sqrt{\frac{h_n}{n}} \sim \left(\sqrt{\frac{n}{h_n}}\right), \tag{C.16}$$

$$tr((\tilde{C}+\tilde{C}')(W+W')W'W) \sim \left(\sqrt{\frac{n}{h_n}}\right) \tag{C.17}$$

and

$$tr((W+W')(\tilde{C}+\tilde{C}')W'W) \sim \left(\sqrt{\frac{n}{h_n}}\right). \tag{C.18}$$

Similarly, the contribution to $\tilde{\kappa}_3$ comes from the term

$$
\begin{aligned}
& -(it)^3(\frac{h_n}{n})^{3/2}\frac{1}{K(0)}\frac{1}{\sigma^2}\sigma^8(\frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}'))tr(W'W(\tilde{C}+\tilde{C}')) \\
+ & \frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}')^2)tr(W'W)+tr((\tilde{C}+\tilde{C}')^2(W+W')W'W) \\
+ & tr((W+W')(\tilde{C}+\tilde{C}')^2W'W)+tr((\tilde{C}+\tilde{C}')(W+W')(\tilde{C}+\tilde{C}')W'W)).
\end{aligned}
$$

Now,

$$tr(W'W(\tilde{C}+\tilde{C}')) \sim \sqrt{\frac{h_n}{n}}2tr(WW'W), \tag{C.19}$$

$$tr(W(\tilde{C}+\tilde{C}')^2) \sim \frac{h_n}{n}tr((W+W')^3), \tag{C.20}$$

$$tr((\tilde{C}+\tilde{C}')^2(W+W')W'W) = o\left(\sqrt{\frac{n}{h_n}}\right), \tag{C.21}$$

$$tr((W+W')(\tilde{C}+\tilde{C}')^2W'W) = o\left(\sqrt{\frac{n}{h_n}}\right) \tag{C.22}$$

and

$$tr((\tilde{C}+\tilde{C}')(W+W')(\tilde{C}+\tilde{C}')W'W) = o\left(\sqrt{\frac{n}{h_n}}\right). \tag{C.23}$$

Using (C.16), (C.17), (C.19)-(C.23), and after some tedious but straightforward algebra we conclude that the contribution to $\tilde{\kappa}_3$ is

$$
\begin{aligned}
& -6\left(\frac{h_n}{n}\right)^{5/2}\frac{1}{K(0)}\sigma^6(2tr(W^2)tr(WW'W)+5tr(W'W)tr(WW'W) \\
+ & tr(W'W)tr(W^3))+o\left(\sqrt{\frac{h_n}{n}}\right).
\end{aligned} \tag{C.24}
$$

When W is symmetric (e.g. W given in (4.1)), the latter expression simplifies to

$$-24\left(\frac{h_n}{n}\right)^{3/2}\sigma^6tr(W^3)+o\left(\sqrt{\frac{h_n}{n}}\right),$$

as

$$K(0) = 2\frac{h_n}{n}trW^2+O\left(\frac{1}{n}\right),$$

according to (B.9).

Contribution from term Q_{12}

By independence between the ratio $\epsilon' W \epsilon / \epsilon' \epsilon$ and $\epsilon' \epsilon$, we have

$$\begin{aligned}
Q_{12} &= 2it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}h_n E\left(\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^3 \epsilon'\epsilon\right) = 2it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}h_n E\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^3 E(\epsilon'\epsilon) \\
&= 2it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}h_n \frac{E\left(\frac{1}{2}\epsilon'(W+W')\epsilon\right)^3}{E(\epsilon'\epsilon)^3} E(\epsilon'\epsilon) \\
&= 2it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}h_n \{(\frac{1}{2}tr((W+W')\Sigma))^3 + 6tr\left(\frac{1}{2}(W+W')\Sigma\right) tr\left(\left(\frac{1}{2}(W+W')\Sigma\right)^2\right) \\
&+ 8tr\left(\left(\frac{1}{2}\Sigma(W+W')\right)^3\right)\} \frac{tr(\Sigma)}{(tr\Sigma)^3 + 6tr(\Sigma)tr(\Sigma^2) + 8tr(\Sigma^3)},
\end{aligned}$$

where Σ is defined according to (B.14). We have

$$(tr\Sigma)^3 + 6tr(\Sigma)tr(\Sigma^2) + 8tr(\Sigma^3) \sim \sigma^6 n^3$$

and $tr(\Sigma) \sim \sigma^2 n$.

