Working Paper Series
Department of Economics
University of Verona

Higher-Order Least Squares Inference for Spatial Autoregressions

Francesca Rossi, Peter M. Robinson

WP Number: 4 March 2020

ISSN: 2036-2919 (paper), 2036-4679 (online)

Higher-Order Least Squares Inference for Spatial Autoregressions

Francesca Rossi * Peter M. Robinson †

March 3, 2020

Abstract

We develop refined inference for spatial regression models with predetermined regressors. The ordinary least squares estimate of the spatial parameter is neither consistent, nor asymptotically normal, unless the elements of the spatial weight matrix uniformly vanish as sample size diverges. We develop refined testing of the hypothesis of no spatial dependence, without requiring negligibility of spatial weights, by formal Edgeworth expansions. We also develop higher-order expansions for both an unstudentized and a studentized transformed estimator, where the studentized one can be used to provide refined interval estimates. A Monte Carlo study of finite sample performance is included.

JEL Classifications: C12, C13, C21

Keywords: Spatial autoregression; least squares estimation; higher-order inference; Edgeworth expansion; testing spatial independence.

1 Introduction

Spatial autoregressions (SARs, henceforth) have been broadly applied in various fields of economics over the past few decades. The main advantage of SARs is their parsimonious functional form, which embeds the notion of pairwise spatial proximity between units in the so-called weight matrix, exogenously chosen by the practitioner in terms of a general economic distance. Thus, SARs are particularly appealing as they allow a straightforward interpretation of estimates in terms of marginal effects, accounting for the feedback generated by the network structure described by the weight matrix.

Various estimation methods for parameters of standard SARs for cross-sectional data and their asymptotic theory have been broadly developed in the recent literature. These include instrumental variables/two-stage least squares methods (e.g. Kelejian and Prucha (1998)), Gaussian maximum

*Corresponding author. Department of Economics, University of Verona. e-mail: francesca.rossi_02@univr.it. Address: via Cantarane 24, 37129, Verona.

†Department of Economics, London School of Economics and Political Science. e-mail: p.m.robinson@lse.ac.uk. Address: Houghton Street, WC2A 2AE, London.

likelihood/quasi-maximum likelihood estimation (e.g. Ord (1975) and Lee (2004)) and techniques based on generalised methods of moments (e.g. Kelejian and Prucha (1999) and Lee (2007)). Also, asymptotic theory for various tests for lack of spatial correlation has been widely derived and discussed, e.g. Burridge (1980), Cliff and Ord (1981), Kelejian and Prucha (2001), Anselin (2001), Robinson (2008), Lee and Yu (2012), Martellosio (2012) and Delgado and Robinson (2015).

A considerable body of literature has focussed on finite-sample behaviour of estimates and test statistics for SARs. Although asymptotic properties are favourable under general conditions, performance of estimates and test statistics in small/moderate samples might not be very satisfactory. More specifically, Bao and Ullah (2007) derived a Nagar-type of expansion to evaluate the second-order bias and mean square error of the Gaussian maximum likelihood estimator for the spatial parameter of SAR models without regressors, and their work has been extended in Bao (2013) to accommodate exogenous regressors and non-Gaussian errors. More generally, Yang (2015) developed higher-order bias and variance corrections by means of stochastic expansions and bootstrap for a class of non-linear models that includes SAR as a special case. More recently, Martellosio and Hillier (2019) derived refined estimates of the spatial parameter of SARs by centring the associated profile score function, and constructed confidence sets using a Lugannani-Rice approximation. So far as improved tests are concerned, various refinements of test statistics have been derived by Cliff and Ord (1981), Robinson (2008), Baltagi and Yang (2013), Robinson and Rossi (2014, 2015), Liu and Yang (2015) and Jin and Lee (2015).

In addition to the aforementioned estimation methods and test statistics, Lee (2002) developed asymptotic theory for inference on parameters of SARs based on the ordinary least squares (OLS, henceforth) principle. The OLS estimator is desirable as it enjoys a simple closed-form, but it is consistent for the spatial parameter only under some stringent assumptions on the network structure in the limit. Specifically, OLS estimates of spatial parameters are consistent only if the elements of the weight matrix are uniformly negligible as sample size increases, and have a limiting standard normal distribution only if they vanish at a suitably faster rate. Such conditions are restrictive and, even more importantly, they are difficult to check in practical cases when only a finite number of observations is available. By means of a formal Edgeworth expansion, Robinson and Rossi (2015) developed a refined t-type test for lack of spatial correlation in SARs without exogenous regressors based on the OLS estimate of the spatial parameter, and they showed its consistency under general assumptions on the weight matrix. However, the framework of Robinson and Rossi (2015) did not allow the construction of improved confidence sets for the spatial parameter. On the other hand, Kyriacou et al. (2017) derived a new OLS-based estimator for the spatial parameter of SARs without exogenous regressors, by means of an indirect inference transformation that restored consistency and asymptotic normality under general network structures.

In this paper, we consider the standard spatial autoregression with predetermined, thus exogenous, regressors

$$y = \lambda Wy + X\beta + \epsilon, \tag{1.1}$$

where W is an $n \times n$ spatial weight matrix with $(i,j)-$th element W_{ij}, y is an $n \times 1$ vector of observations, X is an $n \times k$ matrix of exogenous regressors, and ϵ is an $n \times 1$ vector of independent and identically distributed (i.i.d.) disturbances, with zero mean and unknown variance σ^2. Throughout, we drop the subscript n in $y_n = y$, $W_n = W$, $X_n = X$ and $\epsilon_n = \epsilon$, even though such quantities are, in general, triangular arrays. We focus on estimation and inference on the unknown scalar λ, which is often the parameter of interest in model (1.1). We first derive a formal second-order Edgeworth expansion for the cumulative distribution function (cdf, in the sequel) of the OLS estimate of λ, suitably centred so that the expansion is justified even without uniform negligibility of W_{ij} as sample size increases. Such a formal expansion provides the basis to derive improved tests on λ, after suitable studentization, under the null hypothesis of interest, by means of either Edgeworth-corrected quantiles or corrected test statistics. In order to construct point and interval estimates for λ which are consistent under general assumptions on W, we then introduce a monotonic transformation of the OLS estimate of λ and derive the second-order Edgeworth expansion for its cdf. A studentized version of this expansion provides second-order corrected confidence sets for λ. The advantage of our method over ones based on implicitly-defined estimators is that the simple closed form of the OLS of λ allows straightforward implementation regardless of the W structure, which is reflected by very satisfactory Monte Carlo results in small/moderately-sized samples.

The derivation of the expansion for the suitably centred, standardized cdf of the OLS estimate of λ in (1.1) is presented in Section 2, while its studentized version and application to testing follows in Section 3. In Section 4 we derive the formal expansion for the cdf of the standardized transformed-OLS, and in Section 5 we construct Edgeworth-refined confidence sets by means of the studentized variant of the expansion presented in the previous section. A brief discussion of SAR models with no exogenous regressors is presented in Section 6, while a Monte Carlo exercise to assess the finite sample performance of our refined, OLS-based tests and confidence sets is reported in Section 7.

Throughout, B_{ij} indicates $i-j$th element of the generic $p \times q$ matrix. $\eta_i(A)$, $i = 1, ...q$ denote the eigenvalues of a generic $q \times q$ matrix A, while $\bar{\eta}(A) = \max_{i=1,....q} \{|\eta_i(A)|\}$ and $\underline{\eta}(A) = \min_{i=1,....q} \{|\eta_i(A)|\}$. Also, for $s = 0, 1, ..., A^s$ denotes the $s-$th power of A, with a similar convention for A'^s, in turn denoting the $s-$th power of A', which is the transpose of A. Also, $||.||$ indicates the spectral norm, i.e. for any $p \times q$ matrix B, $||B||^2 = \bar{\eta}(B'B)$, whereas the absolute row sum norm of a $q \times q$ matrix A with $(i, j)-$th element A_{ij} is $||A||_\infty = \max_i \sum_{j=1}^{q} |A_{ij}|$. Let K be a finite, positive, generic constant, c an arbitrarily small positive constant, and $I = I_n$ the $n \times n$ identity matrix. Also, let $f^{(i)}(x) = d^i f(x)/dx^i$ for a generic scalar function $f(\cdot)$ and scalar x, while for a generic $k \times 1$ vector z, $\partial f(z)/\partial z = (\partial f(z)/\partial z_1, \partial f(z)/\partial z_2, ...\partial f(z)/\partial z_k)'$. Finally, $f^{-1}(\cdot)$ denotes the inverse function of $f(\cdot)$, with the obvious implication that $f^{-1(i)}(\cdot)$ represents the total derivative of order i of $f^{-1}(\cdot)$.

2 An Edgeworth expansion for the ordinary least squares estimator

We consider the model (1.1) and its reduced form

$$y = S^{-1}(\lambda)(X\beta + \epsilon), \tag{2.1}$$

provided that the inverse of $S(\lambda) = I - \lambda W$ exists (as implied by Assumptions 2 and 3 below).

The OLS estimates $\hat{\lambda}$, $\hat{\beta}$ of λ and β are given by

$$\begin{pmatrix} \hat{\lambda} - \lambda \\ \hat{\beta} - \beta \end{pmatrix} = M^{-1}u, \tag{2.2}$$

where

$$M = \begin{bmatrix} m_{11} & m_{12} \\ m_{12}' & m_{22} \end{bmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \tag{2.3}$$

and

$$m_{11} = y'W'Wy, \quad m_{12} = y'W'X, \quad m_{22} = X'X, \quad u_1 = y'W'\epsilon \quad u_2 = X'\epsilon. \tag{2.4}$$

Lee (2002) showed that $\hat{\beta}$ is consistent under very general model assumptions, while $\hat{\lambda}$, is consistent under

$$\overline{\lim_{n\to\infty}} \max_{1\le i,j\le n} |W_{ij}| = O\left(\frac{1}{h}\right), \quad \text{where} \quad \frac{1}{h} + \frac{h}{n} \to 0 \text{ as } n \to \infty, \tag{2.5}$$

with $h = h_n$ being a positive, deterministic, sequence.

Denote by ϵ_i the $i-$th element if ϵ, and introduce the following assumptions.

Assumption 1 *The ϵ_i are independent normal random variables with mean zero and unknown variance*

σ^2.

Assumption 2 *$\lambda \in \Lambda$, where $\Lambda = [b_1, b_2]$ with $-1 < b_1 < b_2 < 1$.*

Assumption 3

(i) *For all i, n, $W_{ii} = 0$.*

(ii) *For all n, $||W|| = 1$.*

(iii) *For all sufficiently large n, $||W||_\infty + ||W'||_\infty \leq K$.*

(iv) *$\overline{\lim}_{n\to\infty} \max_{1\leq i,j\leq n} |W_{ij}| = O\left(h^{-1}\right)$, where $h/n \to 0$ as $n \to \infty$.*

Assumption 4 *For all sufficiently large n, $\sup_{\lambda\in\Lambda} \left(||S^{-1}(\lambda)||_\infty + ||S^{-1}(\lambda)'||_\infty\right) < K$.*

Assumption 5 *For all $i, j = 1, ..., n$ and for all n, each element X_{ij} of X is predetermined and $|X_{ij}| \leq K$. Moreover,*

$$0 < c < \underline{\eta}\left(\frac{X'X}{n}\right)$$

for all sufficiently large n.

Normality of the ϵ_i is an unnecessarily strong condition to derive first-order results, but it is familiar in higher-order asymptotic theory as Edgeworth expansions would otherwise be complicated by the presence of cumulants of ϵ_i. Assumptions 2 and 3(ii) guarantee that $S^{-1}(\lambda)$ exists for all $\lambda \in \Lambda$. It is well documented (e.g. Kelejian and Prucha (2010)) that either Assumptions 2 and 3(ii) or similar ones are necessary in order to justify SAR as an equilibrium model and develop asymptotic theory. Assumption 3(iv) establishes the limit behaviour of W as $n \to \infty$. Unlike (2.5), we allow h to be either divergent or bounded.

Let $P = I - X(X'X)^{-1}X'$, $G(\lambda) \equiv G = WS^{-1}(\lambda)$ and $g_{st} = tr(G^s G'^t)/n$, for $s, t = 0, 1, ...$, such that, for instance, $g_{10} = tr(G)/n$. Also, let $\delta(A) = \beta'X'G'PAPGX\beta/n$, for a generic $n \times n$ matrix A such that $||A||_\infty + ||A'||_\infty < K$. Under Assumptions 2-5, for all $s, t = 0, ..., n$, uniformly in $\lambda \in \Lambda$, $||G(\lambda)||_\infty + ||G'(\lambda)||_\infty < K$, $g_{st} = O(1)$ and $\delta(A) = O(1)$, as $n \to \infty$. We also impose

Assumption 6 $\underline{\lim}_{n\to\infty} \ \delta(I) > 0$.

Assumption 6 does not require the limit of $\delta(I)$ to exist, but it ensures that the leading terms of the expansion in Theorems 1-4 below are well defined. Also, Assumption 6 rules out the case $\beta = 0$ a priori, which was covered by Robinson and Rossi (2015), and the possibility that the columns of G and X are perfectly collinear in the limit, e.g. Assumption 6 implies rank$(G'PG) \sim n$ as $n \to \infty$, where $\alpha \sim n$ denotes $\alpha/n \to K \in (0, \infty)$ for a generic sequence $\alpha = \alpha_n$.

In general, $\hat{\lambda}$ is inconsistent for λ under Assumptions 1-6, since

$$\hat{\lambda} - \lambda \underset{p}{\to} \lim_{n\to\infty} \frac{\sigma^2 tr(G)/n}{\beta' X' G' P G X \beta/n + \sigma^2 tr(G'G)/n}, \tag{2.6}$$

where the RHS tends to zero under (2.5) as $n \to \infty$, but not under the weaker Assumption 3(iv).

The requirement on h to derive a central limit theorem for $\sqrt{n}(\hat{\lambda} - \lambda)$ is even stricter than that imposed by (2.5), as $h/\sqrt{n} \to \infty$ as $n \to \infty$. In practical terms, the RHS of (2.6) vanishes if the number of neighbours of each unit increases with n and $\sqrt{n}(\hat{\lambda} - \lambda)$ is asymptotically normal only if this number increases faster than $\sqrt{n}$. Such conditions are difficult to check in practice, where n is finite. As an illustration, $\hat{\lambda}$ would not be consistent under the popular contiguity-based choice of W, in which $W_{ij} = 1$ if unit i and unit j share a border, and $W_{ij} = 0$ otherwise. On the other hand, though, OLS desirably has a simple closed form and is unaffected by sparsity of W.

We introduce

$$\psi_n(x) = \psi(x) = x + \frac{y' S(x)' P S(x) y}{y' W' P W y} \frac{1}{n} tr(G(x)), \tag{2.7}$$

which is well defined as n increases under Assumption 6. As $n \to \infty$,

$$\psi(\lambda) - \lambda \overset{p}{\to} \lim_{n\to\infty} \frac{\sigma^2 g_{10}}{\delta(I) + \sigma^2 g_{11}}, \tag{2.8}$$

so long as the limit exists, this corresponding to the RHS of (2.6). In particular, $\psi(\lambda) = \lambda$ for any n when $\lambda = 0$ and $\psi(\lambda) \to_p \lambda$ as long as $h \to \infty$ as n increases. Since $\hat{\lambda}$ is inconsistent for general λ and h under Assumptions 1-6, in the sequel we focus on deriving the Edgeworth expansion of an adjusted OLS estimator (e.g. Maekawa (1985)), namely $\hat{\lambda} - \psi(\lambda)$.

Define

$$a = \delta(I) + \sigma^2(g_{20} + g_{11} - 2g_{10}^2), \tag{2.9}$$

$$t = \frac{\delta(I) + \sigma^2 g_{11}}{(\sigma^2 a)^{1/2}}, \tag{2.10}$$

$$b = \delta(G) - g_{10}\delta(I) + \frac{1}{3}\sigma^2(g_{30} + 3g_{21}) + \frac{8}{3}\sigma^2 g_{10}^3 - 2\sigma^2 g_{10}(g_{20} + g_{11}) \tag{2.11}$$

and the quadratic polynomial

$$e(x) = \frac{\sigma}{a^{1/2}} \left(tr(G' X (X'X)^{-1} X') - k g_{10} \right) + \frac{2}{ta}(\delta(G) + \sigma^2 g_{21} - \sigma^2 g_{11} g_{10}) x^2 - \frac{\sigma b}{a^{3/2}}(x^2 - 1). \tag{2.12}$$

In the Appendix we prove

Theorem 1 *Suppose that model (1.1) and Assumptions 1-6 hold. For any real number ζ, the cdf of $\hat{\lambda}$ admits the second order formal Edgeworth expansion*

$$Pr(\sqrt{n}\left(\hat{\lambda}-\psi(\lambda)\right) \leq \zeta) = \Phi(t\zeta) + \frac{1}{\sqrt{n}} e(t\zeta)\phi(t\zeta) + o\left(\frac{1}{\sqrt{n}}\right), \tag{2.13}$$

where $e(t\zeta) = O(1)$ as $n \to \infty$.

