Journal of Econometrics 189 (2015) 447–456

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Refinements in maximum likelihood inference on spatial autocorrelation in panel data

Peter M. Robinson *, Francesca Rossi

London School of Economics, United Kingdom
University of Southampton, United Kingdom

ARTICLE INFO

Article history:
Available online 21 March 2015

JEL classification:
C12
C21
C31

Keywords:
Panel data
Fixed effects
Spatial autoregression
Edgeworth expansion
Interval estimates
Tests for cross-sectional independence

ABSTRACT

In a panel data model with fixed effects, possible cross-sectional dependence is investigated in a spatial autoregressive setting. An Edgeworth expansion is developed for the maximum likelihood estimate of the spatial correlation coefficient. The expansion is used to develop more accurate interval estimates for the coefficient, and tests for cross-sectional independence that have better size properties, than corresponding rules of statistical inference based on first order asymptotic theory. Comparisons of finite sample performance are carried out using Monte Carlo simulations.

1. Introduction

Cross-sectional dependence is an increasingly important issue in the analysis of panel data. Much of the machinery for conducting statistical inference on panel data models has been established under the simplifying assumption of cross-sectional independence. This assumption may be unwarranted, due to various causes such as spillovers and competition. Even when dependence does not entail a loss of consistency of point estimates of parameters of interest, such as regression coefficients, it will typically invalidate interval estimates and hypothesis tests. To remedy matters, various approaches have been proposed to incorporate cross-sectional dependence in panel data models. A nonparametric approach is only feasible when the number, T, of time series observations, is large relative to the number, n, of cross-sectional ones. In other situations, including when T is very small, even 2, parametric models have been employed, including factor models and, when information on spatial distances is available, spatial autoregressive models. Using such models, tests for cross-sectional dependence can be carried out, and estimates of parameters describing dependence obtained, along with measures of variability. These methods are usually based on large-n first order asymptotic approximations, finite sample theory being intractable. When n is not very large such approximations may be unreliable.

The present paper derives rules of statistical inference that promise to be more accurate, in the setting of a panel data model with fixed effects and first-order spatial autoregressive (SAR(1)) cross-sectional dependence,

$$Y_t = c + \lambda_0 W Y_t + V_t, \quad t = 1, \dots, T. \tag{1.1}$$

Here, $Y_t = (y_{1t}, \dots, y_{nt})'$ is an $n \times 1$ vector of observations, c is an $n \times 1$ vector of unknown fixed effects, W is an $n \times n$ non-null matrix of nonstochastic spatial weights with zero diagonal elements, $V_t = (v_{1t}, \dots, v_{nt})'$ is an $n \times 1$ vector of disturbances with v_{it} being independent and identically distributed (i.i.d.) $\mathcal{N}(0, \sigma_0^2)$ across $i = 1, \dots, n$ and $t = 1, \dots, T$, for unknown $\sigma_0^2 > 0$, and the spatial correlation parameter λ_0 is unknown. Asymptotic properties for large n are developed, but for notational simplicity we omit the subscript n from Y_t, V_t, W and c, as well as from various other n-dependent quantities. The vector c can be stochastically generated, in which case it can induce cross-sectional dependence within Y_t, but conditional on c there is dependence if and only if $\lambda_0 \neq 0$, and in any case c introduces an incidental parameters problem. As is standard we get around this by eliminating c at the outset by a linear transformation, so no regularity conditions are imposed on c. This requires $T \geq 2$, and indeed in the case $T = 2$ our transformed model is formally equivalent to the pure cross-sectional one in which $T = 1$ and $c = 0$ *a priori*, and our results are

* Correspondence to: Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. Tel.: +44 20 7955 7516; fax: +44 20 7955 6592.
E-mail address: p.m.robinson@lse.ac.uk (P.M. Robinson).