The contribution from Q_{12} to $\tilde{\kappa}_1$ is then

$$2\sigma^2\sqrt{\frac{h_n}{n}}\frac{1}{K(0)}\frac{h_n}{n^2}tr((W+W')^3) = o\left(\sqrt{\frac{h_n}{n}}\right), \tag{C.25}$$

A similar argument holds also for the contribution from Q_{12} to both $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$.

Contribution from term Q_{13}

We have

$$Q_{13} = it\left(\frac{h_n}{n}\right)^{3/2}\frac{1}{K(0)}(tr(W'W) - \frac{1}{n}tr((W+W')^2)(1+\frac{2}{n})^{-1})tr(\frac{1}{2}(W+W')\Sigma).$$

It is straightforward to see that there are no contributions to $\tilde{\kappa}_1$, since

$$\sigma^2\left(\frac{h_n}{n}\right)^{3/2}\frac{1}{K(0)}(tr(W'W) - \frac{1}{n}tr(W^2)(1+\frac{2}{n})^{-1})tr(W) = 0. \tag{C.26}$$

The contribution to $\tilde{\kappa}_2$ comes from

$$(it)^2\sigma^4\left(\frac{h_n}{n}\right)^{3/2}\frac{1}{K(0)}(tr(W'W) - \frac{4}{n}tr(W^2)(1+\frac{2}{n})^{-1})\frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}'))$$

and by (C.16) we conclude that Q_{13} contributes to κ_2 with the term

$$\sigma^4\left(\frac{h_n}{n}\right)^2\frac{1}{K(0)}tr((W+W')^2)tr(W'W) + o_p\left(\sqrt{\frac{h_n}{n}}\right). \tag{C.27}$$

The contribution to $\tilde{\kappa}_3$ comes from

$$(it)^3\sigma^6\left(\frac{h_n}{n}\right)^{3/2}\frac{1}{K(0)}(tr(W'W)-\frac{1}{n}tr(W^2)(1+\frac{2}{n})^{-1})\frac{1}{2}tr((W+W')(\tilde{C}+\tilde{C}')^2)$$

and hence, from (C.20), we conclude that Q_{13} contributes to κ_3 with the term

$$6\sigma^6\left(\frac{h_n}{n}\right)^{5/2}\frac{1}{K(0)}tr(W'W)(tr(W^3)+3tr(W(W')^2))+o\left(\sqrt{\frac{h_n}{n}}\right). \tag{C.28}$$

When W is symmetric, the latter simplifies to

$$12\sigma^6(\frac{h_n}{n})^{3/2}tr(W^3)+o\left(\sqrt{\frac{h_n}{n}}\right).$$

From (C.14), (C.25) and (C.26) we conclude that Q_1 contributes to $\tilde{\kappa}_1$ with the term

$$-2\sigma^2\left(\frac{h_n}{n}\right)^{3/2}\frac{1}{K(0)}tr(WW'W)+o\left(\sqrt{\frac{h_n}{n}}\right). \tag{C.29}$$

From (C.15) and (C.27) we conclude that any contribution to $\tilde{\kappa}_2$ from Q_1 is neglegible, while collecting (C.24) and (C.28) we have that the contribution to $\tilde{\kappa}_3$ from Q_1 is

$$\begin{aligned} & -12\sigma^6\left(\frac{h_n}{n}\right)^{5/2}\frac{1}{K(0)}tr(WW'W)(tr(W^2)+tr(W'W))+o\left(\sqrt{\frac{h_n}{n}}\right) \\ = \quad & -12\sigma^6\left(\frac{h_n}{n}\right)^{3/2}tr(WW'W)+o\left(\sqrt{\frac{h_n}{n}}\right), \end{aligned} \tag{C.30}$$

where the last equality follows by substituting

$$K(0)=\frac{h_n}{n}(trW^2+tr(W'W))+O(\frac{1}{n}),$$

according to (B.9).