The RHS of (2.13) is well defined under Assumption 6. We stress that in general Theorem 1 is infeasible as it depends on unknown parameters. Equivalently, we can write Theorem 1 for $\zeta = t^{-1}\eta$ and we can re-write (2.13) as

$$Pr(\sqrt{n}t\left(\hat{\lambda}-\psi(\lambda)\right) \leq \eta) = \Phi(\eta) + \frac{1}{\sqrt{n}} e(\eta)\phi(\eta) + o\left(\frac{1}{\sqrt{n}}\right). \tag{2.14}$$

A remark on the random centring sequence $\psi(\lambda)$ is appropriate. Instead of $\hat{\lambda} - \psi(\lambda)$, we could consider

$$\hat{\lambda} - \lambda - \frac{\sigma^2 g_{10}}{\delta(I) + \sigma^2 g_{11}}, \tag{2.15}$$

but this involves also the unknown σ^2 and β, which would need to be consistently estimated in order to obtain an operational version of (2.13). As already mentioned, under Assumptions 2-6 and for general values of λ, OLS would be consistent for β, but not for λ and σ^2. If the main application of Theorem 1 is limited to the derivation of corrected tests on λ, which is the focus of Section 3, β and σ^2 could be easily estimated under the null hypothesis. A feasible corrected OLS estimator for λ in (1.1) and its higher order expansion will be discussed in Section 4.

3 Refined test for lack of spatial correlation

In this section, we derive the second-order expansion of the studentized version of (2.14) to develop refined tests of

$$H_0: \quad \lambda = 0 \tag{3.1}$$

against

$$H_1: \quad \lambda > (<)\ 0. \tag{3.2}$$

Under H_0, several simplifications to the expressions displayed in Section 2 are possible. Specifically, $G = W$, $g_{st} = tr(W^s W'^t)/n$, $g_{10} = 0$ and thus $\psi(\lambda) = \lambda = 0$, $\delta(A) = \beta' X' W' PAPWX\beta/n$ (for any generic $n \times n$ matrix A) and $a = \delta(I) + \sigma^2(g_{20} + g_{11})$, with analogous simplifications for b and $e(\eta)$. Define

$$\hat{t} = \frac{g_{11}\hat{\sigma}^2 + \hat{\delta}(I)}{(\hat{\sigma}^2\hat{a})^{1/2}}, \tag{3.3}$$

where $\hat{\delta}(I)$ and $\hat{a}$ are the estimates of $\delta(I)$ and a obtained by replacing the unknown β and σ^2 by their OLS estimates under H_0 in (3.1).

The following theorem is proved in the Appendix

Theorem 2 *Suppose that model (1.1) and Assumptions 1-6 hold. For any real number ζ, under H_0 in (3.1) the studentized cdf of $\hat{\lambda}$ admits the second order formal Edgeworth expansion*

$$Pr(\sqrt{n}\hat{t}\hat{\lambda} \leq \eta) = \Phi(\eta) + \frac{1}{\sqrt{n}}e(\eta)\phi(\eta) + o\left(\frac{1}{\sqrt{n}}\right), \tag{3.4}$$

with $e(\eta)$ defined in (2.12) and $e(\eta) = O(1)$ as $n \to \infty$.

The RHS of (3.4) is well defined under Assumption 6. Theorem 2 shows that studentization of $\hat{\lambda}$ does not alter higher order terms compared to (2.14) for $\lambda = 0$. Also, we can replace unknown λ, β and σ^2 in $e(\cdot)$ defined in (2.12) by $\lambda = 0$, $\hat{\beta}$ and $\hat{\sigma}^2$ respectively, where $\hat{\beta}$ and $\hat{\sigma}^2$ are OLS estimates under H_0, to obtain a feasible variant of $e(\cdot)$, $\hat{e}(\cdot)$. As $\hat{\beta}$ and $\hat{\sigma}^2$ are consistent under H_0, the order of the remainder in (3.4) is not affected when we replace $e(\cdot)$ by $\hat{e}(\cdot)$.

According to standard first order asymptotic theory, we reject H_0 in (3.1) if

$$\sqrt{n}\hat{t}\hat{\lambda} > z_{1-\alpha}, \tag{3.5}$$

where $z_{1-\alpha}$ denote the $1-\alpha$ quantile of the standard normal distribution. From (3.4), a test based on (3.5) has approximate size α. From (3.4), by inversion we can derive a refined test that rejects H_0 in (3.1) when

$$\sqrt{n}\hat{t}\hat{\lambda} > z_{1-\alpha} - \frac{1}{\sqrt{n}}\hat{e}(z_{1-\alpha}). \tag{3.6}$$

Rather than correcting the critical value, from (3.4) we can construct a transformation of the test static itself (i.e. Yanagihara et al. (2005)), such as

$$l(x) = x + \frac{1}{\sqrt{n}}e(x) + \frac{1}{4n}\int \left(e^{(1)}(x)\right)^2 dx = x + \frac{1}{\sqrt{n}}e(x) + \frac{1}{12n}\left(\frac{4}{at}(\delta(W) + \sigma^2 g_{21}) - \frac{\sigma b}{a^{3/2}}\right)^2 x^3, \tag{3.7}$$

where again $e(\cdot)$ is defined according to (2.12). We indicate by $\hat{l}(\cdot)$ the feasible version of $l(\cdot)$, obtained by replacing unknown parameters by their estimates under H_0.

We denote by $w_{1-\alpha}$ the true $1-\alpha$ quantile of $\sqrt{n}\hat{t}\hat{\lambda}$, so that a test that rejects when $\sqrt{n}\hat{t}\hat{\lambda} > w_{1-\alpha}$ has exact size α. From Theorem 2 we deduce

Corollary 1 *Suppose that model (1.1) and Assumptions 1-5 hold.*

$$w_{1-\alpha} = z_{1-\alpha} + O\left(\left(\frac{1}{n}\right)^{1/2}\right) \tag{3.8}$$

$$= z_{1-\alpha} - \frac{1}{\sqrt{n}}\hat{e}(z_{1-\alpha}) + o\left(\left(\frac{1}{n}\right)^{1/2}\right) \tag{3.9}$$

and

$$P(\hat{l}(\sqrt{n}\hat{t}\hat{\lambda}) > z_{1-\alpha}) = \alpha + o\left(\left(\frac{1}{n}\right)^{1/2}\right). \tag{3.10}$$

A similar standardization procedure, and thus similar refined tests, could also be derived for $H_0 : \lambda = \lambda_0 \neq 0$, at expense of some extra computational burden, by using the estimates of β and σ^2 under H_0, i.e.

$$\hat{\beta} = (X'X)^{-1}X'S(\lambda_0)y, \quad \hat{\sigma}^2 = \frac{1}{n}y'S(\lambda_0)'PS(\lambda_0)y. \tag{3.11}$$

Since $\psi(\lambda_0)$ does not contain any unknown parameters under H_0, the whole derivation reported in the proof of Theorem 2 would go through with virtually no modification in case $\lambda_0 \neq 0$, apart from some slightly more cumbersome algebraic expressions.

4 Corrected ordinary least squares estimator

As already mentioned, Theorems 1 and 2 are useful to deduce refined inference procedures when the main interest is testing the significance of the spatial parameter. However, the studentized version of (2.13), and thence the refined confidence sets, cannot be derived in general cases, as neither λ nor σ^2 can be consistently estimated by OLS unless (2.5) holds. Therefore, in this section we introduce a corrected estimator, denoted $\hat{\lambda}_C$ in the sequel, and a higher order expansion for its cdf. We impose

Assumption 7 *For all sufficiently large n, $\psi(\lambda)$ defined in (2.7) is strictly increasing for all $\lambda \in \Lambda$ with probability one.*

From (2.7), $\psi(\lambda)$ is continuously differentiable for all $\lambda \in \Lambda$, with positive first derivative under Assumption 7. The latter is a high level assumption that can be verified numerically in each empirical case, as $\psi(\cdot)$ does not depend on any nuisance parameter. In Section 7 we will report some plots of $\psi(\cdot)$ for a few data generating processes.

Under Assumption 7, define $\hat{\lambda}_C = \psi^{-1}(\hat{\lambda})$. Heuristically, for any real ζ we can write

$$
\begin{aligned}
Pr\left(\sqrt{n}(\hat{\lambda} - \psi(\lambda)) \leq \zeta\right) =& Pr\left(\hat{\lambda} \leq \psi(\lambda) + \frac{\zeta}{\sqrt{n}}\right) = Pr\left(\psi^{-1}(\hat{\lambda}) \leq \psi^{-1}\left(\psi(\lambda) + \frac{\zeta}{\sqrt{n}}\right)\right) \\
=& Pr\left(\hat{\lambda}_C \leq \psi^{-1}(\psi(\lambda)) + \psi^{-1\ (1)}(\psi(\lambda))\frac{\zeta}{\sqrt{n}} + o_p\left(\frac{1}{\sqrt{n}}\right)\right) \\
=& Pr\left(\hat{\lambda}_C \leq \lambda + \frac{\zeta}{\sqrt{n}\psi^{(1)}(\lambda)} + o_p\left(\frac{1}{\sqrt{n}}\right)\right) \\
=& Pr\left(\sqrt{n}\psi^{(1)}(\lambda)(\hat{\lambda}_C - \lambda) + o_p(1) \leq \zeta\right) \\
=& Pr\left(\sqrt{n}\psi^{(1)}(\lambda)(\hat{\lambda}_C - \lambda) \leq \zeta\right) + o(1). \qquad (4.1)
\end{aligned}
$$

By combining Theorem 1 and the equivalence in (4.1), we deduce that under Assumptions 1-7

$$
\sqrt{n}(\hat{\lambda}_C - \lambda) \xrightarrow{d} \mathcal{N}(0, v), \quad \text{with} \quad v = p\lim_{n\to\infty}(t\psi^{(1)}(\lambda))^{-2} \qquad (4.2)
$$

provided that the limit v exists, where

$$
v = \lim_{n\to\infty}\left(\frac{\delta(I) + \sigma^2 g_{11}}{(\sigma^2 a)^{1/2}}\left(1 - \frac{2\sigma^2 g_{10}^2 - \sigma^2 g_{20}}{\delta(I) + \sigma^2 g_{11}}\right)\right)^{-2} = \lim_{n\to\infty}\left(\left(\frac{a}{\sigma^2}\right)^{1/2}\right)^{-2} = \lim_{n\to\infty}\frac{\sigma^2}{a}, \qquad (4.3)
$$

which is equivalent to the variance of the Gaussian maximum likelihood estimator (MLE) of λ (Lee (2004)). The asymptotic distribution result in (4.2) is expected to be robust to some departures from normality of the ϵ_i.

Let

$$
\bar{\iota} = \frac{2}{a^2}\left(\delta(I)g_{10} + \sigma^2(g_{11}g_{10} - 2g_{20}g_{10} + g_{30})\right) \qquad (4.4)
$$

and

$$
b_C = \sigma^2\left(\frac{g_{30}}{3} + g_{21} + \frac{8}{3}g_{10}^3 - 2g_{10}(g_{20} + g_{11})\right) + \delta(G) - g_{10}\delta(I) + \frac{\bar{\iota}}{2}(a - \delta(I))^2. \qquad (4.5)
$$

Define the quadratic polynomial

$$
\begin{aligned}
e_C(x) =& \left(\frac{\sigma^2}{a}\right)^{1/2}\left(tr(GX(X'X)^{-1}X') - g_{10}k - \frac{\bar{\iota}a}{2}\right) + \frac{2\sigma}{a^{3/2}}\left(\sigma^2(2g_{10}^3 + g_{21} - 2g_{11}g_{10} - g_{20}g_{10})\right. \\
&\left. + \delta(G) - \delta(I)g_{10}\right)x^2 - \frac{\sigma b_C}{a^{3/2}}(x^2 - 1), \qquad (4.6)
\end{aligned}
$$

where a, $\bar{\iota}$ and b_C are as in (2.9), (4.4) and (4.5), respectively.

From (4.1) and under Assumption 7, in order to derive the higher order expansion of the cdf of

$\hat{\lambda}_C - \lambda$, we consider the equivalence

$$Pr\left(\hat{\lambda}_C - \lambda \leq \frac{\zeta}{\sqrt{n}}\right) = Pr\left(\psi^{-1}(\hat{\lambda}) - \psi^{-1}(\psi(\lambda)) \leq \frac{\zeta}{\sqrt{n}}\right) \tag{4.7}$$

and prove in the Appendix the following

Theorem 3 *Suppose that model (1.1) and Assumptions 1-7 hold. For any real number ζ, the cdf of $\hat{\lambda}_C$ admits the second order formal Edgeworth expansion*

$$Pr(\sqrt{n}\left(\hat{\lambda}_C - \lambda\right) \leq \zeta) = \Phi\left(\left(\frac{a}{\sigma^2}\right)^{1/2}\zeta\right) + \frac{1}{\sqrt{n}}e_C\left(\left(\frac{a}{\sigma^2}\right)^{1/2}\zeta\right)\phi\left(\left(\frac{a}{\sigma^2}\right)^{1/2}\zeta\right) + o\left(\frac{1}{\sqrt{n}}\right), \tag{4.8}$$

with $e_C(\cdot)$ defined as in (4.6) and $e_C\left(\left(\frac{a}{\sigma^2}\right)^{1/2}\zeta\right) = O(1)$ as $n \to \infty$.

Again, non-singularity of a as n increases is guaranteed under Assumption 6. The first-order limit result in (4.2) is contained in (4.8), provided that the limit of a exists. The expansion in Theorem 3 is infeasible, as it depends on unknown λ, β and σ. In order to construct improved confidence sets we need to derive the studentized version of Theorem 3, as developed in the following section.

5 Refined confidence intervals for the spatial parameter

Let $t_C = \left(a/\sigma^2\right)^{1/2}$, where a is defined as in (2.9). We estimate β and σ^2 by

$$\hat{\beta}_C = (X'X)^{-1}X'S(\hat{\lambda}_C)y, \qquad \hat{\sigma}^2_C = \frac{1}{n}y'S(\hat{\lambda}_C)'PS(\hat{\lambda}_C)y, \tag{5.1}$$

and plug these in a, defined in (2.9), to approximate t_C by $\hat{t}_C = \left(\hat{a}_C/\hat{\sigma}^2_C\right)^{1/2}$.

Let

$$\nu_{CS} = \frac{1}{(a^3\sigma^2)^{1/2}}(\delta(G) - \delta(I)g_{10} + \sigma^2(g_{21} - 2g_{11}g_{20} - g_{10}(g_{11} - 2g_{20}))), \tag{5.2}$$

$$b_{CS} = \sigma^2\left(\frac{g_{30}}{3} + g_{21} + \frac{8}{3}g_{10}^3 - 2g_{10}(g_{20} + g_{11})\right) + \delta(G) + a^{3/2}\sigma\nu_{CS} - \frac{\delta(I)}{a^{1/2}}\left(\sigma\nu_{CS}\delta(I) + \frac{g_{10}}{a^{3/2}}(a - \delta(I))^2\right) \tag{5.3}$$

and the quadratic polynomial

$$e_{CS} = \left(\frac{\sigma^2}{a^3}\right)^{1/2} \left(atr(GX(X'X)^{-1}X') - ag_{10}k - \delta(G) - \sigma^2(g_{21} - 2g_{11}g_{20} - g_{10}(g_{11} - 2g_{20})\right) \quad (5.4)$$

$$+2\left(\frac{\sigma^2}{a^3}\right)^{1/2} \left((2g_{10}^3 + g_{21} - 2g_{11}g_{10} - g_{20}g_{10})\sigma^2 + \delta(G) - \delta(I)g_{10}\right) x^2 - \frac{\sigma b_{CS}}{a^{3/2}}(x^2 - 1). \quad (5.5)$$

In the Appendix we prove

Theorem 4 *Suppose that model (1.1) and Assumptions 1-7 hold. For any real number* η*, the cdf of* $\sqrt{n}\hat{t}_C(\hat{\lambda}_C - \lambda)$ *admits the second order formal Edgeworth expansion*

$$Pr(\sqrt{n}\hat{t}_C\left(\hat{\lambda}_C - \lambda\right) \leq \eta) = \Phi(\eta) + \frac{1}{\sqrt{n}} e_{CS}(\eta)\,\phi(\eta) + o\left(\frac{1}{\sqrt{n}}\right), \quad (5.6)$$

and $e_{CS}(\eta) = O(1)$ *as* $n \to \infty$.