http://dx.doi.org/10.1016/j.jeconom.2015.03.036

new for this case also. Larger T affords greater statistical efficiency, though it could also allow extension to a more elaborate structure, such as time trends with unknown coefficients varying over the cross section dimension i. It would be possible to extend (1.1) to include explanatory variables with coefficients that are constant over i, but as even (1.1) entails relatively complicated formulae we do not pursue the details here. In fact a regression component could in some respects simplify matters, because having eliminated c we could consistently estimate λ_0, with $n \to \infty$, by instrumental variables or even least squares (cf Kelejian and Prucha, 1998, Lee, 2002), but in (1.1) least squares is inconsistent. Instead we employ the maximum likelihood estimate (MLE), which is only implicitly-defined but is asymptotically efficient. In a simple non-panel SAR(1), i.e. with $T = 1$ and $c = 0$ *a priori* in (1.1), Lee (2004) established consistency and asymptotic normality of the MLE, and this theory is straightforwardly extendable to (1.1) with $T \geq 2$ and $c \neq 0$. Lee and Yu (2010) considered panel data models that incorporate a regression component in (1.1), and a possible time effect, and also allowed V_t to have SAR(1) structure, deriving first order asymptotic theory for the pseudo MLE of the parameters, using two different approaches for eliminating the fixed effects.

We develop higher-order asymptotics for the MLE, using an Edgeworth expansion. Though it is possible to justify validity of Edgeworth expansions for implicitly-defined estimates (see e.g. Bhattacharya and Ghosh, 1978), we focus on practically useful aspects by presenting formal expansions. First-order asymptotics are available under much milder distributional conditions than normality (as in Lee and Yu, 2010, for example) but as in much of the Edgeworth literature we impose normality in order to keep formulae simple. Bao and Ullah (2007) derived the second-order bias and mean squared error of the MLE in (1.1) with $T = 2$ and $c = 0$ *a priori*. Recently, Robinson and Rossi (2014a,b) have developed Edgeworth-improved tests for no spatial correlation in SAR(1) models for pure cross-sectional data based on least squares estimation and Lagrange multiplier tests. It would be possible to extend our results to develop refined inference on the MLE of the spatial correlation parameter in models including explanatory variables (cf e.g. Lee, 2004), though the formulae for interval estimates and tests would be more complicated. It would also be possible to develop refined inference for higher-order SAR models (cf e.g. Lee and Liu, 2010), though the multiparameter aspect would complicate proofs (cf e.g. Taniguchi, 1988 in the Gaussian time series case).

In the following section the MLE is described, regularity conditions are listed, and a formal Edgeworth expansion for its cumulative distribution function (cdf) is presented, whereas Section 3 reports a formal Edgeworth expansion for the cdf of a studentized MLE and deduces confidence intervals for λ_0 that are more accurate than ones based on first-order asymptotics. Section 4 deduces tests of the null hypothesis $\lambda_0 = 0$ that have better size properties than ones based on first-order asymptotics. Section 5 compares our methods with first-order ones in Monte Carlo simulations.

2. Edgeworth expansion

The log-likelihood for (1.1) is given by

$$l(\lambda, \sigma^2) = -\frac{nT}{2}\ln(2\pi) - \frac{nT}{2}\log\sigma^2 + T\log(\det(S(\lambda))) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\|S(\lambda)Y_t - c\|^2, \quad (2.1)$$

where $S(\lambda) = I_n - \lambda W$, I_n is the $n \times n$ identity matrix, $\|.\|$ denotes spectral norm, det(.) is the determinant operator and λ and σ^2 denote any admissible parameter values. Define

$$\tilde{Y}_t = Y_t - \sum_{t=1}^{T} Y_t/T, \qquad \tilde{V}_t = V_t - \sum_{t=1}^{T} V_t/T. \quad (2.2)$$

On concentrating c and σ^2 out, and defining

$$\hat{\sigma}^2(\lambda) = \frac{1}{nT}\sum_{t=1}^{T}\tilde{Y}_t' S(\lambda)' S(\lambda)\tilde{Y}_t, \quad (2.3)$$

the MLE of λ_0 is given by

$$\hat{\lambda} = \arg\max_{\lambda \in \Lambda} l(\lambda),$$

where

$$l(\lambda) = l(\lambda, \hat{\sigma}^2(\lambda)) = -\frac{nT}{2}(\ln(2\pi) + 1) - \frac{nT}{2}\log(\hat{\sigma}^2(\lambda)) + T\log\det(S(\lambda)), \quad (2.4)$$

and Λ is the set of admissible values for λ, assumed compact.