Finally, we report the main steps for the evaluation of the contribution to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ from

$$Q_2=\frac{1}{2}it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)^2}\frac{1}{n}E((Z^{(1)}(0))^2J(0)\epsilon'\epsilon).$$

Substituting (B.3) and (B.8), we write

$$Q_2=Q_{21}+Q_{22}+Q_{23},$$

where

$$Q_{21}\sim -3it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)^2}h_n^2\frac{E((\epsilon'W\epsilon)^3\epsilon'W'W\epsilon)}{E(\epsilon'\epsilon)^3},$$

$$
\begin{aligned}
Q_{22} &= 4it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)^2}h_n^2 E\left(\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^5 \epsilon'\epsilon\right) \\
&= 4it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)^2}h_n^2 \frac{E(\epsilon' W\epsilon)^5}{E(\epsilon'\epsilon)^5}E(\epsilon'\epsilon)
\end{aligned}
$$

and

$$
\begin{aligned}
Q_{23} &= -it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)^2}\frac{h_n^2}{n}tr(W^3)E\left(\left(\frac{\epsilon' W\epsilon}{\epsilon'\epsilon}\right)^2 \epsilon'\epsilon\right) \\
&= -it\sqrt{\frac{h_n}{n}}\frac{1}{K(0)^2}\frac{h_n^2}{n}tr(W^3)\frac{E(\epsilon' W\epsilon)^2}{E(\epsilon'\epsilon)^2}E(\epsilon'\epsilon).
\end{aligned}
$$

Contribution from term Q_{21}

Using some standard results on the expectations of quadratic forms in normal random variables, we have

$$
\begin{aligned}
E((\frac{1}{2}\epsilon'(W+W')\epsilon)^3\epsilon' W' W\epsilon) &= (tr(\Sigma\frac{1}{2}(W+W')))^3 tr(\Sigma W' W) \\
&+ 6(tr(\Sigma\frac{1}{2}(W+W')))^2 tr(\Sigma\frac{1}{2}(W+W')\Sigma W' W) \\
&+ 6tr((\Sigma\frac{1}{2}(W+W'))^2)tr(\Sigma\frac{1}{2}(W+W'))tr(\Sigma W' W) \\
&+ 8tr(\Sigma\frac{1}{2}(W+W'))tr((\Sigma\frac{1}{2}(W+W'))^2\Sigma W' W) \\
&+ 12tr(\Sigma\frac{1}{2}(W+W')\Sigma W' W)tr((\Sigma\frac{1}{2}(W+W'))^2) \\
&+ 48tr((\Sigma\frac{1}{2}(W+W'))^3\Sigma W' W)
\end{aligned}
$$

and $E(\epsilon'\epsilon)^3 \sim n^3\sigma^6$. Therefore, the contribution to $\tilde{\kappa}_1$ is

$$
\begin{aligned}
&-3\sigma^2\sqrt{\frac{h_n}{n}}\frac{1}{(K(0))^2}\frac{h_n^2}{n^3}(\frac{3}{2}tr((W+W')^2)tr((W+W')WW') \\
&+ \quad 6tr((W+W')^3W'W)) = o\left(\sqrt{\frac{h_n}{n}}\right).
\end{aligned}
$$

A similar argument holds for the contribution to both $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$.

Contribution from term Q_{22}

We have

$$
\frac{E(\epsilon'\epsilon)}{E(\epsilon'\epsilon)^5} \sim \frac{\sigma^2 n}{\sigma^{10}n^5} = \frac{1}{\sigma^8 n^4}.
$$

Also, we can evaluate the fifth moment of $\epsilon' W\epsilon$ by the recursive formula given in (C.7). By tedious, but straightforward algebra, it is possible to show that the contribution to $\tilde{\kappa}_1$, $\tilde{\kappa}_2$ and $\tilde{\kappa}_3$ are $o(\sqrt{h_n/n})$. Intuitively, this is because no term in $E((\epsilon' W\epsilon)^5)$ is large enough to offset the factor h_n^2/n^4.