From Theorem 4 we can construct refined confidence sets for λ. We focus on the one-sided interval (L, ∞), where L is a suitable lower-end point, although the same type of correction can be deduced for $(-\infty, U)$, U being an upper-end point. Define $\mathcal{I} = \left(\hat{\lambda}_C - w^C_{1-\alpha}/(\hat{t}^C\sqrt{n}), \infty\right)$, where w^C_α denotes the true $\alpha-$quantile of the c.d.f. of $\sqrt{n}\hat{t}_C(\hat{\lambda}_C - \lambda)$, such that $Pr(\lambda \in \mathcal{I}) = 1 - \alpha$, and

$$\mathcal{I}^{\mathcal{N}} = \left(\hat{\lambda}_C - z_{1-\alpha}/(\hat{t}^C\sqrt{n}), \infty\right). \quad (5.7)$$

From Theorem 4, we define the Edgeworth-corrected $\alpha-$quantile of $\sqrt{n}\hat{t}_C(\hat{\lambda}_C - \lambda)$ as

$$z_\alpha - \frac{1}{\sqrt{n}} e_{CS}(z_\alpha) \quad (5.8)$$

and the corresponding refined interval

$$\mathcal{I}^{Ed} = \left(\hat{\lambda}_C - \frac{z_{1-\alpha} - n^{-1/2} e_{CS}(z_{1-\alpha})}{\sqrt{n}\hat{t}^C}, \infty\right). \quad (5.9)$$

From Theorem 4 we deduce

Corollary 2 *Suppose that model (1.1) and Assumptions 1- 7 hold. As* $n \to \infty$*,*

$$Pr\left(\lambda \in \mathcal{I}^{\mathcal{N}}\right) = Pr\left(\lambda \in \mathcal{I}\right) + O\left(\frac{1}{\sqrt{n}}\right) = 1 - \alpha + O\left(\frac{1}{\sqrt{n}}\right)$$

$$Pr\left(\lambda \in \mathcal{I}^{Ed}\right) = Pr\left(\lambda \in \mathcal{I}\right) + o\left(\frac{1}{\sqrt{n}}\right) = 1 - \alpha + o\left(\frac{1}{\sqrt{n}}\right). \quad (5.10)$$

The interval in (5.9) is infeasible, as the lower end-point depends on unknown λ, β and σ^2 through $e_{CS}(\cdot)$. We can construct a feasible version of $\mathcal{I}^{Ed}$ by replacing the unknowns with their respective estimates $\hat{\lambda}_C$, $\hat{\beta}_C$ and $\hat{\sigma}^2_C$, and thus $e_{CS}(\cdot)$ with its estimated version $\hat{e}_{CS}(\cdot)$. Since $\hat{\lambda}_C - \lambda = o_p(1)$, $\hat{\beta}_C - \beta = o_p(1)$ and $\hat{\sigma}^2_C - \sigma^2 = o_p(1)$, the result in (5.10) holds with $\mathcal{I}^{Ed}$ replaced by

$$\hat{\mathcal{I}}^{Ed} = \left(\hat{\lambda}_C - \frac{z_{1-\alpha} - n^{-1/2}\hat{e}_{CS}(z_{1-\alpha})}{\sqrt{n}\hat{t}^C}, \infty \right). \tag{5.11}$$

The practical performance of the corrected confidence sets is assessed by a Monte Carlo experiment reported in Section 7.

6 Discussion on the pure SAR case

A particular case of model (1.1) is the so-called pure SAR model

$$y = \lambda W y + \epsilon, \tag{6.1}$$

where $\beta = 0$ a priori. Estimation of λ in (6.1) is generally more problematic than that in (1.1) as the rate of convergence of standard estimators, such as MLE/QMLE, can be slower than the usual $\sqrt{n}$ depending on the choice of W (e.g. Lee (2004)). Moreover, the OLS estimator of λ in (6.1),

$$\hat{\lambda} = \frac{y'W'y}{y'W'Wy}, \tag{6.2}$$

is inconsistent unless $\lambda = 0$ even under (2.5).

Since the framework outlined in Sections 2-5 cannot be directly applied to (6.2) as the leading terms of (2.13) and (4.8) would be singular under Assumption 6, in this section we outline how results in Sections 2-5 can be adapted to accommodate the case $\beta = 0$.

Following our derivation in Section 2, we define

$$\psi_{P,n}(x) = \psi_P(x) = x + \frac{y'S(x)'S(x)y/n}{hy'W'Wy/n}\frac{h}{n}tr(G(x)). \tag{6.3}$$

For $s, t = 0, 1, ...$, let $g_{st,P} = htr(G^s G'^t)/n$, $a_P = g_{11,P} + g_{20,P} - 2g^2_{10,P}/h$, $t_P = g_{11,P}/a_P^{1/2}$,

$$b_P = g_{30,P} + 3g_{21,P} + \frac{8g^3_{10,P}}{h^2} - \frac{6}{h}g_{10,P}(g_{11,P} + g_{20,P}) \tag{6.4}$$

and

$$e_P(x) = \frac{2}{t_P a_P}\left(g_{21,P} - \frac{g_{11,P} g_{10,P}}{h}\right) x^2 - \frac{1}{3}\frac{b_P}{a_P^{3/2}}(x^2 - 1). \qquad (6.5)$$

In order to ensure that the leading term of the expansions in Theorems 5-7 are well defined, we replace Assumption 6 by

Assumption 6' $\underline{\lim}_{n\to\infty} (g_{11,P} + g_{20,P}) > 0$.

Similarly to Theorem 1, we derive

Theorem 5 *Suppose that model (6.1), Assumptions 1- 4 and 6' hold. For any real number ζ, the cdf of $\hat{\lambda}$ admits the second order formal Edgeworth expansion*

$$Pr\left(\left(\frac{n}{h}\right)^{1/2}\left(\hat{\lambda} - \psi_P(\lambda)\right) \le \zeta\right) = \Phi(t_P\zeta) + \left(\frac{h}{n}\right)^{1/2} e_P(t_P\zeta)\phi(t_P\zeta) + o\left(\left(\frac{h}{n}\right)^{1/2}\right), \qquad (6.6)$$

where $e_P(t_P\zeta) = O(1)$ as $n \to \infty$.

The proof of Theorem 5 is similar to that of Theorem 1 and is omitted here to avoid repetition. If $\lambda = 0$, the expansion in (6.6) corresponds to Theorem 1 in Robinson and Rossi (2015). Unlike (2.13), the expansion in (6.6) does not depend on σ^2 and can be used to derive improved tests on λ. Unlike Theorem 1 in Robinson and Rossi (2015), Theorem 5 can be used to improve tests of $H_0: \ \lambda = \lambda_0$ against $H_1: \ \lambda > \lambda_0 \ \ (\lambda < \lambda_0)$ with $\lambda_0 \neq 0$.

Similarly to Assumption 7, we introduce

Assumption 7' *For all sufficiently large n, $\psi_P(\lambda)$ defined in (6.3) is strictly increasing for all $\lambda \in \Lambda$ with probability one.*

Under Assumption 7, we can estimate λ in (6.1) by $\hat{\lambda}_{P,C} = \psi_P^{-1}(\hat{\lambda})$, and deduce

$$Pr\left(\left(\frac{n}{h}\right)^{1/2}(\hat{\lambda} - \psi_P(\lambda)) \le \zeta\right) = Pr\left(\left(\frac{n}{h}\right)^{1/2}\psi_P^{(1)}(\lambda)(\hat{\lambda}_{P,C} - \lambda) \le \zeta\right) + o(1), \qquad (6.7)$$

such that

$$\left(\frac{n}{h}\right)^{1/2}(\hat{\lambda}_{P,C} - \lambda) \xrightarrow{d} \mathcal{N}(0, v_P), \quad \text{with} \quad v_P = p\lim_{n\to\infty}(t_P\psi_P^{(1)}(\lambda))^{-2} = \lim_{n\to\infty} a_P^{-1}, \qquad (6.8)$$

which is equivalent to the asymptotic variance of the MLE of λ in (6.1) (e.g. Lee (2004), Theorem 5.2).

We also report (without proofs, to avoid repetition), results corresponding to Theorems 3 and 4 for model (6.1). For this purpose, we define

$$\bar{\iota}_P = \frac{2}{a_P^2}\left(\frac{g_{11,P}g_{10,P}}{h} - \frac{2g_{20,P}g_{10,P}}{h} + g_{30,P}\right) \tag{6.9}$$

and

$$b_{P,C} = \frac{g_{30,P}}{3} + g_{21,P} + \frac{8}{3}\frac{g_{10,P}^3}{h^2} - \frac{2g_{10,P}(g_{20,P}+g_{11,P})}{h} + \frac{\bar{\iota}_P}{2}a_P^2 \tag{6.10}$$

and

$$e_{P,C}(x) = -\frac{\bar{\iota}_P a_P^{1/2}}{2} + \frac{2}{a_P^{3/2}}\left(\frac{2g_{10,P}^3}{h^2} + g_{21,P} - \frac{2g_{11,P}g_{10,P}}{h} - \frac{g_{20,P}g_{10,P}}{h}\right)x^2 - \frac{b_{P,C}}{a_P^{3/2}}(x^2-1). \tag{6.11}$$

Theorem 6 *Suppose that model (6.1), Assumptions 1- 4, 6' and 7' hold. For any real number* ζ*, the cdf of* $\hat{\lambda}_{P,C}$ *admits the second order formal Edgeworth expansion*

$$Pr\left(\left(\frac{n}{h}\right)^{1/2}\left(\hat{\lambda}_{P,C}-\lambda\right)\le\zeta\right) = \Phi\left(a_P^{1/2}\zeta\right) + \left(\frac{h}{n}\right)^{1/2} e_{P,C}\left(a_P^{1/2}\zeta\right)\phi\left(a_P^{1/2}\zeta\right) + o\left(\left(\frac{h}{n}\right)^{1/2}\right), \tag{6.12}$$

with $e_{P,C}\left(a_P^{1/2}\zeta\right) = O(1)$ *as* $n\to\infty$.

Theorem 7 *Suppose that model (6.1), Assumptions 1- 4, 6' and 7' hold. For any real number* η*, the cdf of* $((n\hat{a}_P/h)^{1/2}(\hat{\lambda}_{P,C}-\lambda)$ *admits the second order formal Edgeworth expansion*

$$\begin{aligned} Pr\left(\left(\frac{n}{h}\right)^{1/2}\hat{a}_P^{1/2}\left(\hat{\lambda}_{P,C}-\lambda\right)\le\eta\right) &= \Phi(\eta) + \left(\frac{h}{n}\right)^{1/2} e_{P,C}(\eta)\phi(\eta) \\ + \left(\frac{h}{n}\right)^{1/2}\frac{1}{a_P^{3/2}}\left(g_{21,P}+g_{30,P}-\frac{2g_{20,P}g_{10,P}}{h}\right)&\eta^2\phi(\eta) + o\left(\left(\frac{h}{n}\right)^{1/2}\right), \end{aligned} \tag{6.13}$$

with $e_{P,C}(\eta) = O(1)$ and $g_{21,P}+g_{30,P}-2g_{20,P}g_{10,P}/h = O(1)$ *as* $n\to\infty$.

Similarly to the discussion in Section 6, Theorem 7 can be used to derive improved confidence intervals for λ in (6.1).

7 Monte Carlo results

In this section, we report a small Monte Carlo exercise to assess the finite sample behaviour of our corrected tests. The number of exogenous regressors is set at $k=3$, with X_1 being a $n\times 1$ column of ones, while for each $i=1,...n$ $X_{2i}\sim U[0,1]$ and $X_{3i}\sim U[0,1]$, $U[a,b]$ denoting a uniform distribution on support $[a, b]$. Regressors are generated once for each scenario and kept fixed across 1000 Monte Carlo replications. We set $\beta = (0.3, 0.5, -0.5)$, $n = 30, 50, 100, 200$ and generate ϵ according to Assumption 1 with $\sigma^2 = 1$. We choose two different specifications for W, both of which are symmetric and satisfy Assumptions 3 and 4

a) W is based on an exponential distance criterion, with initial weights equal to $exp(-|\ell_i - \ell_j|)1(|\ell_i - \ell_j| < \log n)$, where ℓ_i is the $i-$th location along the interval $[0, n]$ which is randomly generated from $U[0, n]$. The resulting matrix is then rescaled to produce a W with elements in each row summing to one. Such W is empirically motivated as it mimics a distance-based weight matrix constructed from real data. W is generated once for each sample size and is kept fixed across replications and across the different scenarios.

b) W is generated as a circulant structure with two neighbours ahead and two behind, normalised so that elements in each row sum to one. More specifically,

$$W = \frac{1}{4} \begin{pmatrix} 0 & 1 & 1 & 0 & .... & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & .... & 0 & 1 \\ 1 & 1 & 0 & .... & .... & .... & 1 & 1 \\ ... & ... & ... & .... & .... & ... & ... & .... \\ 1 & 1 & ... & .... & ... & 1 & 1 & 0 \end{pmatrix}. \tag{7.1}$$

Across tables, “standard”, “corrected” and “transformed” denote respectively empirical size/power of tests (3.5), (3.6) and (3.10). For comparison purpose, we also report size and power of the t-test of (3.1) against (3.2) based on the maximum likelihood estimator $\hat{\lambda}_{MLE}$ of λ (Lee (2004)), and specifically of a test that rejects H_0 in (3.1) when

$$\sqrt{n} \left(\frac{\hat{a}_{MLE}}{\hat{\sigma}^2_{MLE}} \right)^{1/2} \hat{\lambda}_{MLE} > z_{1-\alpha}, \tag{7.2}$$

where $\hat{\sigma}^2_{MLE}$ and $\hat{a}_{MLE}$ respectively denote the MLE of σ^2 under H_0 and the estimated version of a in (2.9) with the unknown σ^2 and β replaced by their MLE under H_0. In the tables we indicate by “MLE” the empirical size/power obtained from (7.2).

We also report size and power derived by a parametric bootstrap algorithm, denoted by “bootstrap” in the tables. More specifically, under H_0, we estimate β in (1.1) by OLS and obtain the “restricted” residuals $\hat{\epsilon} = y - X\hat{\beta}$. We then generate $B = 999$ $n-$dimensional vectors $\hat{\epsilon}^*_j$, $j = 1, ...B$, where each component of ϵ_j^* is extracted from $N(0, \hat{\epsilon}'\hat{\epsilon}/n)$, and hence B pseudo-data under H_0, $y_j^* = X\hat{\beta} + \hat{\epsilon}_j^*$, $j = 1, ...B$. For every $j = 1, ...B$, we compute OLS bootstrap estimates λ_j^* and restricted bootstrap estimates of β and σ^2 as $\beta_j^* = (X'X)^{-1}X'y_j^*$ and $\sigma_j^2 = y_j^{*'}Py_j^*/n$. We then obtain B bootstrap OLS null statistics as $\sqrt{n}t_j^*\lambda_j^*$, where t_j^* is obtained from (2.10) by replacing unknowns by β_j^* and σ_j^* for each $j = 1, ...B$, and compute the $1-\alpha$ bootstrap quantile $w_{1-\alpha}^*$ as the solution of

$$\frac{1}{B}\sum_{j=1}^{B} 1\left(\sqrt{n}t_j^*\lambda_j^* \leq w_{1-\alpha}^*\right) \leq 1-\alpha$$

under H_0 in (3.1). The size of the test of (3.1) based on $w_{1-\alpha}^*$ is obtained as

$$Pr\left(\sqrt{n}\hat{t}\hat{\lambda} > w_{1-\alpha}^*\right). \tag{7.3}$$

[Tables 1-2 about here]

In Tables 1 and 2 we report empirical sizes of the various tests of (3.1) against (3.2), where nominal size is $\alpha = 0.05$, and W is chosen as both a) and b) above. In both tables, sizes for "normal" and "MLE" are substantially lower than the nominal 0.05, while those obtained by "corrected" and "transformed" seem significantly better. Specifically, sizes for "corrected" in both Tables 1 and 2 are slightly higher than 0.05 for the smallest samples, but results improve as n increases. Sizes for "transformed", instead, are very close to the nominal for all sample sizes. Interestingly, for both scenarios, "transformed" outperforms "bootstrap" for all sample sizes, while for larger n "corrected" returns values that are closer to the nominal 0.05 than "bootstrap" . This pattern is roughly preserved even if we implement the bootstrap algorithm by resampling from restricted residuals, rather than using the parametric version described above.