Note that (2.2) transforms (1.1) to

$$S(\lambda_0)\tilde{Y}_t = \tilde{V}_t, \quad t = 1, \ldots, T, \quad (2.5)$$

where $\tilde{V}_t$ is correlated across t, indeed $\sum_{t=1}^{T}\tilde{V}_t \equiv 0$. As in Lee and Yu (2010), for example, define $J = I_T - l_T l_T'/T$, where l_T denotes a $T \times 1$ column of ones, and $V = (V_1', V_2', \ldots, V_T')'$, and for a $T \times (T-1)$ matrix P such that $J = PP'$ and $P'P = I_{T-1}$, let $\epsilon = (P \otimes I_n)'V$, so $E(\epsilon\epsilon') = I_{n(T-1)}\sigma_0^2$. With respect to quadratic forms such as (2.3), it is then useful to note that, for any $n \times n$ matrix D

$$\sum_{t=1}^{T}\tilde{V}_t' D\tilde{V}_t = V'(J \otimes I_n)(I_T \otimes D)(J \otimes I_n)V = r(D), \quad (2.6)$$

where

$$r(D) = \epsilon(I_{T-1} \otimes D)\epsilon. \quad (2.7)$$

We introduce a series of regularity conditions. These are in part motivated by large-n asymptotics, with T kept fixed, in line with the discussion in the previous section. We could develop asymptotics with T increasing with n, or sequential asymptotics with T increasing after n, but there is little practical value in doing so here because in our model $T \to \infty$ is not needed for consistent estimation or to materially simplify the theory. We only mention that we could on the other hand develop theory with T increasing and n held fixed, but this would be relatively trivial as (2.3) then becomes a multivariate model, with unknown but finite-dimensional location c, for T independent observations, and indeed there is no theoretical reason for imposing a parsimonious model such as SAR(1). We will however keep T in normalizing factors to demonstrate the improved rate of convergence that would result in letting $T \to \infty$ with or after n. For a matrix D with (i,j)th element d_{ij}, define the maximum absolute row sum norm $\|D\|_\infty = \max_i \sum_j |d_{ij}|$.

Assumption 1. The v_{it}, $i = 1, \ldots, n$; $t = 1, \ldots T$, are i.i.d. $\mathcal{N}(0, \sigma_0^2)$ random variables.

Assumption 2. $\Lambda = [b_1, b_2]$, where $-1 < b_1 < b_2 < 1$, and λ_0 is an interior point of Λ.

Assumption 3. (i) For all n, $w_{ii} = 0$, $i = 1, \ldots, n$.
(ii) For all n, $\|W\| \leq 1$.
(iii) As $n \to \infty$, $\|W\|_\infty + \|W'\|_\infty = O(1)$.
(iv) As $n \to \infty$, uniformly in $i, j = 1, \ldots, n$, $w_{ij} = O(1/h)$, where $h = h_n$ is bounded away from zero for all n and $h/n \to 0$ as $n \to \infty$.

Assumption 4. As $n \to \infty$, $\sup_{\lambda\in\Lambda}\|S^{-1}(\lambda)\|_\infty + \sup_{\lambda\in\Lambda}\|S^{-1}(\lambda)'\|_\infty = O(1)$.

Assumption 5. For all $\lambda \in \Lambda - \{\lambda_0\}$

$$\underset{n\to\infty}{\underline{\lim}} \frac{n^{-1} tr\left(S^{-1\prime}S(\lambda)'S(\lambda)S^{-1}\right)}{\left(\det\left(S^{-1\prime}S(\lambda)'S(\lambda)S^{-1}\right)\right)^{1/n}} > 1, \tag{2.8}$$

where $S = S(\lambda_0)$.