Contribution from term Q_{23}

We have

$$E((\epsilon' W\epsilon)^2) = \frac{1}{4}E(\epsilon'(W+W')\epsilon)^2 = \frac{1}{4}(tr(\Sigma(W+W')))^2 + \frac{1}{2}tr((\Sigma(W+W'))^2)$$

and

$$\frac{E(\epsilon'\epsilon)}{E(\epsilon'\epsilon)^2} \sim \frac{n\sigma^2}{n^2\sigma^4} = \frac{1}{n\sigma^2}.$$

Therefore, the contribution to $\tilde{\kappa}_1$ is

$$\begin{aligned} & -\sigma^2\sqrt{\frac{h_n}{n}}\frac{1}{K(0)^2}\frac{h_n^2}{n^2}(tr(W^2)+tr(W'W))tr(W^3) \\ = \quad & -\sigma^2\sqrt{\frac{h_n}{n}}\frac{tr(W^3)}{(tr(W^2)+tr(W'W))} + o\left(\sqrt{\frac{h_n}{n}}\right), \end{aligned} \tag{C.31}$$

where the last equality follows since

$$K(0) = \frac{h_n}{n}(trW^2 + tr(W'W)) + O(\frac{1}{n}).$$

Similarly, the contribution to $\tilde{\kappa}_2$ comes from the term

$$-(it)^2\sigma^4\sqrt{\frac{h_n}{n}}\frac{1}{K(0)^2}\frac{h_n^2}{n^2}tr(W^3)tr((W+W')^2(\tilde{C}+\tilde{C}'))$$

and, by (C.17), is $o(\sqrt{h_n/n})$.

Finally, the contribution to $\tilde{\kappa}_3$ comes from the term

$$-\frac{1}{4}(it)^3\sigma^6\left(\frac{h_n}{n}\right)^{7/2}\frac{1}{K(0)^2}tr(W^3)(tr((W+W')^2))^2$$

and hence the actual contribution to $\tilde{\kappa}_3$ is

$$\begin{aligned} - \quad & 6\sigma^6\left(\frac{h_n}{n}\right)^{7/2}\frac{1}{K(0)^2}tr(W^3)(tr(W^2)+tr(W'W))^2 + o\left(\sqrt{\frac{h_n}{n}}\right) \\ = \quad & -6\sigma^6\left(\frac{h_n}{n}\right)^{3/2}tr(W^3) + o\left(\sqrt{\frac{h_n}{n}}\right), \end{aligned} \tag{C.32}$$

since

$$K(0) = \frac{h_n}{n}(trW^2 + tr(W'W)) + O(\frac{1}{n}).$$

Collecting (C.29) and (C.31), we conclude that the contribution to $\tilde{\kappa}_1$ from Q_1+Q_2 is

$$-2\sigma^2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(WW'W)}{K(0)} - \sigma^2\left(\frac{h_n}{n}\right)^{3/2}\frac{tr(W^3)}{K(0)} + o(\sqrt{\frac{h_n}{n}}).$$

The overall contribution to $\tilde{\kappa}_2$ from $Q_1 + Q_2$ is neglegible, while that to $\tilde{\kappa}_3$ is

$$-12\sigma^6 \left(\frac{h_n}{n}\right)^{3/2} tr(WW'W) - 6\sigma^6 \left(\frac{h_n}{n}\right)^{3/2} tr(W^3) + o\left(\sqrt{\frac{h_n}{n}}\right),$$

by collecting (C.30) and (C.32).

Acknowledgement

This research was supported by ESRC Grant RES-062-23-0036. I am particularly grateful to my Ph.D. advisor, P.M. Robinson, for invaluable guide and advice. I wish to thank Javier Hidalgo, Oliver Linton, Marcia Schafgans, Myung Seo, Jungyoon Lee, Abhimanyu Gupta and the participants at the work in progress seminars at LSE for valuable comments and suggestions.

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.

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

Bhattacharya, R.N. and R.R. Rao (1976). *Normal Approximation and Asymptotic Expansions.* John Wiley & Sons.

Buonanno, P., D. Montolio and P.Vanin (2009). Does Social Capital Reduce Crime? *Journal of Law and Economics* **52**, 145-70.

Case, A.C. (1991). Spatial Patterns in Household Demand. *Econometrica*, **59**, 953-65.

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

Conniffe, D. and J.E. Spencer (2001). When moments of ratios are ratios of moments. *Journal of the Royal Statistical Society. Series D (The Statistician)* **50**, 161-8.

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

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.

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

Jones, M.C. (1987). On moments of ratios of quadratic forms in normal variables. *Statistics and Probability letters* **6**, 129-36.

Kelejian, H.H. and I. Prucha. On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* **104**, 219-57.

Kelejian, H.H. and D.P. Robinson (1992). Spatial Autocorrelation - A new computationally simple test with an application to the per capita county police expenditures. *Regional Science and Urban Economics* **22**, 317-31.