[Tables 3-4 about here]

In Tables 3 and 4 we report power of the various tests against a fixed one-sided alternative (3.2),

$$H_1: \quad \lambda = \bar{\lambda} > 0, \tag{7.4}$$

with $\bar{\lambda} = 0.2, 0.5, 0.8$ where, again, $\alpha = 0.05$. In Table 3 we report results for scenario a), while Table 4 displays empirical powers for scenario b). In both Tables and for all $\bar{\lambda}$, "corrected" and "transformed" significantly outperform "standard" and "MLE" for all sample sizes. Also, in both scenarios, tests based on "corrected" and "transformed" appear to be more powerful than "bootstrap". Comparing results in Tables 3 and 4, we detect a very similar pattern for the performance of various tests.

[Figures 1-2 about here]

In Figures 1 and 2 we report plots of $\psi(\cdot)$ in (2.7) for a few data generating processes, when $n = 100$. In particular, Figure 1 reports $\psi(x)$ for $-1 < x < 1$, when y is generated as in (1.1) with $\lambda = 0.2, 0.5, 0.8$ and W is a) above. Figure 2 reports $\psi(x)$ over the same support, when y is generated according to the same values of λ, but with W as in b). In all scenarios, β and X are generated as in the Monte Carlo experiment described at the beginning of this section. From the plots of all curves for both scenarios, we notice that $\psi(x)$ is strictly increasing over the support. There is evidence of a discontinuity at $x = 1$, but $\psi(\cdot)$ is well-behaved in the interior of the support implied by Assumption 2.

We also assess the practical performance of results in Section 5 by comparing the lower-end-points (LEPs, henceforth) of intervals (5.7) and (5.11) averaged across Monte Carlo replications, and their empirical coverage probabilities. We also report corresponding quantities for the confidence intervals

$$\mathcal{I}^{MLE} = \left(\hat{\lambda}_{MLE} - z_{1-\alpha} \left(\frac{\hat{\sigma}^2_{MLE}}{n\hat{a}_{MLE}} \right)^{1/2}, \infty \right) \tag{7.5}$$

and

$$\mathcal{I}^{B} = \left(\hat{\lambda}_C - \frac{w^{C*}_{1-\alpha}}{\sqrt{n}\hat{t}_C}, \infty \right), \tag{7.6}$$

where $w^{C*}_{1-\alpha}$ is obtained by a bootstrap algorithm in which the pseudo-sample is generated as $y^*_j = S^{-1}(\hat{\lambda}_C)(X\hat{\beta_C} + \hat{\epsilon}^*_j)$, $j = 1, ...B$, with $B = 999$. Again, each component of the $n-$dimensional vectors $\hat{\epsilon}^*_j$, $j = 1, ...B$, is extracted from $N(0, \hat{\epsilon}'\hat{\epsilon}/n)$, with $\hat{\epsilon} = y - \hat{\lambda}_C Wy - \hat{\beta}_C X$. For every $j = 1, ...B$, we compute corrected-OLS bootstrap estimates λ^*_{Cj}, β^*_{Cj} and σ^{2*}_{Cj} . We then obtain B bootstrap statistics as $\sqrt{n}t^*_{Cj}(\lambda^*_{Cj} - \hat{\lambda}_C)$, where t^*_{Cj} is obtained from (2.10) and replacing unknowns by λ^*_{Cj}, β^*_{Cj} and σ^{2*}_{Cj} for each $j = 1, ...B$, and compute the $1 - \alpha$ bootstrap quantile $w^{C*}_{1-\alpha}$ as the solution of

$$\frac{1}{B}\sum_{j=1}^{B} 1 \left(\sqrt{n}t^*_{Cj}(\lambda^*_{Cj} - \hat{\lambda}_C) \leq w^{C*}_{1-\alpha} \right) \leq 1 - \alpha.$$

Tables 5 and 6 reports average LEPs and empirical coverage probabilities (in brackets) for $\mathcal{I}^{\mathcal{N}}$, $\hat{\mathcal{I}}^{Ed}$, $\mathcal{I}^{MLE}$ and $\mathcal{I}^{B}$ (respectively, denoted as "normal", "corrected", "MLE" and "bootstrap" in the Tables) with $\alpha = 0.05$ for $\lambda = 0.3, 0.5, 0.7$.

[Tables 5-6 about here]

Results in both Tables 5 an 6 reveal that, although not identical, LEPs and coverage probabilities of "normal" and "MLE" are often the same up to the fourth decimal place, confirming that the asymptotic equivalence of $\hat{\lambda}_C$ and $\hat{\lambda}_{MLE}$ reported after (4.2) is preserved reasonably well in small samples. Across all scenarios, as expected LEP becomes a sharper lower bound for λ as n increases, although the empirical coverage probabilities are often higher than the nominal 0.95. In all cases, "corrected" produce LEPs that are closer to the true value of λ compared to that of either "normal" or "MLE", with associated coverage probabilities that are generally closer to 0.95 compared to "normal" (with few exceptions, e.g. for $\lambda = 0.7$ and small n, in scenario b)). The comparison of "corrected" and "bootstrap", instead, reveal that "bootstrap" generally offers sharper LEPs than "corrected" for $\lambda = 0.3, 0.5$, while the opposite holds for $\lambda = 0.7$. In terms of coverage probabilities, the performance of "corrected" is comparable to, and often superior to "bootstrap". Across all scenarios and for all n, Tables 5 and 6 show that "corrected" and "bootstrap" outperform "normal" and "MLE" in terms of providing sharper bounds for λ and in terms of coverage probabilities.

8 Concluding remarks

In this paper we revisited standard OLS estimation for the spatial parameter of a standard SAR model with or without exogenous regressors, and derived formal higher-order expansions that can be used to develop improved inference under general network structures. In particular, we suggest improved tests starting from the Edgeworth expansion of the cdf of a suitably centred OLS estimate of λ, and construct refined confidence sets from the higher-order expansion of a transformed OLS estimate. The transformed estimate of λ is consistent and it is asymptotically first-order equivalent to the MLE under normality of the error term. A small Monte Carlo study shows that our new improved tests and confidence sets enjoy a very satisfactory finite sample performance, which is comparable (or sometimes superior) to a suitable bootstrap algorithm.

Appendix

Proof of Theorem 1

We have

$$M^{-1} = \begin{bmatrix} m^{11} & m^{12} \\ m^{12\prime} & m^{22} \end{bmatrix}, \tag{A.1}$$

where $m^{11} = (m_{11} - m_{12}m_{22}^{-1}m_{21})^{-1}$, $m^{12} = m^{21\prime} = -m^{11}m_{12}m_{22}^{-1}$, $m^{22} = m_{22}^{-1}(I + m_{21}m^{11}m_{12}m_{22}^{-1})$, and m_{11}, $m_{12} = m_{21}'$ and m_{22} are defined in (2.3).

We proceed as in Phillips (1977). From (2.2) and (2.7),

$$\begin{aligned} \hat{\lambda} - \psi(\lambda) &= m^{11}u_1 + m^{12}u_2 - g_{10}m^{11}\epsilon' P\epsilon = m^{11}(u_1 - m_{12}m_{22}^{-1}u_2) - g_{10}m^{11}\epsilon' P\epsilon \\ &= ((X\beta + \epsilon)' G' P G (X\beta + \epsilon))^{-1} \left((X\beta + \epsilon)' G' P\epsilon - g_{10}\epsilon' P\epsilon\right), \end{aligned} \tag{A.2}$$

since, from (2.3) and after substituting (2.1), we obtain

$$u_1 - m_{12}m_{22}^{-1}u_2 = (X\beta + \epsilon)' G'\epsilon - (X\beta + \epsilon)' G' X (X'X)^{-1} X'\epsilon = (X\beta + \epsilon)' G' P\epsilon \tag{A.3}$$

and

$$m^{11} = ((X\beta + \epsilon)' G' P G (X\beta + \epsilon))^{-1}. \tag{A.4}$$

For any real number ζ,

$$\Pr\left(\hat{\lambda} - \psi(\lambda) \leq \frac{\zeta}{\sqrt{n}}\right) = \Pr\left(\frac{1}{2}\epsilon'(C + C')\epsilon + c'\epsilon + d \leq 0\right), \tag{A.5}$$

with

$$C = G'P - Pg_{10} - \frac{\zeta}{\sqrt{n}} G' P G, \quad c' = \beta' X' G' P \left(I - \frac{2\zeta}{\sqrt{n}} G\right), \quad d = -\sqrt{n}\zeta\delta(I). \tag{A.6}$$

For ease of notation, let $f = \epsilon'(C + C')\epsilon/2 + c'\epsilon + d$.

Under Assumption 1, we can derive the characteristic function of f as

$$\begin{aligned} E(e^{itf}) =& E(e^{it(\frac{1}{2}\epsilon'(C+C')\epsilon + c'\epsilon + d)}) = \frac{e^{itd}}{(2\pi)^{n/2}\sigma^n} \int_{\Re^n} e^{it\frac{1}{2}\xi'(C+C')\xi + itc'\xi} e^{-\frac{\xi'\xi}{2\sigma^2}} d\xi \\ =& \frac{e^{itd}}{(2\pi)^{n/2}\sigma^n} \int_{\Re^n} e^{-\frac{1}{2}(\frac{\xi}{\sigma} - q)'(I - it\sigma^2(C+C'))(\frac{\xi}{\sigma} - q)} e^{\frac{1}{2}q'(I - it\sigma^2 C)q} d\xi, \end{aligned} \tag{A.7}$$

where q satisfies $itc' = q'(I - it\sigma^2(C + C'))/\sigma$. By standard algebra, $q = (I - it\sigma^2(C + C'))^{-1} itc\sigma$, and hence $q'(I - it\sigma^2(C+C'))q = -t^2c'(I - it\sigma^2(C+C'))^{-1}c\sigma^2$. By Gaussian integration, (A.7) becomes

$$\begin{aligned} E(e^{itf}) =& e^{itd - \frac{\sigma^2}{2}t^2c'(I-it\sigma^2(C+C'))^{-1}c} det(I - it\sigma^2(C+C'))^{-1/2} \\ =& e^{itd - \frac{\sigma^2}{2}t^2c'(I-it\sigma^2(C+C'))^{-1}c} \prod_{j=1}^{n}(1 - it\sigma^2\eta_j(C+C'))^{-1/2}. \end{aligned} \tag{A.8}$$

From (A.8), the cumulant generating function of f is

$$\begin{aligned} \ln\left(E(e^{itf})\right) =& itd - \frac{\sigma^2}{2}t^2c'(I - it\sigma^2(C+C'))^{-1}c - \frac{1}{2}\sum_{j=1}^{n}\ln(1 - it\sigma^2\eta_j(C+C')) \\ =& itd - \frac{\sigma^2}{2}t^2c'\sum_{s=0}^{\infty}(it\sigma^2(C+C'))^s c + \frac{1}{2}\sum_{s=1}^{\infty}\frac{1}{s}(it\sigma^2)^s tr((C+C')^s), \end{aligned} \tag{A.9}$$

where the last displayed equality follows since

$$(I - it\sigma^2(C+C'))^{-1} = \sum_{s=0}^{\infty}(it\sigma^2(C+C'))^s, \quad \ln(1 - it\sigma^2\eta_j(C+C')) = -\sum_{s=1}^{\infty}\frac{(it\sigma^2\eta_j(C+C'))^s}{s} \tag{A.10}$$

and hence

$$\begin{aligned} -\frac{1}{2}\sum_{j=1}^{n}ln(1 - it\sigma^2\eta_j(C+C')) =& \frac{1}{2}\sum_{j=1}^{n}\sum_{s=1}^{\infty}\frac{1}{s}(it\sigma^2)^s\eta_j(C+C')^s = \frac{1}{2}\sum_{s=1}^{\infty}\frac{1}{s}(it\sigma^2)^s\sum_{j=1}^{n}\eta_j(C+C')^s \\ =& \frac{1}{2}\sum_{s=1}^{\infty}\frac{1}{s}(it\sigma^2)^s tr((C+C')^s). \end{aligned} \tag{A.11}$$

With κ_s denoting the $s-$th cumulant of f, (A.9) gives

$$\kappa_1 = d + \sigma^2 trC, \quad \kappa_2 = \sigma^2(c'c + \frac{\sigma^2}{2}tr((C+C')^2)), \tag{A.12}$$

and

$$\kappa_s = \frac{\sigma^{2s}s!}{2}\left(\frac{1}{\sigma^2}c'(C+C')^{s-2}c + \frac{tr((C+C')^s)}{s}\right), s > 2. \tag{A.13}$$

Let $f^c = (f - \kappa_1)/\kappa_2^{1/2}$ and $\kappa_s^c = \kappa_s/\kappa_2^{s/2}$, so that

$$\ln\left(E(e^{itf^c})\right) = -\frac{1}{2}t^2 + \sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!} \tag{A.14}$$

and thus

$$
\begin{aligned}
E(e^{itf^c}) &= e^{-\frac{1}{2}t^2}\exp\left(\sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!}\right)\\
&=e^{-\frac{1}{2}t^2}\left(1+\sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!}+\frac{1}{2!}\left(\sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!}\right)^2+\frac{1}{3!}\left(\sum_{s=3}^{\infty}\frac{\kappa_s^c(it)^s}{s!}\right)^3+.....\right)\\
&= e^{-\frac{1}{2}t^2}\left(1+\frac{\kappa_3^c(it)^3}{3!}+\frac{\kappa_4^c(it)^4}{4!}+\frac{\kappa_5^c(it)^5}{5!}+\left(\frac{\kappa_6^c}{6!}+\frac{(\kappa_3^c)^2}{(3!)^2}\right)(it)^6+.....\right). \qquad \text{(A.15)}
\end{aligned}
$$

By Fourier inversion, formally,

$$
Pr(\sqrt{n}\left(\hat{\lambda}-\psi(\lambda)\right)\leq\zeta)=Pr(f^c\leq-\kappa_1^c)=\Phi(-\kappa_1^c)-\frac{\kappa_3^c}{3!}\Phi^{(3)}(-\kappa_1^c)+\frac{\kappa_4^c}{4!}\Phi^{(4)}(-\kappa_1^c)+.... \qquad \text{(A.16)}
$$

Now,

$$
\kappa_1^c=\frac{d+\sigma^2tr(C)}{\sigma\left(c'c+\frac{\sigma^2}{2}tr((C+C')^2)\right)^{1/2}}. \qquad \text{(A.17)}
$$

By standard algebra, $tr(C)=-\sqrt{n}g_{11}\zeta-tr(G'X(X'X)^{-1}X')+g_{10}k+O\left(1/\sqrt{n}\right)$, and hence

$$
d+\sigma^2tr(C)=-\sqrt{n}\left(g_{11}\sigma^2+\delta(I)\right)\zeta-\sigma^2tr(G'X(X'X)^{-1}X')+\sigma^2kg_{10}+O\left(\frac{1}{\sqrt{n}}\right). \qquad \text{(A.18)}
$$

Also, $\sigma^2tr((C+C')^2)/2=\sigma^2n(g_{20}+g_{11}-2g_{10}^2)+4\sigma^2\sqrt{n}(g_{10}g_{11}-g_{21})\zeta+o(\sqrt{n})$ and $c'c=n\delta(I)-4\sqrt{n}\delta(G)\zeta+o(\sqrt{n})$, so that

$$
\begin{aligned}
\left(c'c+\frac{\sigma^2}{2}tr((C+C')^2)\right)^{1/2} &=n^{1/2}a^{1/2}\left(1+\frac{4(\sigma^2g_{10}g_{11}-\sigma^2g_{21}-\delta(G))\zeta}{\sqrt{n}a}+o\left(\frac{1}{\sqrt{n}}\right)\right)^{1/2}\\
&=n^{1/2}a^{1/2}\left(1+\frac{2(\sigma^2g_{10}g_{11}-\sigma^2g_{21}-\delta(G))\zeta}{\sqrt{n}a}\right)+o\left(\frac{1}{\sqrt{n}}\right), \qquad \text{(A.19)}
\end{aligned}
$$

with a defined in (2.9). Combining (A.18) and (A.19),

$$
\begin{aligned}
\kappa_1^c=&-\frac{g_{11}\sigma^2+\delta(I)}{\sigma a^{1/2}}\zeta-\frac{1}{\sqrt{n}}\frac{\sigma(tr(G'X(X'X)^{-1}X')-kg_{10})}{a^{1/2}}\\
&+\frac{1}{\sqrt{n}}\frac{2(g_{11}\sigma^2+\delta(I))(\sigma^2g_{11}g_{10}-\delta(G)-\sigma^2g_{21})\zeta^2}{\sigma a^{3/2}}+o\left(\frac{1}{\sqrt{n}}\right). \qquad \text{(A.20)}
\end{aligned}
$$

Similarly, $c'(C+C')c=2n\delta(G)-2ng_{10}\delta(I)+o(n)$ and $tr((C+C')^3)/3=2n\left(g_{30}+3g_{21}\right)/3+$

$5ng_{10}^3/3 - 2ng_{10}(g_{20}+g_{11}) + o(n)$, and hence

$$\kappa_3^c = \frac{3\sigma\left(c'(C+C')c + \frac{\sigma^2}{3}tr((C+C')^3)\right)}{\left(c'c + \frac{\sigma^2}{2}tr((C+C')^2)\right)^{3/2}} = \frac{6\sigma}{\sqrt{n}}\frac{b}{a^{3/2}} + o\left(\frac{1}{\sqrt{n}}\right), \tag{A.21}$$

where b is defined according to (2.11).