Assumptions 2 and 3(ii) imply that the series

$$S^{-1}(\lambda) = \sum_{s=0}^{\infty} (\lambda W)^s \tag{2.9}$$

converges and thus that $S(\lambda)$ is nonsingular, indeed $\det(S(\lambda)) > 0$, on Λ. These, or some other suitable restrictions on W and Λ, are also necessary for existence of $\hat{\lambda}$. If W is symmetric with non-negative elements and $Wl = l$, as in the block-diagonal districts-farmers W of Case (1991), Assumption 3(iii) is automatically satisfied and $\|W\|_\infty = 1$. In the latter case, by (2.9) and under Assumption 2, it follows that Assumption 4 holds. The sequence h defined in Assumption 3(iv) can be bounded or divergent, and such a condition on w_{ij} as $n \to \infty$ is generally required to develop asymptotic theory for estimates of parameters in (1.1). Assumption 5 is an identifiability condition, necessary for consistency of $\hat{\lambda}$; the ratio in (2.8) is in any case guaranteed to be no less than 1 by the inequality between arithmetic and geometric means. While these conditions, and Assumption 6 below, are designed for the development of only formal Edgeworth expansions, and are insufficient to justify validity, Assumptions 1–5 are sufficient for consistency of $\hat{\lambda}$, and indeed for $\hat{\lambda} = \lambda_0 + O_p\left((nT/h)^{-1/2}\right)$ as $n \to \infty$, a property used in our proofs.

Define

$$G(\lambda) = WS^{-1}(\lambda), \qquad A(\lambda) = G(\lambda) - \frac{trG(\lambda)}{n} I_n, \tag{2.10}$$

$$a(\lambda) = \frac{h(T-1)}{nT}\left(tr\left(G(\lambda)^2 + G(\lambda)' G(\lambda)\right) - \frac{2}{n}\left(trG(\lambda)\right)^2\right) = \frac{h(T-1)}{2nT} tr\left(\left(A(\lambda) + A(\lambda)'\right)^2\right), \tag{2.11}$$

$$G = G(\lambda_0), \qquad A = A(\lambda_0), \qquad a = a(\lambda_0) \tag{2.12}$$

and

$$f(u) = a^{-3/2}\frac{h(T-1)}{3nT}\left(\frac{8(trG)^3}{n^2} - \frac{6trGtr(G^2+G'G)}{n} + tr(G^3+3G^2G') + \left(tr(2G^3+3G'G^2) - \frac{3trGtr(2G^2+G'G)}{n} + \frac{4(trG)^3}{n^2}\right)u^2\right). \tag{2.13}$$

Under Assumptions 3 and 4 $\|G\|_\infty + \|G'\|_\infty = O(1)$ and $tr(WD) = O(n/h)$ as $n \to \infty$ for any $n \times n$ matrix D such that $\|D\|_\infty + \|D'\|_\infty = O(1)$. Thus $a = O(1)$ as $n \to \infty$. We avoid pathological situations by requiring

Assumption 6.

$$\underset{n\to\infty}{\underline{\lim}}\; a > 0. \tag{2.14}$$

We have the following result.

Theorem 1. *Let model (1.1) and Assumptions 1–6 hold. For any real x the cdf of $(nT/h)^{1/2}(\hat{\lambda} - \lambda_0)$ admits the second order formal Edgeworth expansion*

$$P\left(\left(\frac{nT}{h}\right)^{1/2}(\hat{\lambda} - \lambda_0) \le x\right) = \Phi\left(a^{1/2}x\right) + \left(\frac{h}{nT}\right)^{1/2} f(a^{1/2}x)\phi\left(a^{1/2}x\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right), \tag{2.15}$$

and

$$f(a^{1/2}x) = O(1) \tag{2.16}$$

as $n \to \infty$.

The expansion in (2.15) is justified whether $h = O(1)$ or $h \to \infty$ as $n \to \infty$. In the latter case some simplifications would be possible. We stress that relaxing the assumption of normality would lead to a different, more complicated approximation to the cdf.

3. Improved confidence intervals

In order to derive Edgeworth-corrected confidence intervals we need the second order Edgeworth expansion of the studentized MLE of λ_0, i.e.

$$\left(\frac{nT}{h}\right)^{1/2}\widehat{a}^{1/2}(\hat{\lambda} - \lambda_0), \tag{3.1}$$

where $\widehat{a} = a(\hat{\lambda})$. Define

$$d(\lambda) = \frac{T-1}{T}\frac{h}{n}\left(tr\left(G(\lambda)^3 + G(\lambda)^2 G(\lambda)'\right) - \frac{2}{n} trG(\lambda)\, tr\left(G(\lambda)^2\right)\right) \tag{3.2}$$