Koopmans, T. (1942). Serial correlation and quadratic forms in normal variables. *The Annals of Mathematical Statistics***13**, 1, 14-33.

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.

Lieberman, O. (1994). A Laplace approximation to the moments of a ratio of quadratic forms. *Biometrika* **81**, 681-90.

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.

Roberts, L.A. (1995). On the existence of moments of ratios of quadratic forms. *Econometric Theory*, **11** 750-74.

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

Smith, M.D. (1989). On the expectation of a ratio of quadratic forms in normal variables. *Journal of multivariate analysis* **31**, 244-57.

Stanca, L. (2009). The geography of Economics and Happiness: spatial patterns in the effects of economic conditions on well being. Forthcoming, *Social Indicators Research.*

Tanaka, K. (1984). An asymptotic expansion associated with the Maximum Likelihood Estimators in ARMA models. *Journal of the Royal Statistical Society. Series B* **46**, 58-67.

Taniguchi, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. *Annals of Statistics* **11**, 157-69.

Taniguchi, M. (1986). Third order asymptotic properties of Maximum Likelihood Estimators for Gaussian ARMA processes. *Journal of multivariate analysis* **18**, 1-31.

Taniguchi, M. (1988). Asymptotic expansion of the distributions of some test statistics for Gaussian ARMA processes. *Journal of multivariate analysis* **27**, 494-511.

Taniguchi, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*. Springer-Verlag, Berlin.

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.

	$m=8$ $r=5$	$m=12$ $r=8$	$m=18$ $r=11$	$m=28$ $r=14$
$Pr(a\hat{\lambda} > z_\alpha \mid H_0)$	0	0	0.001	0.001
$Pr(a\hat{\lambda} > t^{Ed} \mid H_0)$	0.1250	0.1170	0.1100	0.0990
$Pr(g(a\hat{\lambda}) > z_\alpha \mid H_0)$	0.0560	0.0550	0.0520	0.0480
$Pr(a\hat{\lambda} > w^*_\alpha \mid H_0)$	0.0390	0.0610	0.0530	0.0540

Table 1: Empirical sizes of the tests of H_0 in (1.2) when λ is estimated by OLS and the sequence h_n is divergent. The reported values have to be compared with the nominal 0.05.

	$m=5$ $r=8$	$m=5$ $r=20$	$m=5$ $r=40$	$m=5$ $r=80$
$Pr(a\hat{\lambda} > z_\alpha \mid H_0)$	0.0010	0.0010	0.0010	0.0110
$Pr(a\hat{\lambda} > t^{Ed} \mid H_0)$	0.0960	0.0700	0.0570	0.0520
$Pr(g(a\hat{\lambda}) > z_\alpha \mid H_0)$	0.0550	0.0570	0.0550	0.0510
$Pr(a\hat{\lambda} > w^*_\alpha \mid H_0)$	0.0430	0.0400	0.057	0.055

Table 2: Empirical sizes of the tests of H_0 in (1.2) when λ is estimated by OLS and the sequence h_n is bounded. The reported values have to be compared with the nominal 0.05.

	$m=8$ $r=5$	$m=12$ $r=8$	$m=18$ $r=11$	$m=28$ $r=14$
$Pr\left(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > z_\alpha \vert H_0\right)$	0.0050	0.0060	0.0040	0.0130
$Pr\left(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed} \vert H_0\right)$	0.1180	0.0910	0.0740	0.0600
$Pr\left(\tilde{g}(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda}) > z_\alpha) \vert H_0\right)$	0.0560	0.0520	0.0520	0.0450
$Pr\left(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{w}^*_\alpha \vert H_0\right)$	0.0580	0.0520	0.0540	0.0460

Table 3: Empirical sizes of the tests of H_0 in (1.2) when λ is estimated by MLE and the sequence h_n is divergent. The reported values have to be compared with the nominal 0.05

.

	$m=5$ $r=8$	$m=5$ $r=20$	$m=5$ $r=40$	$m=5$ $r=80$
$Pr\left(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > z_\alpha \vert H_0\right)$	0.0120	0.0250	0.0320	0.0380
$Pr\left(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed} \vert H_0\right)$	0.0900	0.0750	0.0680	0.0490
$Pr\left(\tilde{g}(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda}) > z_\alpha) \vert H_0\right)$	0.0570	0.0550	0.0490	0.0510
$Pr\left(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{w}^*_\alpha \vert H_0\right)$	0.0620	0.0561	0.0582	0.0523

Table 4: Empirical sizes of the tests of H_0 in (1.2) when λ is estimated by MLE and the sequence h_n is bounded. The reported values have to be compared with the nominal 0.05.