Thus, by Taylor expansion of (A.20), (A.16) becomes

$$Pr(\sqrt{n}\left(\hat{\lambda} - \psi(\lambda)\right) \leq \zeta) = \Phi(t\zeta) + \frac{1}{\sqrt{n}}e_1(t\zeta)\phi(t\zeta) - \frac{1}{\sqrt{n}}\frac{\sigma b}{a^{3/2}}\Phi^{(3)}(t\zeta) + o\left(\frac{1}{\sqrt{n}}\right), \tag{A.22}$$

where $e_1(x) = \sigma\left(tr(G'X(X'X)^{-1}X') - kg_{10}\right)/a^{1/2} + 2(\delta(G) + \sigma^2 g_{21} - \sigma^2 g_{11}g_{10})x^2/ta$ and t defined as in (2.10).

The claim in (2.13) of Theorem 1 holds using $\Phi^{(3)}(x) = (x^2-1)\phi(x)$ and by letting $e(x) = e_1(x) - \sigma b(x^2-1)/a^{3/2}$.

Proof of Theorem 2

By the mean value theorem,

$$\begin{aligned}\sqrt{n}\hat{t}\hat{\lambda} =& \sqrt{n}t\hat{\lambda} + \sqrt{n}\hat{\lambda}\frac{\partial t}{\partial \beta'}|_{\beta,\sigma^2}(\hat{\beta}-\beta) + \sqrt{n}\hat{\lambda}\frac{\partial t}{\partial \sigma^2}|_{\beta,\sigma^2}(\hat{\sigma}^2-\sigma^2)\\ &+\frac{1}{2}\sqrt{n}\hat{\lambda}\begin{pmatrix}\hat{\beta}-\beta\\ \hat{\sigma}^2-\sigma^2\end{pmatrix}'\bar{H}\begin{pmatrix}\hat{\beta}-\beta\\ \hat{\sigma}^2-\sigma^2\end{pmatrix},\end{aligned} \tag{A.23}$$

where $\bar{H}$ is the $(k+1)\times(k+1)$ matrix of second derivatives, i.e.

$$H = \begin{pmatrix}\frac{\partial^2 t}{\partial\beta\partial\beta'} & \frac{\partial^2 t}{\partial\beta\partial\sigma^2}\\ \frac{\partial^2 t}{\partial\sigma^2\partial\beta'} & \frac{\partial^2 t}{\partial(\sigma^2)^2}\end{pmatrix}, \tag{A.24}$$

evaluated at $\bar{\beta}$ and $\bar{\sigma}^2$ such that $||\bar{\beta}-\beta|| < ||\hat{\beta}-\beta||$ and $|\bar{\sigma}^2-\sigma^2| < |\hat{\sigma}^2-\sigma^2|$, respectively, and

$$\hat{\beta}-\beta = (X'X)^{-1}X'\epsilon, \quad \hat{\sigma^2}-\sigma^2 = \frac{1}{n}y'Py - \sigma^2 = \frac{1}{n}\epsilon'\epsilon - \sigma^2 + o_p\left(\frac{1}{\sqrt{n}}\right) \tag{A.25}$$

under H_0 in (3.1).

Let

$$v = \frac{\delta(I) + \sigma^2(2g_{20} + g_{11})}{(\sigma^2 a)^{3/2}}. \tag{A.26}$$

By standard algebra we derive

$$\frac{\partial t}{\partial \beta'} = \frac{\sigma^2}{n} v\beta' X'W'PWX, \quad \frac{\partial t}{\partial \sigma^2} = -\frac{\delta(I)}{2} v \tag{A.27}$$

and

$$\begin{aligned}\frac{\partial^2 t}{\partial \beta \partial \beta'} &= \frac{\sigma^2}{n} vX'W'PWX + \frac{\sigma^2}{n} X'W'PWX\beta \frac{\partial v}{\partial \beta'}, \\ \frac{\partial^2 t}{\partial (\sigma^2)^2} &= -\frac{\delta(I)}{2} \frac{\partial v}{\partial \sigma^2}, \qquad \frac{\partial^2 t}{\partial \beta \partial \sigma^2} = -\frac{v}{n} X'W'PWX\beta - \frac{\delta(I)}{2} \frac{\partial v}{\partial \beta},\end{aligned} \tag{A.28}$$

with

$$\frac{\partial v}{\partial \beta} = -\frac{1}{na}\left(\frac{2\sigma^2 g_{20}}{(\sigma^2 a)^{3/2}} + v\right) X'W'PWX\beta, \quad \frac{\partial v}{\partial \sigma^2} = -\frac{3v(g_{20} + g_{11})}{2a} - \frac{a(3\delta(I) + \sigma^2(2g_{20} + g_{11}))}{2(a\sigma^2)^{5/2}}. \tag{A.29}$$

Under H_0 in (3.1), $\hat{\beta} - \beta = O_p(1/\sqrt{n})$, $\hat{\sigma}^2 - \sigma^2 = O_p(1/\sqrt{n})$ and $\hat{\lambda} = O_p(1/\sqrt{n})$, so that the last term in (A.23) is $O_p(1/n)$ as long as each element of $\bar{H}$ is $O_p(1)$. Standard calculations under Assumptions 3-5 show that each element of $X'W'PWX/n$, as well as the numerators of v, $\partial v/\partial \sigma^2$ and $\partial v/\partial \beta$, are $O(1)$ as $n \to \infty$ under H_0 in (3.1). Each element of $\bar{H}$ is $O_p(1)$ so long as $\bar{a} = \bar{\beta}' X'W'PWX\bar{\beta}/n + \bar{\sigma}^2(g_{20} + g_{11}) > 0$ as $n \to \infty$. Under H_0, with probability approaching 1 as $n \to \infty$, $||\bar{\beta} - \beta|| < ||\hat{\beta} - \beta|| < \varepsilon$ and $|\bar{\sigma}^2 - \sigma^2| < |\hat{\sigma}^2 - \sigma^2| < \varepsilon$, for any $\varepsilon > 0$. We can write

$$\bar{a} \geq a - |\bar{a} - a|, \tag{A.30}$$

and under Assumption 6, as $n \to \infty$

$$a = \delta(I) + \frac{\sigma^2}{2n} tr(W + W')^2 \geq \delta(I) > 0. \tag{A.31}$$

Also,

$$\bar{a} - a = \frac{1}{n}\left(\bar{\beta}'X'W'PWX\bar{\beta} - \beta'X'W'PWX\beta\right) + (g_{20} + g_{11})(\bar{\sigma}^2 - \sigma^2)$$
$$= \frac{1}{n}\beta'X'W'PX(\bar{\bar{\beta}} - \beta) + (g_{20} + g_{11})(\bar{\sigma}^2 - \sigma^2) + O_p(||\bar{\bar{\beta}} - \beta||^2) = O_p(\varepsilon) = o_p(1), \quad \text{(A.32)}$$

with $||\bar{\bar{\beta}} - \beta|| < ||\bar{\beta} - \beta|| < \varepsilon$ and where the last equality in (A.32) holds from arbitrariness of ε. Hence, combining (A.30)-(A.32), $\bar{a} > 0$ as $n \to \infty$, and (A.23) becomes

$$\sqrt{n}\hat{t}\hat{\lambda} = \sqrt{n}t\hat{\lambda} + \sqrt{n}\hat{\lambda}\frac{\partial t}{\partial \beta'}|_{\beta,\sigma^2}(\hat{\beta} - \beta) + \sqrt{n}\hat{\lambda}\frac{\partial t}{\partial \sigma^2}|_{\beta,\sigma^2}(\hat{\sigma}^2 - \sigma^2) + O_p\left(\frac{1}{n}\right). \quad \text{(A.33)}$$

By substituting (A.25) and (A.27) into (A.33),

$$\sqrt{n}\hat{t}\hat{\lambda} = \sqrt{n}t\hat{\lambda} + \sqrt{n}\hat{\lambda}\frac{\sigma^2 v}{n}\beta'X'W'PWX(X'X)^{-1}X'\epsilon - \sqrt{n}\hat{\lambda}\frac{\delta(I)v}{2}\left(\frac{1}{n}\epsilon'\epsilon - \sigma^2\right) + o_p\left(\frac{1}{\sqrt{n}}\right), \quad \text{(A.34)}$$

and therefore, for any real η,

$$P\left(\sqrt{n}\hat{t}\hat{\lambda} \le \eta\right) = P\left(\sqrt{n}t\hat{\lambda} + \sqrt{n}\hat{\lambda}\frac{\sigma^2 v}{n}\beta'X'W'PWX(X'X)^{-1}X'\epsilon - \sqrt{n}\hat{\lambda}\frac{\delta(I)v}{2}\left(\frac{1}{n}\epsilon'\epsilon - \sigma^2\right)\right.$$
$$\left. + o_p\left(\frac{1}{\sqrt{n}}\right) \le \eta\right)$$
$$= P\left(\sqrt{n}t\hat{\lambda} + \sqrt{n}\hat{\lambda}\frac{\sigma^2 v}{n}\beta'X'W'PWX(X'X)^{-1}X'\epsilon - \sqrt{n}\hat{\lambda}\frac{\delta(I)v}{2}\left(\frac{1}{n}\epsilon'\epsilon - \sigma^2\right) \le \eta\right)$$
$$+ o\left(\frac{1}{\sqrt{n}}\right).$$

So, by substituting $\hat{\lambda} = m^{11}(X\beta + \epsilon)'W'P\epsilon$, where m^{11} is defined in (A.4), into the last displayed expression and rearranging,

$$P\left(\sqrt{n}\hat{t}\hat{\lambda} \le \eta\right) = P\left(\xi + \frac{1}{\sqrt{n}}\rho \le 0\right) + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{(A.35)}$$

with $\xi = \epsilon' C_s \epsilon + c_s'\epsilon + d_s$, where

$$C_s = \frac{t}{2\sqrt{n}}(WP + PW') - \frac{\eta}{n}W'PW, \quad c_s' = \frac{t}{\sqrt{n}}\beta'X'W'P - \frac{2\eta}{n}\beta'X'W'PW, \quad d_s = -\eta\delta(I) \quad \text{(A.36)}$$

and $\rho = \rho_1 + \rho_2 + \rho_3 + \rho_4$, with

$$\rho_1 = \frac{\sigma^2 v}{2n}\epsilon'(W'P + PW)\epsilon\beta' X' W' PW(I-P)\epsilon, \tag{A.37}$$

$$\rho_2 = \frac{\sigma^2 v}{2n}\epsilon'\left((I-P)W'PWX\beta\beta'X'W'P + PWX\beta\beta'X'W'PW(I-P)\right)\epsilon, \tag{A.38}$$

$$\rho_3 = -\frac{v\delta(I)}{4n}\epsilon'(W'P + PW)\epsilon\left(\epsilon'\epsilon - n\sigma^2\right), \quad \rho_4 = -\frac{v\delta(I)}{2n}\beta'X'W'P\epsilon\left(\epsilon'\epsilon - n\sigma^2\right). \tag{A.39}$$

By standard algebra involving expectations of quadratic forms in normal random variables we can show that $\xi = O_p(1)$ and $\rho = O_p(1)$ as $n \to \infty$.

We approximate the characteristic function of $\xi + \frac{1}{\sqrt{n}}\rho$ by $1 + \chi$, where

$$\chi = itE(\xi + \frac{1}{\sqrt{n}}\rho) + \frac{1}{2}(it)^2 E((\xi + \frac{1}{\sqrt{n}}\rho)^2) + \frac{1}{6}(it)^3 E((\xi + \frac{1}{\sqrt{n}}\rho)^3), \tag{A.40}$$

and deduce its approximate cumulant generating function

$$\log(1+\chi) = \sum_{s=1}^{\infty}(-1)^{s+1}\frac{\chi^s}{s}. \tag{A.41}$$

Using the notation adopted in the proof of Theorem 1, we indicate by κ_s the sth cumulant of $\xi + \frac{1}{\sqrt{n}}\rho$ and by $\kappa_s^c = \kappa_s/\kappa_2^{s/2}$ its scaled version. Using standard formulae for moments of linear and quadratic forms in normal random variables (details of full derivation can be obtained from the authors upon request), we deduce $\kappa_1 = -\eta(\sigma^2 g_{11} + \delta(I)) - \sigma^2 tr(W'X(X'X)^{-1}X')t/\sqrt{n} + o\left(1/\sqrt{n}\right)$, and $\kappa_2 = \sigma^2 t^2 a - 4\sigma^2 t(\delta(W) + \sigma^2 g_{21})\eta/\sqrt{n} + o\left(1/\sqrt{n}\right)$, such that

$$\kappa_1^c = -\eta - \frac{1}{\sqrt{n}}\left(\frac{2}{at}(\delta(W) + \sigma^2 g_{21})\eta^2 + \frac{\sigma}{a^{1/2}}tr(W'X(X'X)^{-1}X')\right) + o\left(\frac{1}{\sqrt{n}}\right). \tag{A.42}$$

Also, from (A.40) and (A.41), we deduce

$$\kappa_3 = 8\sigma^6 tr(C_s^3) + 6\sigma^4 c_s' C_s c_s + o\left(\frac{1}{\sqrt{n}}\right) = \frac{6}{\sqrt{n}}t^3\sigma^4\left(\delta(W) + \sigma^2\left(\frac{g_{30}}{3} + g_{21}\right) + o\left(\frac{1}{\sqrt{n}}\right)\right) \tag{A.43}$$

and $\kappa_3^c = 6\sigma b/\sqrt{n}a^{3/2} + o\left(1/\sqrt{n}\right)$. The claim in (3.4) follows by setting $e(\eta)$ as in (2.12).

Proof of Theorem 3

By standard algebra, using (2.1), $\psi^{(1)}(\lambda) = 1 - m^{11}\left(2g_{10}\epsilon' G' P\epsilon + 2g_{10}\beta' X' G' P\epsilon - g_{20}\epsilon' P\epsilon\right)$ and $\psi^{(2)}(\lambda) = 2m^{11}\left((m^{11})^{-1}g_{10} - 2g_{20}\epsilon' G' P\epsilon - 2g_{20}\beta' X' G' P\epsilon + g_{30}\epsilon' P\epsilon\right)$. We write

$$\begin{aligned}\sqrt{n}(\hat{\lambda}_C - \lambda) =& \sqrt{n}(\psi^{-1}(\hat{\lambda}) - \psi^{-1}(\psi(\lambda))) \\ =& \sqrt{n}\frac{\hat{\lambda} - \psi(\lambda)}{\psi^{(1)}(\lambda)} - \frac{1}{2}\frac{\psi^{(2)}(\lambda)}{(\psi^{(1)}(\lambda))^3}\sqrt{n}(\hat{\lambda} - \psi(\lambda))^2 + o_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned} \tag{A.44}$$

where $\psi(\lambda)$, $\psi^{(1)}(\lambda)$ and $\psi^{(2)}(\lambda)$ are well defined under Assumption 6 and are $O_p(1)$. The order of the remainder in (A.44) follows from standard algebra under Assumptions 3-7, after observing that, from the mean value theorem, the negligible term is

$$\left(\frac{3(\psi^{(2)}(\psi^{-1}(\bar{\lambda})))^2}{(\psi^{(1)}(\psi^{-1}(\bar{\lambda})))^5} - \frac{\psi^{(3)}(\psi^{-1}(\bar{\lambda}))}{(\psi^{(1)}(\psi^{-1}(\bar{\lambda})))^4}\right)\frac{\sqrt{n}(\hat{\lambda} - \psi(\lambda))^3}{6} \tag{A.45}$$

with $|\bar{\lambda} - \psi(\lambda)| < |\hat{\lambda} - \psi(\lambda)| = O_p(1/\sqrt{n})$, from Theorem 1. By $\bar{\lambda} = \psi(\lambda) + O_p(1/\sqrt{n})$ and $\psi^{-1}(\psi(\lambda) + O_p(1/\sqrt{n})) = \lambda + O_p(1/\sqrt{n})$ under Assumption 7, (A.45) becomes

$$\left(\frac{3(\psi^{(2)}(\lambda))^2}{(\psi^{(1)}(\lambda))^5} - \frac{\psi^{(3)}(\lambda)}{(\psi^{(1)}(\lambda))^4}\right)\frac{\sqrt{n}(\hat{\lambda} - \psi(\lambda))^3}{6} + o_p\left(\sqrt{n}(\hat{\lambda} - \psi(\lambda))^3\right) = o_p\left(\frac{1}{\sqrt{n}}\right). \tag{A.46}$$