and

$$d = d(\lambda_0). \tag{3.3}$$

We obtain

Theorem 2. *Let model (1.1) and Assumptions 1–6 hold. For any real ζ the cdf of $(nT/h)^{1/2}\widehat{a}^{1/2}(\hat{\lambda} - \lambda_0)$ admits the second order formal Edgeworth expansion*

$$P\left(\left(\frac{nT}{h}\right)^{1/2}\widehat{a}^{1/2}(\hat{\lambda} - \lambda_0) \le \zeta\right) = \Phi(\zeta) + \left(\frac{h}{nT}\right)^{1/2}\left(f(\zeta) - \frac{d}{a^{3/2}}\zeta^2\right)\phi(\zeta) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right), \tag{3.4}$$

where $f(.)$ is defined in (2.13) and

$$f(\zeta) - \frac{d}{a^{3/2}}\zeta^2 = O(1) \tag{3.5}$$

as $n \to \infty$.

Again our approximate cdf is not robust to departures from normality. A robust one would involve cumulants, which would be likely estimated imprecisely in modest samples.

From Theorem 2 we can derive Edgeworth-improved confidence intervals. We focus on intervals of the form $(-\infty, U)$, where U is a suitable upper end-point, but similar results hold for (L, ∞), where L is a lower end-point. For $\alpha \in (0, 1)$, let $I = (-\infty, \hat{\lambda} - (h/nT)^{1/2}\widehat{a}^{-1/2}w_{1-\alpha})$ such that

$$P(\lambda_0 \in I) = 1 - \alpha, \tag{3.6}$$

where $w_{1-\alpha}$ denotes the true $\alpha-$quantile of the cdf of $(nT/h)^{1/2}\widehat{a}^{1/2}(\hat{\lambda} - \lambda_0)$, and

$$I^{\mathcal{N}} = (-\infty, \hat{\lambda} - (h/nT)^{1/2}\widehat{a}^{-1/2}z_{1-\alpha}), \tag{3.7}$$

where $\Phi(z_\alpha) = 1 - \alpha$. Also, we define the (infeasible) Edgeworth-corrected interval as $I^{Ed} = (-\infty, \hat{\lambda} - (h/nT)^{1/2}\widehat{a}^{-1/2}v_{1-\alpha})$, where

$$v_{1-\alpha} = z_{1-\alpha} - \left(\frac{h}{nT}\right)^{1/2}\left(f(z_{1-\alpha}) - \frac{d}{a^{3/2}}z_{1-\alpha}^2\right)$$
$$= -z_\alpha - \left(\frac{h}{nT}\right)^{1/2}\left(f(z_\alpha) - \frac{d}{a^{3/2}}z_\alpha^2\right), \tag{3.8}$$

which depends on the unknown λ_0. Let $\hat{d} = d(\hat{\lambda})$ and $\hat{f}(.)$ be as defined in (2.13) with G and a replaced by $\hat{G} = G(\hat{\lambda})$ and $\widehat{a}$. Since $\hat{\lambda}$ converges to λ_0 at rate $(nT/h)^{1/2}$, we expect (e.g. Hall, 1992) the feasible version of J^{Ed}, obtained by respectively replacing $f(.)$, d and a in (3.8) with $\hat{f}(.)$, $\hat{d}$ and $\widehat{a}$, to retain the same higher-order properties. Define

$$\hat{I}^{Ed} = (-\infty, \hat{\lambda} - (h/nT)^{1/2}\widehat{a}^{-1/2}\hat{v}_{1-\alpha}) \tag{3.9}$$

where

$$\hat{v}_{1-\alpha} = -z_\alpha - \left(\frac{h}{nT}\right)^{1/2}\left(\hat{f}(z_\alpha) - \frac{\hat{d}}{\widehat{a}^{3/2}}z_\alpha^2\right). \tag{3.10}$$

From Theorem 2 we deduce

Corollary 1. *Let model (1.1) and Assumptions 1–6 hold.*

$$P(\lambda_0 \in I^{\mathcal{N}}) = P(\lambda_0 \in I) + O\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$
$$= 1 - \alpha + O\left(\left(\frac{h}{nT}\right)^{1/2}\right) \tag{3.11}$$
$$P(\lambda_0 \in \hat{I}^{Ed}) = P(\lambda_0 \in I) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$
$$= 1 - \alpha + o\left(\left(\frac{h}{nT}\right)^{1/2}\right) \tag{3.12}$$

as $n \to \infty$.