	$m=8$ $r=5$		$m=12$ $r=8$		$m=18$ $r=11$		$m=28$ $r=14$	
$Pr(a\hat{\lambda} > z_\alpha \vert H_1)$	$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$	
	0.1	0	0.1	0	0.1	0.0050	0.1	0.0090
	0.5	0	0.5	0.3350	0.5	0.6730	0.5	0.8540
	0.8	0.2570	0.8	0.9940	0.8	1	0.8	1
$Pr(a\hat{\lambda} > t^{Ed} \vert H_1)$	$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$	
	0.1	0.5610	0.1	0.6100	0.1	0.6630	0.1	0.6930
	0.5	0.9520	0.5	0.9860	0.5	0.9930	0.5	1
	0.8	1	0.8	1	0.8	1	0.8	1
$Pr(a\hat{\lambda} > w^*_\alpha \vert H_1)$	$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$		$\bar{\lambda}$	
	0.1	0.1110	0.1	0.1190	0.1	0.1550	0.1	0.1640
	0.5	0.7250	0.5	0.8730	0.5	0.9380	0.5	0.9660
	0.8	0.9960	0.8	1	0.8	1	0.8	1

Table 5: Empirical powers of the tests of H_0 in (4.4) with $\bar{\lambda} = 0.1, 0.5, 0.8$ when λ is estimated by OLS and the sequence h_n is divergent. α is set to 0.95.

	$m=5$ $r=8$	$m=5$ $r=20$	$m=5$ $r=40$	$m=5$ $r=80$
$Pr(a\hat{\lambda} > z_\alpha \mid H_1)$	$\bar{\lambda}$ 0.1 0.0100 0.5 0.5510 0.8 0.9990	$\bar{\lambda}$ 0.1 0.0830 0.5 0.9880 0.8 1	$\bar{\lambda}$ 0.1 0.1870 0.5 1 0.8 1	$\bar{\lambda}$ 0.1 0.3630 0.5 1 0.8 1
$Pr(a\hat{\lambda} > t^{Ed} \mid H_1)$	λ 0.1 0.6400 0.5 0.9910 0.8 1	λ 0.1 0.7390 0.5 1 0.8 1	λ 0.1 0.8520 0.5 1 0.8 1	λ 0.1 0.6930 0.5 1 0.8 1
$Pr(a\hat{\lambda} > w^*_\alpha \mid H_1)$	$\bar{\lambda}$ 0.1 0.1390 0.5 0.8880 0.8 1	$\bar{\lambda}$ 0.1 0.2030 0.5 0.9920 0.8 1	$\bar{\lambda}$ 0.1 0.2960 0.5 1 0.8 1	$\bar{\lambda}$ 0.1 0.4510 0.5 1 0.8 1

Table 6: Empirical powers of the tests of H_0 in (4.4) with $\bar{\lambda} = 0.1, 0.5, 0.8$ when λ is estimated by OLS and the sequence h_n is bounded. α is set to 0.95.

	$m=8$ $r=5$	$m=12$ $r=8$	$m=18$ $r=11$	$m=28$ $r=14$
$Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > z_\alpha \mid H_1)$	λ 0.1 0.0100 0.5 0.4740 0.8 0.9850	λ 0.1 0.0370 0.5 0.7270 0.8 0.9990	λ 0.1 0.0380 0.5 0.8640 0.8 1	λ 0.1 0.0560 0.5 0.8930 0.8 1
$Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed} \mid H_1)$	$\bar{\lambda}$ 0.1 0.1270 0.5 0.7600 0.8 0.9900	$\bar{\lambda}$ 0.1 0.1300 0.5 0.8710 0.8 1	$\bar{\lambda}$ 0.1 0.1410 0.5 0.9270 0.8 1	$\bar{\lambda}$ 0.1 0.1740 0.5 0.9750 0.8 1
$Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{w}^*_\alpha \mid H_1)$	$\bar{\lambda}$ 0.1 0.0940 0.5 0.7480 0.8 0.9980	$\bar{\lambda}$ 0.1 0.1220 0.5 0.8560 0.8 1	$\bar{\lambda}$ 0.1 0.1300 0.5 0.9180 0.8 1	$\bar{\lambda}$ 0.1 0.1450 0.5 0.9990 0.8 1

Table 7: Empirical powers of the tests of H_0 in (4.4) with $\bar{\lambda} = 0.1, 0.5, 0.8$ when λ is estimated by MLE and the sequence h_n is divergent. α is set to 0.95.