We define

$$\iota_1 = \frac{1}{n}(m^{11})^{-1}(\hat{\lambda} - \psi(\lambda)) = \frac{1}{n}\left(\epsilon' G' P\epsilon + \beta' X' G' P\epsilon - g_{10}\epsilon' P\epsilon\right), \tag{A.47}$$

$$\iota_2 = \frac{1}{n}(m^{11})^{-1}\psi^{(1)}(\lambda) = \frac{1}{n}\left((m^{11})^{-1} - 2g_{10}\epsilon' G' P\epsilon - 2g_{10}\beta' X' G' P\epsilon + g_{20}\epsilon' P\epsilon\right) \tag{A.48}$$

and

$$\iota_3 = \frac{1}{n}(m^{11})^{-1}\psi^{(2)}(\lambda) = \frac{2}{n}\left((m^{11})^{-1}g_{10} - 2g_{20}\epsilon' G' P\epsilon - 2g_{20}\beta' X' G' P\epsilon + g_{30}\epsilon' P\epsilon\right), \tag{A.49}$$

where, by standard algebra of quadratic forms in normal variates, under Assumptions 1-5, $\iota_1 = O_p(1/\sqrt{n})$, $\iota_2 = O_p(1)$ (and non zero under Assumption 7) and $\iota_3 = O_p(1)$. Also, let

$$\bar{\iota}_3 = 2\left(\delta(I)g_{10} + \sigma^2(g_{11}g_{10} - 2g_{20}g_{10} + g_{30})\right). \tag{A.50}$$

Thus, (A.44) becomes

$$\sqrt{n}(\hat{\lambda}_C-\lambda)=\sqrt{n}\frac{\iota_1}{\iota_2}-\frac{\sqrt{n}}{2}\frac{\iota_3}{\iota_2^3}\iota_1^2+o_p\left(\frac{1}{\sqrt{n}}\right)=\sqrt{n}\frac{\iota_1}{\iota_2}-\frac{\sqrt{n}}{2}\frac{\bar{\iota}_3}{a^3}\iota_1^2+o_p\left(\frac{1}{\sqrt{n}}\right)$$
$$=\sqrt{n}\frac{\iota_1}{\iota_2}-\frac{\sqrt{n}}{2}\frac{\bar{\iota}_3}{a^2}\frac{\iota_1^2}{\iota_2}+o_p\left(\frac{1}{\sqrt{n}}\right), \tag{A.51}$$

where the second equality is deduced by the delta method expansion

$$\frac{\iota_3}{\iota_2^3}-\frac{\bar{\iota}_3}{a^3}=\frac{1}{a^3}(\iota_3-\bar{\iota}_3)-\frac{3\bar{\iota}_3}{a^4}(\iota_2-a)+....=O_p\left(\frac{1}{\sqrt{n}}\right), \tag{A.52}$$

with the last displayed bound following from Lemma 1, and from a and $\bar{\iota}_3$ being $O(1)$. Similarly, the third equality in (A.51) follows since

$$\frac{\sqrt{n}}{2}\frac{\bar{\iota}_3}{a^3}\iota_1^2=\frac{\sqrt{n}}{2}\frac{\bar{\iota}_3}{a^3}\iota_1^2\frac{a+(\iota_2-a)}{\iota_2}=\frac{\sqrt{n}}{2}\frac{\bar{\iota}_3}{a^2}\frac{\iota_1^2}{\iota_2}+o_p\left(\frac{1}{\sqrt{n}}\right), \tag{A.53}$$

by Lemma 1 and from $\iota_1^2/\iota_2=O_p(1/n)$.

Setting $\bar{\iota}=\bar{\iota}_3/a^2$, for any real ζ,

$$Pr\left(\sqrt{n}(\hat{\lambda}_C-\lambda)\leq\zeta\right)=Pr\left(\sqrt{n}\iota_1-\frac{\sqrt{n}}{2}\bar{\iota}\iota_1^2-\zeta\iota_2\leq 0\right)+o\left(\frac{1}{\sqrt{n}}\right), \tag{A.54}$$

since ι_2 is strictly positive over $\lambda\in\Lambda$ under Assumption 7. Therefore, similarly to the proof of Theorem 2,

$$Pr\left(\sqrt{n}(\hat{\lambda}_C-\lambda)\leq\zeta\right)=Pr\left(\xi_C+\frac{1}{\sqrt{n}}\rho_C\leq 0\right)+o\left(\frac{1}{\sqrt{n}}\right), \tag{A.55}$$

where $\xi_C=\epsilon' C_C\epsilon+c_C'\epsilon+d_C$, with

$$C_C=\frac{1}{2\sqrt{n}}\left(1+\frac{2\zeta}{\sqrt{n}}g_{10}\right)(G'P+PG)-\frac{1}{\sqrt{n}}\left(g_{10}+\frac{\zeta}{\sqrt{n}}g_{20}\right)P-\frac{\zeta}{n}G'PG, \tag{A.56}$$

$$c_C'=\frac{1}{\sqrt{n}}\left(1+\frac{2\zeta}{\sqrt{n}}g_{10}\right)\beta'X'G'P-\frac{2\zeta}{n}\beta'X'G'PG, \tag{A.57}$$

$d_C=-\delta(I)\zeta$, and $\rho_C=\rho_{C1}+\rho_{C2}+\rho_{C3}$, with

$$\rho_{C1}=\frac{\bar{\iota}}{2n}\epsilon'\left(\frac{G'P+PG}{2}-g_{10}P\right)\epsilon\epsilon'\left(\frac{G'P+PG}{2}-g_{10}P\right)\epsilon, \tag{A.58}$$

$$\rho_{C2} = \frac{\bar{\iota}}{2n}\beta' X' G' P \epsilon\epsilon' PGX\beta, \quad \rho_{C3} = \frac{\bar{\iota}}{n}\epsilon' \left(\frac{G'P + PG}{2} - g_{10}P\right)\epsilon\epsilon' PGX\beta. \tag{A.59}$$

Proceeding as in (A.40) and (A.41), we derive the first three cumulants of $\xi_C + \rho_C/\sqrt{n}$ (indicated again as κ_1, κ_2 and κ_3) by standard algebra involving moments of quadratic forms in normal random variables, i.e.

$$\begin{aligned}\kappa_1 =& \sigma^2 tr(C_C) + d_C + \frac{1}{\sqrt{n}}E(\rho_{C1}) + \frac{1}{\sqrt{n}}E(\rho_{C2}) \\ =& -\left(\delta(I) + \sigma^2(g_{11} + g_{20} - 2g_{10}^2)\right)\zeta - \frac{\sigma^2}{\sqrt{n}}tr(GX(X'X)^{-1}X') + \frac{\sigma^2}{\sqrt{n}}g_{10}k \\ &+ \frac{\bar{\iota}\sigma^2}{2\sqrt{n}}(\sigma^2(g_{20} + g_{11} - 2g_{10}^2) + \delta(I)) + o\left(\frac{1}{\sqrt{n}}\right) \\ =& -a\zeta - \frac{\sigma^2}{\sqrt{n}}tr(GX(X'X)^{-1}X') + \frac{\sigma^2}{\sqrt{n}}g_{10}k + \frac{\bar{\iota}\sigma^2}{2\sqrt{n}}a + o\left(\frac{1}{\sqrt{n}}\right),\end{aligned} \tag{A.60}$$

with a defined in (2.9). Similarly,

$$\begin{aligned}\kappa_2 =& E\left(\left(\xi_C + \frac{1}{\sqrt{n}}\rho_C\right)^2\right) - \kappa_1^2 = 2\sigma^4 tr(C_C^2) + \sigma^2 c_C' c_C + o\left(\frac{1}{\sqrt{n}}\right) \\ =& \sigma^2 a + \frac{4\sigma^2}{\sqrt{n}}\left((g_{20}g_{10} + 2g_{11}g_{10} - 2g_{10}^3 - g_{21})\sigma^2 + \delta(I)g_{10} - \delta(G)\right)\zeta + o\left(\frac{1}{\sqrt{n}}\right) \\ =& \sigma^2 a\left(1 + \frac{4\zeta}{\sqrt{n}a}\left(\sigma^2(g_{20}g_{10} + 2g_{11}g_{10} - 2g_{10}^3 - g_{21}) + \delta(I)g_{10} - \delta(G)\right)\right) + o\left(\frac{1}{\sqrt{n}}\right),\end{aligned} \tag{A.61}$$

since

$$\begin{aligned}2\sigma^4 tr\left(C_C^2\right) =& \sigma^4\left(g_{20} + g_{11} - 2g_{10}^2\right) + \frac{4\sigma^4}{\sqrt{n}}\left(g_{20}g_{10} + 2g_{11}g_{10} - 2g_{10}^3 - g_{21}\right)\zeta + o\left(\frac{1}{\sqrt{n}}\right), \\ \sigma^2 c_C' c_C =& \sigma^2\delta(I) + \frac{4\sigma^2}{\sqrt{n}}\left(\delta(I)g_{10} - \delta(G)\right)\zeta + o\left(\frac{1}{\sqrt{n}}\right).\end{aligned} \tag{A.62}$$

Denoting $\kappa_s^c = \kappa_s/\kappa_2^{s/2}$, we deduce

$$\begin{aligned}\kappa_1^c =& -\left(\frac{a}{\sigma^2}\right)^{1/2}\zeta - \frac{1}{\sqrt{n}}\left(\frac{\sigma^2}{a}\right)^{1/2}\left(tr(GX(X'X)^{-1}X') - g_{10}k - \frac{\bar{\iota}a}{2}\right) \\ &+ \frac{1}{\sqrt{n}}\frac{2\zeta^2}{(\sigma^2 a)^{1/2}}\left(\sigma^2(g_{20}g_{10} + 2g_{11}g_{10} - 2g_{10}^3 - g_{21}) + \delta(I)g_{10} - \delta(G)\right) + o\left(\frac{1}{\sqrt{n}}\right).\end{aligned} \tag{A.63}$$

Similarly, we deduce

$$\kappa_3 = 8\sigma^6 tr(C_C^3) + 6\sigma^4 c_C' C c_C + \frac{3}{\sqrt{n}}\sigma^4 \bar{\iota}(a-\delta(I))^2 + o\left(\frac{1}{\sqrt{n}}\right)$$
$$= \frac{6\sigma^6}{\sqrt{n}}\left(\frac{g_{30}}{3} + g_{21} + \frac{8}{3}g_{10}^3 - 2g_{10}(g_{20}+g_{11})\right) + \frac{6\sigma^4}{\sqrt{n}}\left(\delta(G) - g_{10}\delta(I) + \frac{\bar{\iota}}{2}(a-\delta(I))^2\right) + o\left(\frac{1}{\sqrt{n}}\right), \tag{A.64}$$

since $c_C' C_C c_C = \left(\delta(G) - g_{10}\delta(I)\right)/\sqrt{n} + o\left(1/\sqrt{n}\right)$ and

$$8tr(C_C^3) = \frac{6}{\sqrt{n}}\left(\frac{g_{30}}{3} + g_{21} + \frac{8}{3}g_{10}^3 - 2g_{10}(g_{20}+g_{11})\right) + o\left(\frac{1}{\sqrt{n}}\right), \tag{A.65}$$

and thus

$$\kappa_3^c = \frac{6\sigma}{\sqrt{n}}\frac{b_C}{a^{3/2}} + o\left(\frac{1}{\sqrt{n}}\right), \tag{A.66}$$

where b_C is defined as (4.5). So, after Taylor expansion of $\Phi(-\kappa_1^c)$,

$$Pr\left(\sqrt{n}(\hat{\lambda}_C - \lambda) \le \zeta\right) = \Phi\left(\left(\frac{a}{\sigma^2}\right)^{1/2}\zeta\right) + \frac{1}{\sqrt{n}}e_{1C}\left(\left(\frac{a}{\sigma^2}\right)^{1/2}\zeta\right)\phi\left(\left(\frac{a}{\sigma^2}\right)^{1/2}\zeta\right)$$
$$- \frac{1}{\sqrt{n}}\frac{\sigma b_C}{a^{3/2}}\Phi^{(3)}\left(\left(\frac{a}{\sigma^2}\right)^{1/2}\zeta\right) + o\left(\frac{1}{\sqrt{n}}\right), \tag{A.67}$$

where

$$e_{1C}(x) = \left(\frac{\sigma^2}{a}\right)^{1/2}\left(tr(GX(X'X)^{-1}X') - g_{10}k - \frac{\bar{\iota}a}{2}\right) + \frac{2\sigma}{a^{3/2}}\left((2g_{10}^3 + g_{21} - 2g_{11}g_{10} - g_{20}g_{10})\sigma^2\right.$$
$$\left. + \delta(G) - \delta(I)g_{10}\right)x^2. \tag{A.68}$$

Thus, the claim in (4.8) follows by $\Phi^{(3)}(x) = (x^2-1)\phi(x)$ and with $e_C(\cdot)$ as in (4.6).

Proof of Theorem 4

We proceed similarly to the proof of Theorem 2. By the mean value theorem,

$$\sqrt{n}\hat{t}_C(\hat{\lambda}_C-\lambda) =\sqrt{n}t_C(\hat{\lambda}_C-\lambda)+\sqrt{n}(\hat{\lambda}_C-\lambda)\frac{\partial t_C}{\partial \beta'}|_{\lambda,\beta,\sigma^2}(\hat{\beta}_C-\beta)+\sqrt{n}(\hat{\lambda}_C-\lambda)\frac{\partial t_C}{\partial \sigma^2}|_{\lambda,\beta,\sigma^2}(\hat{\sigma}^2_C-\sigma^2)$$
$$+\sqrt{n}\frac{\partial t_C}{\partial \lambda}|_{\lambda,\beta,\sigma^2}(\hat{\lambda}_C-\lambda)^2+\frac{1}{2}\sqrt{n}(\hat{\lambda}_C-\lambda)\begin{pmatrix}\hat{\lambda}_C-\lambda\\ \hat{\beta}_C-\beta\\ \hat{\sigma}^2_C-\sigma^2\end{pmatrix}'\bar{H}\begin{pmatrix}\hat{\lambda}_C-\beta\\ \hat{\beta}_C-\beta\\ \hat{\sigma}^2_C-\sigma^2\end{pmatrix}, \tag{A.69}$$

where $\bar{H}$ is the $(k+2)\times(k+2)$ matrix of second derivatives, i.e.

$$H=\begin{pmatrix}\frac{\partial^2 t_C}{\partial\lambda^2} & \frac{\partial^2 t_C}{\partial\lambda\partial\beta'} & \frac{\partial^2 t_C}{\partial\lambda\partial\sigma^2}\\ \frac{\partial^2 t_C}{\partial\beta\partial\lambda} & \frac{\partial^2 t_C}{\partial\beta\partial\beta'} & \frac{\partial^2 t_C}{\partial\beta\partial\sigma^2}\\ \frac{\partial^2 t_C}{\partial\sigma^2\partial\lambda} & \frac{\partial^2 t_C}{\partial\sigma^2\partial\beta'} & \frac{\partial^2 t_C}{\partial(\sigma^2)^2}\end{pmatrix}, \tag{A.70}$$

evaluated at $\bar{\beta}_C$, $\bar{\sigma}^2_C$ and $\bar{\lambda}_C$ such that $||\bar{\beta}_C-\beta||<||\hat{\beta}_C-\beta||$, $|\bar{\sigma}^2_C-\sigma^2|<|\hat{\sigma}^2_C-\sigma^2|$ and $|\bar{\lambda}_C-\lambda|<|\hat{\lambda}_C-\lambda|$ respectively. From (A.51) and (5.1),

$$\hat{\lambda}_C-\lambda=\frac{\iota_1}{\iota_2}-\frac{1}{2}\frac{\bar{\iota}_3}{a^2}\frac{\iota_1^2}{\iota_2}+o_p\left(\frac{1}{n}\right), \tag{A.71}$$

$$\begin{aligned}\hat{\beta}_C-\beta&=\left(X'X\right)^{-1}X'\epsilon-\left(X'X\right)^{-1}X'GX\beta(\hat{\lambda}_C-\lambda)+o_p\left(\frac{1}{\sqrt{n}}\right)\\ &=\left(X'X\right)^{-1}X'\epsilon-\left(X'X\right)^{-1}X'GX\beta\frac{\iota_1}{\iota_2}+o_p\left(\frac{1}{\sqrt{n}}\right)\\ &=\left(X'X\right)^{-1}X'\epsilon-\left(X'X\right)^{-1}X'GX\beta\frac{\iota_1}{a}+o_p\left(\frac{1}{\sqrt{n}}\right)\end{aligned} \tag{A.72}$$

and

$$\begin{aligned}\hat{\sigma}^2_C-\sigma^2&=\frac{1}{n}y'S(\hat{\lambda}_C)'PS(\hat{\lambda}_C)y-\sigma^2=\frac{1}{n}\epsilon'\epsilon-\sigma^2-\frac{2}{n}\epsilon'G'P\epsilon(\hat{\lambda}_C-\lambda)+o_p\left(\frac{1}{\sqrt{n}}\right)\\ &=\frac{1}{n}\epsilon'\epsilon-\sigma^2-\frac{2}{n}\epsilon'G'P\epsilon\frac{\iota_1}{\iota_2}+o_p\left(\frac{1}{\sqrt{n}}\right)=\frac{1}{n}\epsilon'\epsilon-\sigma^2-\frac{2}{n}\epsilon'G'P\epsilon\frac{\iota_1}{a}+o_p\left(\frac{1}{\sqrt{n}}\right)\end{aligned} \tag{A.73}$$

with ι_1, ι_2, $\bar{\iota}_3$ and a defined in (A.47), (A.48), (A.50) and (2.9), respectively. The last equality in both (A.72) and (A.73) follow from the same argument adopted to obtain (A.51) and by Lemma 1.