	$m=5$ $r=8$	$m=5$ $r=20$	$m=5$ $r=40$	$m=5$ $r=80$
$Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > z_\alpha \mid H_1)$	λ 0.1 0.0510 0.5 0.7910 0.8 1	λ 0.1 0.1260 0.5 0.9890 0.8 1	λ 0.1 0.2070 0.5 0.9980 0.8 1	λ 0.1 0.4000 0.5 1 0.8 1
$Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed} \mid H_1)$	$\bar{\lambda}$ 0.1 0.1260 0.5 0.8820 0.8 1	$\bar{\lambda}$ 0.1 0.1920 0.5 0.9950 0.8 1	$\bar{\lambda}$ 0.1 0.2720 0.5 1 0.8 1	$\bar{\lambda}$ 0.1 0.4530 0.5 1 0.8 1
$Pr(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{w}^*_\alpha \mid H_1)$	$\bar{\lambda}$ 0.1 0.1140 0.5 0.8920 0.8 .	$\bar{\lambda}$ 0.1 0.1940 0.5 1 0.8 1	$\bar{\lambda}$ 0.1 0.3020 0.5 1 0.8 1	$\bar{\lambda}$ 0.1 0.5220 0.5 1 0.8 1

Table 8: Empirical powers of the tests of H_0 in (4.4) with $\bar{\lambda} = 0.1, 0.5, 0.8$ when λ is estimated by MLE and the sequence h_n is bounded. α is set to 0.95.

Rejection rule	$\alpha = 0.95$	$\alpha = 0.99$
$a\hat{\lambda} > z_\alpha$	reject H_0 $(1.713 > 1.645)$	fail to reject H_0 $(1.713 < 2.326)$
$a\hat{\lambda} > t^{Ed}$	reject H_0 $(1.713 > 1.287)$	reject H_0 $(1.713 > 1.666)$

Table 9: Outcomes of the tests of H_0 in (1.2) when λ in model (5.2) is estimated by OLS

Rejection rule	$\alpha = 0.95$	$\alpha = 0.99$
$\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > z_\alpha$	reject H_0 $(2.869 > 1.645)$	reject H_0 $(2.869 > 2.326)$
$\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{t}^{Ed}$	reject H_0 $(2.869 > 1.429)$	reject H_0 $(2.869 > 1.922)$

Table 10: Outcomes of the tests of H_0 in (1.2) when λ in model (5.2) is estimated by MLE

Rejection rule	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$
$a\hat{\lambda} > z_\alpha$	reject H_0 $(1.9998 > 1.645)$	fail to reject H_0 $(1.9998 < 2.326)$	fail to reject H_0 $(1.9998 < 2.5776)$
$a\hat{\lambda} > t^{Ed}$	reject H_0 $(1.9998 > 1.4042)$	reject H_0 $(1.9998 > 1.8821)$	fail to reject H_0 $(1.9998 < 2.0410)$

Table 11: Outcomes of the tests of H_0 in (1.2) when λ in model (5.5) is estimated by OLS

Rejection rule	$\alpha = 0.95$	$\alpha = 0.99$	$\alpha = 0.995$
$\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > z_\alpha$	reject H_0 $(2.2934 > 1.645)$	fail to reject H_0 $(2.2934 < 2.326)$	fail to reject H_0 $(2.2934 < 2.5776)$
$\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda} > \tilde{t}^{ed}$	reject H_0 $(2.2934 > 1.5227)$	reject H_0 $(2.2934 > 2.0767)$	reject H_0 $(2.2934 > 2.2704)$

Table 12: Outcomes of the tests of H_0 in (1.2) when λ in model (5.5) is estimated by MLE