Let

$$d_\lambda\equiv\frac{\partial t_C}{\partial\lambda}=\frac{1}{t_C}\left(\frac{1}{n\sigma^2}\beta'X'G'PG^2X\beta+g_{30}+g_{21}-2g_{10}g_{20}\right), \tag{A.74}$$

$$d_\beta \equiv \frac{\partial t_C}{\partial \beta'} = \frac{1}{n\sigma^2 t_C}\beta' X' G' P G X \quad \text{and} \quad d_{\sigma^2} \equiv \frac{\partial t_C}{\partial \sigma^2} = -\frac{\delta(I)}{2\sigma^4 t_C}. \tag{A.75}$$

Thus, (A.69) can be written as

$$\begin{aligned}\sqrt{n}\hat{t}_C(\hat{\lambda}_C - \lambda) =& \sqrt{n} t_C \frac{\iota_1}{\iota_2} + \sqrt{n}\frac{\iota_1^2}{a\iota_2}\left(d_\lambda - d_\beta (X'X)^{-1}X'GX\beta - d_{\sigma^2}\epsilon'(G'P+PG)\epsilon - \frac{t_C\bar{\iota}_3}{2a}\right)\\ &+\sqrt{n}\frac{\iota_1}{\iota_2}\left(d_\beta(X'X)^{-1}X'\epsilon + d_{\sigma^2}\left(\frac{1}{n}\epsilon'\epsilon - \sigma^2\right)\right) + o_p\left(\frac{1}{\sqrt{n}}\right)\\ =& \sqrt{n} t_C \frac{\iota_1}{\iota_2} + \sqrt{n}\frac{\iota_1^2}{\iota_2}\nu_{CS} - \frac{1}{\sqrt{n}}\frac{\iota_1^2}{\iota_2 a} d_{\sigma^2}\epsilon'(G'P+PG)\epsilon + \sqrt{n}\frac{\iota_1}{\iota_2}d_\beta(X'X)^{-1}X'\epsilon\\ &+\frac{1}{\sqrt{n}}d_{\sigma^2}\frac{\iota_1}{\iota_2}(\epsilon'\epsilon - n\sigma^2) + o_p\left(\frac{1}{\sqrt{n}}\right),\end{aligned} \tag{A.76}$$

where

$$\nu_{CS} = \frac{1}{a}\left(d_\lambda - d_\beta(X'X)^{-1}X'GX\beta - \frac{t_C\bar{\iota}_3}{2a}\right) = \frac{1}{(a^3\sigma^2)^{1/2}}(\delta(G) - \delta(I)g_{10} + \sigma^2(g_{21} - 2g_{11}g_{20} - g_{10}(g_{11} - 2g_{20}))), \tag{A.77}$$

as defined in (5.2).

Therefore, under Assumption 7, for any real η

$$Pr\left(\sqrt{n}\hat{t}_C(\hat{\lambda}_C - \lambda) \le \eta\right) = Pr\left(\epsilon' C_{CS}\epsilon + c_{CS}'\epsilon + d_{CS} + \frac{1}{\sqrt{n}}\rho_{CS} \le 0\right) + o_p\left(\frac{1}{\sqrt{n}}\right), \tag{A.78}$$

with

$$C_{CS} = \frac{1}{2\sqrt{n}}\left(t_C + \frac{2\eta}{\sqrt{n}}g_{10}\right)(G'P+PG) - \frac{1}{\sqrt{n}}\left(t_C g_{10} + \frac{\eta}{\sqrt{n}}g_{20}\right)P - \frac{\eta}{n}G'PG, \tag{A.79}$$

$$c_{CS}' = \frac{1}{\sqrt{n}}\left(t_C + \frac{2\eta}{\sqrt{n}}g_{10}\right)\beta' X' G' P - \frac{2\eta}{n}\beta' X' G' P G, \quad d_{CS} = -\eta\delta(I), \tag{A.80}$$

and $\rho_{CS} = \sum_{i=1}^{10}\rho_{CSi}$,

$$\begin{aligned}\rho_{CS1} &= \frac{\nu_{CS}}{n}\epsilon'\left(\frac{PG+G'P}{2} - g_{10}P\right)\epsilon\epsilon'\left(\frac{PG+G'P}{2} - g_{10}P\right)\epsilon, \quad \rho_{CS2} = \frac{\nu_{CS}}{n}\beta'X'G'P\epsilon\epsilon'PGX\beta,\\ \rho_{CS3} &= \frac{2\nu_{CS}}{n}\epsilon'\left(\frac{PG+G'P}{2} - g_{10}P\right)\epsilon\epsilon'PGX\beta, \quad \rho_{CS4} = \frac{\delta(I)}{n^2(a\sigma^2)^{3/2}}\epsilon'\left(\frac{PG+G'P}{2}\right)\epsilon\beta'X'G'P\epsilon\epsilon'PGX\beta,\\ \rho_{CS5} &= \frac{\delta(I)}{n^2(a\sigma^2)^{3/2}}\epsilon'\left(\frac{PG+G'P}{2}\right)\epsilon\epsilon'\left(\frac{PG+G'P}{2} - g_{10}P\right)\epsilon\epsilon'\left(\frac{PG+G'P}{2} - g_{10}P\right)\epsilon,\\ \rho_{CS6} &= \frac{2\delta(I)}{n^2(a\sigma^2)^{3/2}}\epsilon'\left(\frac{PG+G'P}{2}\right)\epsilon\epsilon'\left(\frac{PG+G'P}{2} - g_{10}P\right)\epsilon\beta'X'G'P\epsilon,\end{aligned}$$

$$\rho_{CS7} = \frac{1}{n(\sigma^2 a)^{1/2}} \epsilon' \left(\frac{PG + G'P}{2} - g_{10} P \right) \epsilon \beta' X' G' PGX(X'X)^{-1} X' \epsilon,$$

$$\rho_{CS8} = \frac{1}{n(\sigma^2 a)^{1/2}} \beta' X' G' PGX(X'X)^{-1} X' \epsilon \epsilon' PGX\beta, \quad \rho_{CS9} = -\frac{\delta(I)}{2n\sigma^3 a^{1/2}} \beta' X' G' P \epsilon (\epsilon'\epsilon - n\sigma^2),$$

$$\rho_{CS10} = -\frac{\delta(I)}{2n\sigma^3 a^{1/2}} \epsilon' \left(\frac{PG + G'P}{2} - g_{10} P \right) \epsilon (\epsilon'\epsilon - n\sigma^2). \tag{A.81}$$

Similarly to the arguments employed in the proofs of Theorems 2 and 3, we derive the first three cumulants of $\epsilon' C_{CS} \epsilon + c'_{CS} \epsilon + d_{CS} + \rho_{CS}/\sqrt{n}$ (κ_1, κ_2 and κ_3, respectively) and the centred cumulants κ_1^c and κ_2^c as

$$\begin{aligned}
\kappa_1 =& \sigma^2 tr(C_{CS}) + d_{CS} + \frac{1}{\sqrt{n}} E(\rho_{CS}) \\
=& -\left(\delta(I) + \sigma^2 (g_{11} + g_{20} - 2g_{10}^2) \right) \eta + \frac{t_C \sigma^2}{\sqrt{n}} \left(k g_{10} - tr(G'X(X'X)^{-1}X') \right) + \frac{1}{\sqrt{n}} \nu_{CS} \sigma^2 a \\
&+ \frac{1}{\sqrt{n}} \frac{\delta(I)\sigma^4}{(a\sigma^2)^{3/2}} g_{10} a + o\left(\frac{1}{\sqrt{n}} \right) \\
=& -a\eta + \left(\frac{\sigma^2}{a} \right)^{1/2} \frac{1}{\sqrt{n}} \left(akg_{10} - atr(G'X(X'X)^{-1}X') + \delta(G) + \sigma^2 (g_{21} - 2g_{11}g_{20} - g_{10}(g_{11} - 2g_{20})) \right) \\
&+ o\left(\frac{1}{\sqrt{n}} \right).
\end{aligned} \tag{A.82}$$

Also

$$\begin{aligned}
\kappa_2 =& E\left(\left(\epsilon' C_{CS} \epsilon + c'_{CS} \epsilon + d_{CS} + \frac{1}{\sqrt{n}} \rho_{CS} \right)^2 \right) - \kappa_1^2 \\
=& 2\sigma^4 tr(C_{CS}^2) + \sigma^2 c'_{CS} c_{CS} + \frac{2}{\sqrt{n}} E(\epsilon' c_{CS} \rho_{CS}) + \frac{2}{\sqrt{n}} E(\epsilon' C_{CS} \epsilon \rho_{CS}) - \frac{2\sigma^2}{\sqrt{n}} tr(C_{CS}) E(\rho_{CS}) + o\left(\frac{1}{\sqrt{n}} \right) \\
=& 2\sigma^4 tr(C_{CS}^2) + \sigma^2 c'_{CS} c_{CS} + o\left(\frac{1}{\sqrt{n}} \right),
\end{aligned} \tag{A.83}$$

where the last equality follows from algebra of quadratic forms in normal variates. Thus, similarly to the proof of Theorem 3,

$$\kappa_2 = \sigma^2 t_C^2 a \left(1 + \frac{4\eta}{\sqrt{n} a t_C} \left(\sigma^2 (g_{20} g_{10} + 2g_{11} g_{10} - 2g_{10}^3 - g_{21}) + \delta(I) g_{10} - \delta(G) \right) \right) + o\left(\frac{1}{\sqrt{n}} \right), \tag{A.84}$$

such that

$$\kappa_1^c = -\eta + 2 \left(\frac{\sigma^2}{a^3} \right)^{1/2} \left(\sigma^2 (g_{20} g_{10} + 2g_{11} g_{10} - 2g_{10}^3 - g_{21}) + \delta(I) g_{10} - \delta(G) \right) \eta^2$$

$$+\left(\frac{\sigma^2}{a^3}\right)^{1/2}\left(akg_{10}-atr(G'X(X'X)^{-1}X')+\delta(G)+\sigma^2(g_{21}-2g_{11}g_{20}-g_{10}(g_{11}-2g_{20}))\right)+o\left(\frac{1}{\sqrt{n}}\right). \tag{A.85}$$

Also, by tedious, but straightforward algebra,

$$\begin{aligned}
\kappa_3 =& 8\sigma^6 tr(C_{CS}^3)+6\sigma^4 c_{CS}'C_{CS}c_{CS}+\frac{6a^2}{\sqrt{n}}\left(a\sigma^2\nu_{CS}+\frac{\delta(I)\sigma g_{10}}{a^{1/2}}\right)-\frac{6\delta(I)}{\sqrt{n}}\left(\sigma^2 a\nu_{CS}\delta(I)+\frac{\sigma g_{10}}{a^{1/2}}(a-\delta(I))^2\right)\\
=&\frac{6\sigma^3a^{3/2}}{\sqrt{n}}\left(\frac{g_{30}}{3}+g_{21}+\frac{8}{3}g_{10}^3-2g_{10}(g_{20}+g_{11})\right)+\frac{6\sigma a^{3/2}}{\sqrt{n}}\left(\delta(G)-g_{10}\delta(I)\right)\\
&+\frac{6a^2}{\sqrt{n}}\left(a\sigma^2\nu_{CS}+\frac{\delta(I)\sigma g_{10}}{a^{1/2}}\right)-\frac{6\delta(I)}{\sqrt{n}}\left(\sigma^2 a\nu_{CS}\delta(I)+\frac{\sigma g_{10}}{a^{1/2}}(a-\delta(I))^2\right)+o\left(\frac{1}{\sqrt{n}}\right),
\end{aligned} \tag{A.86}$$

so that

$$\begin{aligned}
\kappa_3^c =& \frac{6}{\sqrt{n}}\left(\frac{\sigma^2}{a}\right)^{3/2}\left(\frac{g_{30}}{3}+g_{21}+\frac{8}{3}g_{10}^3-2g_{10}(g_{20}+g_{11})\right)+\frac{6}{\sqrt{n}}\left(\frac{\sigma^2}{a^3}\right)^{1/2}\left(\delta(G)-g_{10}\delta(I)\right)\\
&+\frac{6}{\sqrt{n}}\left(\frac{\sigma^2}{a^3}\right)^{1/2}\left(a^{3/2}\sigma\nu_{CS}+\delta(I)g_{10}\right)-\frac{6\delta(I)}{\sqrt{n}}\left(\frac{\sigma^2}{a^4}\right)^{1/2}\left(\sigma\nu_{CS}\delta(I)+\frac{g_{10}}{a^{3/2}}(a-\delta(I))^2\right)+o\left(\frac{1}{\sqrt{n}}\right)
\end{aligned}$$

So, after Taylor expansion of $\Phi(-\kappa_1^c)$,

$$Pr\left(\sqrt{n}\hat{t}_C(\hat{\lambda}_C-\lambda)\le\eta\right)=\Phi(\eta)+\frac{1}{\sqrt{n}}e_{1CS}(\eta)\phi(\eta)-\frac{1}{\sqrt{n}}\frac{\sigma b_{CS}}{a^{3/2}}\Phi^{(3)}(\eta)+o\left(\frac{1}{\sqrt{n}}\right), \tag{A.87}$$

where

$$\begin{aligned}
e_{1CS}(x)=&\left(\frac{\sigma^2}{a^3}\right)^{1/2}\left(atr(GX(X'X)^{-1}X')-ag_{10}k-\delta(G)-\sigma^2(g_{21}-2g_{11}g_{20}-g_{10}(g_{11}-2g_{20})\right)\\
&+2\left(\frac{\sigma^2}{a^3}\right)^{1/2}\left((2g_{10}^3+g_{21}-2g_{11}g_{10}-g_{20}g_{10})\sigma^2+\delta(G)-\delta(I)g_{10}\right)x^2
\end{aligned} \tag{A.88}$$

and b_{CS} defined as in (5.3). Thus, the claim in (5.6) follows from $\Phi^{(3)}(x)=(x^2-1)\phi(x)$ and (5.4).