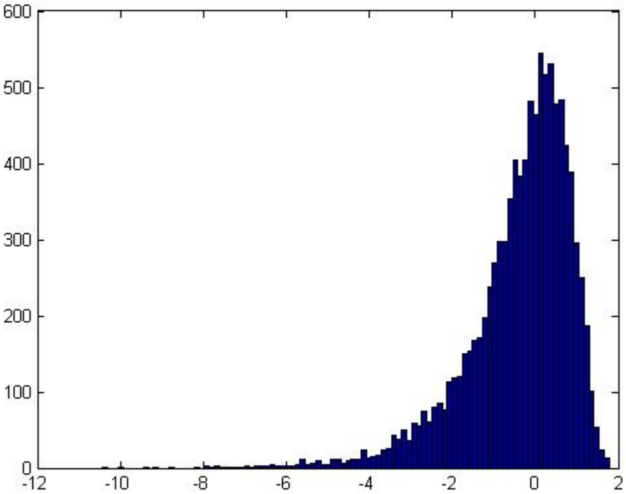

Figure 1: Empirical pdf of $a\hat{\lambda}$ under H_0

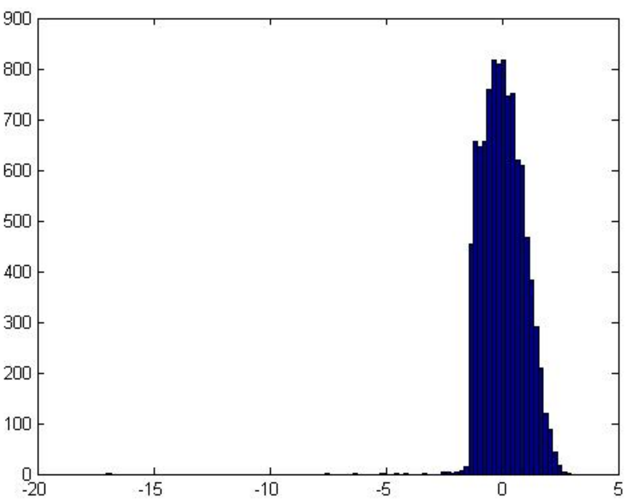

Figure 2: Empirical pdf of $g(a\hat{\lambda})$ under H_0

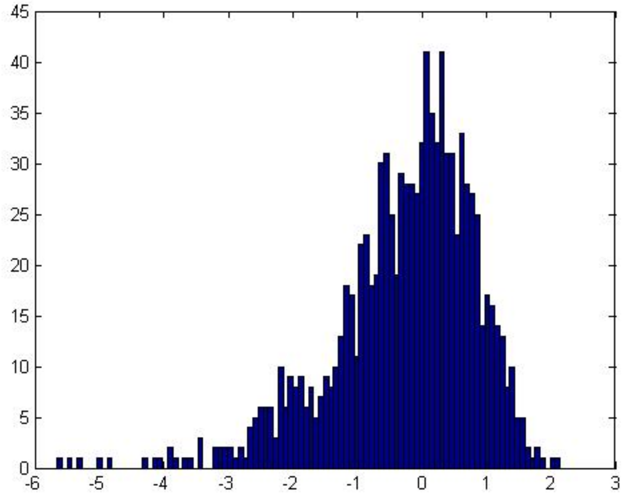

Figure 3: Empirical pdf of $\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda}$ under H_0

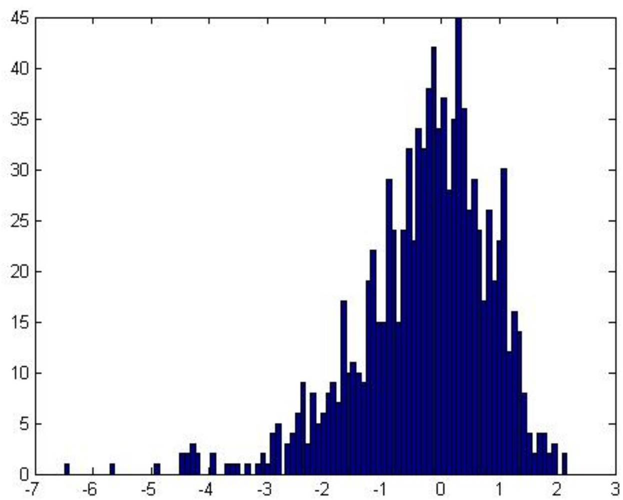

Figure 4: Empirical pdf of $\tilde{g}(\sqrt{\frac{n}{h_n}}\tilde{a}\tilde{\lambda})$ under H_0