Lemma 1 *Under Assumptions 1-5,* $\iota_2-a=O_p(1/\sqrt{n})$ *and* $\iota_3-\bar{\iota_3}=O_p(1/\sqrt{n})$*.*

Proof: By standard algebra

$$
\begin{aligned}
\iota_2 - a =& \frac{1}{n}\left(\epsilon' G' P G \epsilon - n\sigma^2 g_{11}\right) - \frac{1}{n} g_{10}\left(\epsilon'(G'P + PG)\epsilon - 2n\sigma^2 g_{10}\right) + \frac{1}{n} g_{20}\left(\epsilon' P \epsilon - n\sigma^2\right) \\
&+ \frac{2}{n}\beta' X' G' P (G - g_{10} I)\epsilon.
\end{aligned} \tag{A.89}
$$

By the c_r inequality,

$$
\begin{aligned}
E\left(\iota_2 - a\right)^2 \leq& \frac{K}{n^2}\Big(E\left(\epsilon' G' P G \epsilon - n\sigma^2 g_{11}\right)^2 + E\left(\epsilon'(G'P + PG)\epsilon - 2n\sigma^2 g_{10}\right)^2 \\
&+ E\left(\epsilon' P \epsilon - n\sigma^2\right)^2 + \sigma^2 \beta' X' G' P (G - g_{10} I)(G - g_{10} I)' P G X \beta\Big).
\end{aligned} \tag{A.90}
$$

From standard calculation of moments of quadratic forms under Assumption 1, the first term in (A.94) is

$$
\begin{aligned}
&\frac{K}{n^2}\left(\sigma^4 tr^2(G'PG) + 2\sigma^4 tr(G'PGG'PG) + n^2\sigma^4 g_{11}^2 - 2n\sigma^4 g_{11}^2 + 2n\sigma^4 g_{11} tr(G'X(X'X)^{-1}X'G)\right) \\
=& \frac{K}{n^2}\left(2\sigma^4 tr(G'PGG'PG) + \sigma^4 tr^2(G'X(X'X)^{-1}X'G)\right) = O\left(\frac{1}{n}\right),
\end{aligned} \tag{A.91}
$$

under Assumptions 2-5. By Markov's inequality, the first term on the RHS of (A.94) is $O_p(1/\sqrt{n})$. Similarly, we can show that the second and third terms on the RHS of (A.94) are $O_p(1/\sqrt{n})$ as $n \to \infty$. The last term on the RHS of (A.94) is

$$
\frac{K}{n}\left(\delta(GG') + g_{10}^2\delta(I) - 2\delta(G) g_{10}\right) = O\left(\frac{1}{n}\right) \tag{A.92}
$$

under Assumptions 2-5, as n increases. By Markov's inequality $\iota_2 - a = O_p(1/\sqrt{n})$

Similarly

$$
\begin{aligned}
\iota_3 - \bar{\iota}_3 =& \frac{2}{n} g_{10}\left(\epsilon' G' P G \epsilon - n\sigma^2 g_{11}\right) - \frac{2}{n} g_{20}\left(\epsilon'(G'P + PG)\epsilon - 2n\sigma^2 g_{10}\right) \\
&+ \frac{2}{n} g_{30}\left(\epsilon' P \epsilon - n\sigma^2\right) + \frac{4}{n}\beta' X' G' P G \epsilon (g_{10} - g_{20})
\end{aligned} \tag{A.93}
$$

and by c_r inequality

$$
\begin{aligned}
E\left(\iota_3 - \bar{\iota}_3\right)^2 \leq& \frac{K}{n^2}\Big(E\left(\epsilon' G' P G \epsilon - n\sigma^2 g_{11}\right)^2 + E\left(\epsilon'(G'P + PG)\epsilon - 2n\sigma^2 g_{10}\right)^2 \\
&+ E\left(\epsilon' P \epsilon - n\sigma^2\right)^2 + \sigma^2 \beta' X' G' P G G' P G X \beta (g_{10} - g_{20})^2\Big) = O\left(\frac{1}{n}\right),
\end{aligned} \tag{A.94}
$$

from the same argument above and observing that

$$\frac{1}{n}\beta' X' G' PGG' PGX\beta(g_{10} - g_{20})^2 = \delta(GG')(g_{10} - g_{20})^2 = O(1) \tag{A.95}$$

under Assumptions 2-5. Thus, by Markov's inequality, $\iota_3 - \bar{\iota}_3 = O_p(1/\sqrt{n})$.

Tables

	$n = 30$	$n = 50$	$n = 100$	$n = 200$
normal	0.0170	0.0110	0.0210	0.0250
corrected	0.0820	0.0660	0.0640	0.0520
transformed	0.0600	0.0530	0.0560	0.0500
bootstrap	0.0280	0.0410	0.0320	0.0310
MLE	0.0150	0.0130	0.0220	0.0290

Table 1: Empirical sizes (nominal $\alpha = 0.05$) of tests of H_0 in (3.1) against H_1 in (3.2) for model (1.1) when W is derived from the random exponential distance (a).

	$n = 30$	$n = 50$	$n = 100$	$n = 200$
normal	0.0140	0.0160	0.0290	0.0390
corrected	0.0820	0.0650	0.0620	0.0600
transformed	0.0630	0.0540	0.0570	0.0540
bootstrap	0.0310	0.0410	0.0340	0.0370
MLE	0.0020	0.0200	0.0310	0.0350

Table 2: Empirical sizes (nominal $\alpha = 0.05$) of tests of H_0 in (3.1) against H_1 in (3.2) for model (1.1) when W is the circulant (b).

	$\bar{\lambda}$	$n = 30$	$n = 50$	$n = 100$	$n = 200$
	0.2	0.0860	0.1580	0.2550	0.5720
normal	0.5	0.5690	0.8170	0.9740	1
	0.8	0.9840	1	1	1
	0.2	0.2650	0.3620	0.4830	0.6860
corrected	0.5	0.7840	0.9170	0.9910	1
	0.8	0.9950	1	1	1
	0.2	0.2270	0.3250	0.4510	0.6780
transformed	0.5	0.7530	0.9100	0.9910	1
	0.8	0.9950	1	1	1
	0.2	0.1080	0.2010	0.3300	0.6120
bootstrap	0.5	0.5240	0.8680	0.9730	1
	0.8	0.9850	1	1	1
	0.2	0.1020	0.1840	0.2960	0.5970
MLE	0.5	0.5860	0.8340	0.9800	1
	0.8	0.9880	1	1	1

Table 3: Empirical powers of tests of H_0 (3.1) against H_1 (3.2), with nominal size $\alpha = 0.05$ for model (1.1) when W is derived from the exponential distance (a).

	$\bar{\lambda}$	$n = 30$	$n = 50$	$n = 100$	$n = 200$
	0.2	0.0760	0.1530	0.3170	0.5690
normal	0.5	0.3730	0.7160	0.9720	1
	0.8	0.8950	0.9950	1	1
	0.2	0.2750	0.3630	0.4930	0.6880
corrected	0.5	0.7040	0.8700	0.9910	1
	0.8	0.9870	0.9990	1	1
	0.2	0.2190	0.3230	0.4720	0.6710
transformed	0.5	0.6500	0.8550	0.9900	1
	0.8	0.9780	0.9990	1	1
	0.2	0.1390	0.2180	0.3510	0.6120
bootstrap	0.5	0.5450	0.7970	0.9810	0.9950
	0.8	0.9540	0.9970	1	1
	0.2	0.0670	0.1640	0.3490	0.5910
MLE	0.5	0.4630	0.7580	0.9760	1
	0.8	0.9420	0.9940	1	1

Table 4: Empirical powers of tests of H_0 (3.1) against H_1 (3.2), with nominal size $\alpha = 0.05$ in model (1.1) when W is the circulant (b).

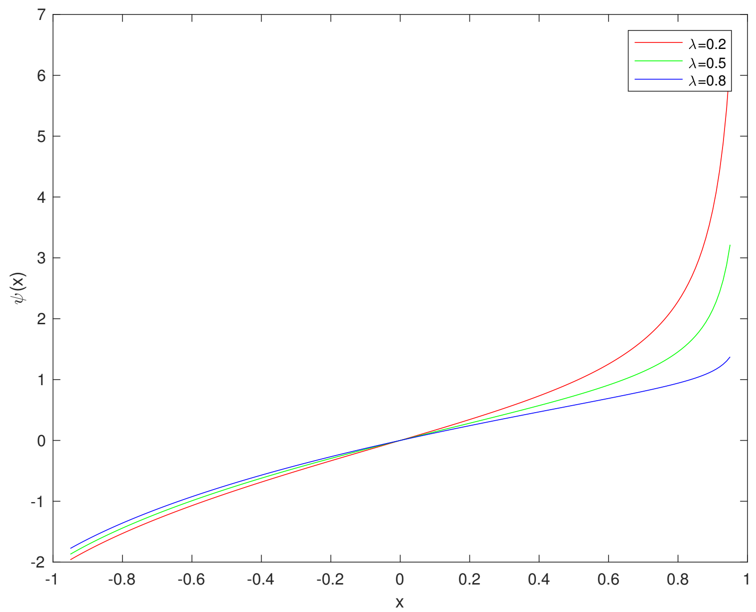

Figure 1: Plots of $\psi(x)$ in (2.7) for $-1 < x < 1$ when W is derived from the exponential distance (a) and $\lambda = 0.2, 0.5, 0.8$. $n = 100$.

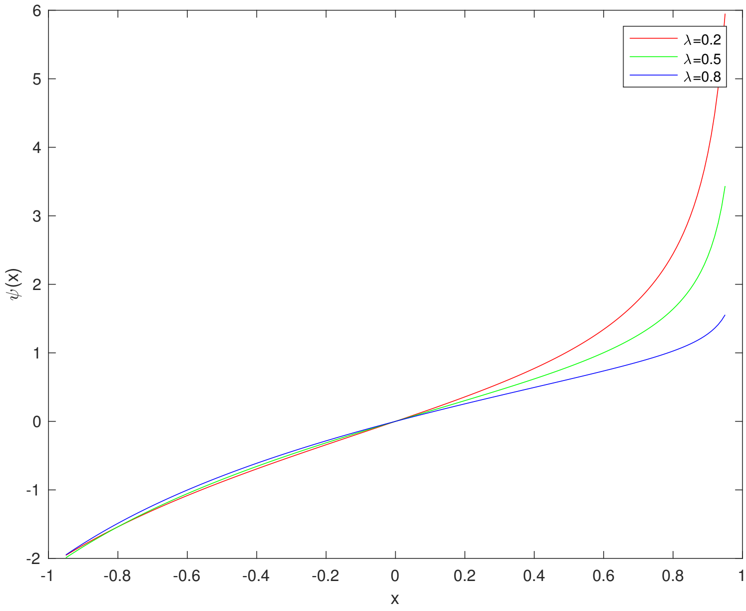

Figure 2: Plots of $\psi(x)$ in (2.7) for $-1 < x < 1$ when W is the circulant (b) and $\lambda = 0.2, 0.5, 0.8$. $n = 100$.

	λ	$n = 30$	$n = 50$	$n = 100$	$n = 200$
normal	0.3	-0.5713 (0.9870)	-0.3625 (0.9940)	-0.2269 (0.9970)	-0.0351 (0.9970)
	0.5	-0.1501 (0.9780)	0.0410 (0.9850)	0.1576 (0.9910)	0.2863 (0.9930)
	0.7	0.3472 (0.9280)	0.4548 (0.9600)	0.5272 (0.9660)	0.5944 (0.9600)
corrected	0.3	-0.3976 (0.9610)	-0.2361 (0.9660)	-0.1378 (0.9810)	0.0026 (0.9870)
	0.5	0.0022 (0.9470)	0.1416 (0.9550)	0.2162 (0.9650)	0.3138 (0.9770)
	0.7	0.4319 (0.9050)	0.5095 (0.9230)	0.5616 (0.9320)	0.6088 (0.9450)
MLE	0.3	-0.5713 (0.9870)	-0.3625 (0.9940)	-0.2269 (0.9970)	-0.0351 (0.9970)
	0.5	-0.1537 (0.9780)	0.0410 (0.9850)	0.1576 (0.9910)	0.2863 (0.9930)
	0.7	0.3580 (0.9280)	0.4496 (0.9600)	0.5269 (0.9660)	0.5944 (0.9600)
bootstrap	0.3	-0.2094 (0.9820)	-0.0662 (0.9720)	0.0467 (0.9780)	0.1399 (0.9680)
	0.5	0.0781 (0.9740)	0.1719 (0.9760)	0.2840 (0.9740)	0.3626 (0.9780)
	0.7	0.3488 (0.9920)	0.4739 (0.9900)	0.5540 (0.9660)	0.6070 (0.9660)

Table 5: Average lower-end-point of intervals (5.7), (5.11), (7.5) and (7.6) across 1000 Monte Carlo replications, and their respective empirical coverage probability (in brackets) for model (1.1) when W is derived from the exponential distance (a). Nominal confidence level $1 - \alpha = 0.95$.

	λ	$n=30$	$n=50$	$n=100$	$n=200$
normal	0.3	-0.7457 (0.9900)	-0.4861 (0.9950)	-0.2105 (0.9970)	-0.0412 (0.9990)
	0.5	-0.3148 (0.9850)	-0.0619 (0.9880)	0.1428 (0.9920)	0.2584 (0.9910)
	0.7	0.1366 (0.9570)	0.3605 (0.9710)	0.5032 (0.9630)	0.5749 (0.9560)
corrected	0.3	-0.5143 (0.9710)	-0.3355 (0.9700)	-0.1349 (0.9810)	0.0038 (0.9870)
	0.5	-0.1330 (0.9450)	0.0561 (0.9660)	0.1991 (0.9670)	0.2882 (0.9770)
	0.7	0.2834 (0.9370)	0.4479 (0.9380)	0.5403 (0.9410)	0.5924 (0.9410)
MLE	0.3	-0.7457 (0.9900)	-0.4861 (0.9950)	-0.2105 (0.9970)	-0.0412 (0.9990)
	0.5	-0.3148 (0.9850)	-0.0619 (0.9880)	0.1428 (0.9920)	0.2584 (0.9910)
	0.7	0.1487 (0.9560)	0.3653 (0.9710)	0.5038 (0.9630)	0.5749 (0.9560)
bootstrap	0.3	-0.2527 (0.9740)	-0.0984 (0.9750)	0.0538 (0.9720)	0.1370 (0.9760)
	0.5	-0.0046 (0.9680)	0.1503 (0.9820)	0.2734 (0.9740)	0.3583 (0.9880)
	0.7	0.2997 (0.9740)	0.4268 (0.9760)	0.5357 (0.9720)	0.5898 (0.9720)

Table 6: Average lower-end-point of intervals (5.7), (5.11), (7.5) and (7.6) across 1000 Monte Carlo replications, and their respective empirical coverage probability (in brackets) for model (1.1) when W is the circulant (b). Nominal confidence level $1-\alpha=0.95$.

References

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

[2] Baltagi, B.H. and Z. Yang (2013). Standardized LM tests for spatial error dependence in linear or panel regressions. *Econometrics Journal* **16**, 103-134.

[3] Bao, Y. (2013) Finite-sample bias of the QMLE in spatial autoregressive models. *Econometric Theory* **29**, 68-88.

[4] Bao, Y. and A. Ullah (2007) Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* **137**, 396-413.

[5] Bao, Y. (2013) Finite-sample bias of the QMLE in spatial autoregressive models. *Econometric Theory* **29**, 68-88.

[6] Bao, Y. and A. Ullah (2007) Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* **137**, 396-413.

[7] Burridge, P. (1980). On the Cliff-Ord test for spatial correlation. *Journal of the Royal Statistical Society, Series B* **42**, 107-8.

[8] Cliff, A. and J.K. Ord (1981). *Spatial Processes: Models & Applications.* Pion.

[9] Delgado, M.A. and P.M. Robinson (2015) Non-nested testing of spatial correlation. *Journal of Econometrics* **187**, 385-401.

[10] Jin, F. and L.F.Lee (2015). On the bootstrap for Moran's I test for spatial dependence. *Journal of Econometrics* **184**, 295-314.

[11] Kelejian, H.H. and I.R. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* **17**, 99-121.

[12] Kelejian, H.H. and I.R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* **40**, 509-533.

[13] Kelejian, H.H. and I.R. Prucha (2001) On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* **104**, 219-257.

[14] Kelejian, H.H. and I.R. Prucha (2010). Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics* **157**, 53-67.

[15] Kyriacou, M, Phillips P.C.B. and Rossi F. (2017). Indirect inference in spatial autoregressions. *The Econometrics Journal* **20**, 168-189.

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

[17] Lee, L.F. (2004). Asymptotic distribution of quasi-maximum likelihood estimates for spatial autoregressive models. *Econometrica* **72**, 1899-1925.

[18] Lee, L.F. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* **137**, 489-514.

[19] Lee, L.F. and J. Yu (2012). The $C(\alpha)$ -type gradient test for spatial dependence in spatial autoregressive models. *Letters in Spatial and Resource Sciences* **5**, 119-135.

[20] Liu, S.F. and Z. Yang (2015). Improved Inferences for Spatial Regression Models. *Regional Science and Urban Economics* **55**, 55-67.

[21] Maekawa, K. (1985). Edgeworth expansion for the OLS estimator in a time series regression model. *Econometric Theory* **1**, 223-239.

[22] Martellosio, F. (2012) Testing for spatial autocorrelation: The regressors that make the power disappear. *Econometric Reviews* 31, 215-240.

[23] Martellosio, F. and G. Hillier (2019). Adjusted QMLE for the Spatial Autoregressive Parameter. Forthcoming in *Journal of Econometrics*.

[24] Ord, J.K. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* **70**, 120-6.

[25] Phillips, P.C.B. (1977). Approximations to some finite sample distributions associated with a first-order stochastic difference equation. *Econometrica* **45**, 463-85.

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

[27] Robinson, P. M. and F. Rossi (2014). Improved Lagrange multiplier tests in spatial autoregressions. *Econometrics Journal* **17**, 139-164.

[28] Robinson, P.M. and F. Rossi (2015). Refined tests for spatial correlation. *Econometric Theory* **31**, 1249-1280.

[29] 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.

[30] Yang, Z. L. (2015). A general method for third order bias and variance corrections on a non-linear estimator. *Journal of Econometrics* **186**, 178-200.