Note that the interval $\hat{I}^{Ed}$, while more complicated than $I^{\mathcal{N}}$, is a closed form function of $\hat{\lambda}$ and given quantities, and can be rapidly computed.

Two-sided improved confidence intervals could be constructed similarly starting from a third-order Edgeworth expansion of the cdf of (3.1). We focus here on one-sided intervals since very often in practical applications the sign of λ_0 can be conjectured. Moreover, from parity properties of the second-order term in (3.4), the standard two-sided confidence interval based on the asymptotic critical values is expected to have coverage probability $1 - \alpha + O(h/(nT))$, unlike the result displayed in (3.11), and thus the derivation of Edgeworth corrections seems more necessary in case of one-sided intervals.

4. Improved tests

We are interested in testing

$$H_0: \ \lambda_0 = 0 \tag{4.1}$$

against a one-sided alternative

$$H_1: \ \lambda_0 > 0. \tag{4.2}$$

We define (2.13) under H_0 as

$$f_0(u) = \left(\frac{h(T-1)}{nT}\right)^{-1/2} \times \frac{(tr(W^3 + 3W^2W') + tr(2W^3 + 3W'W^2)u^2)}{3tr^{3/2}(W^2 + W'W)}, \tag{4.3}$$

since under H_0 in (4.1) $G = W$ so that $trG = trW = 0$ under Assumption 3(i), and

$$a = a(0) = \frac{h(T-1)}{nT}tr(W^2 + W'W). \tag{4.4}$$

Thus f_0 is a completely known function. By choosing $x = a^{-1/2}\zeta$ in (2.15) and (2.16), we deduce

Corollary 2. *Let model (1.1) and Assumptions 1–6 hold. For any real ζ, under H_0 in (4.1), the cdf of $(nT/h)^{1/2}a^{1/2}\hat{\lambda}$ admits the second order formal Edgeworth expansion*

$$P\left(\left(\frac{nT}{h}\right)^{1/2}a^{1/2}\hat{\lambda} \le \zeta\right) = \Phi(\zeta) + \left(\frac{h}{nT}\right)^{1/2}f_0(\zeta)\phi(\zeta) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right), \tag{4.5}$$

and

$$f_0(\zeta) = O(1) \tag{4.6}$$

as $n \to \infty$.

Corollary 2 can be used to deduce improved tests of (4.1). Let u_α be the $(1-\alpha)$ quantile of the cdf of $(nT/h)^{1/2}a^{1/2}\hat{\lambda}$, and

$$s_\alpha = z_\alpha - \left(\frac{h}{nT}\right)^{1/2}f_0(z_\alpha). \tag{4.7}$$

From Corollary 2 we deduce

Corollary 3. *Let model (1.1) and Assumptions 1–6 hold. Under H_0 in (4.1), as $n \to \infty$*

$$u_\alpha = z_\alpha + O\left(\left(\frac{h}{nT}\right)^{1/2}\right) \tag{4.8}$$
$$= s_\alpha + o\left(\left(\frac{h}{nT}\right)^{1/2}\right). \tag{4.9}$$

Thus, the test that rejects (4.1) against (4.2) when

$$\left(\frac{nT}{h}\right)^{1/2}a^{1/2}\hat{\lambda} > s_\alpha \tag{4.10}$$

is more accurate than the standard

$$\left(\frac{nT}{h}\right)^{1/2}a^{1/2}\hat{\lambda} > z_\alpha \tag{4.11}$$

implied by first-order asymptotic theory.

Rather than correcting critical values, we can construct a transformation such that the cdf of the transformed statistic is closer to the standard normal than that of $(nT/h)^{1/2}a^{1/2}\hat{\lambda}$ (e.g. Yanagihara and Yuan, 2005). Define

$$F(\zeta) = \zeta + \left(\frac{h}{nT}\right)^{1/2} f_0(\zeta) + \left(\frac{h}{3nTa}\right)^3 (T-1)^2 \times (tr(2W^3 + 3W^2W'))^2\zeta^3. \quad (4.12)$$

Since

$$\frac{dF(\zeta)}{d\zeta} = \left(1 + \frac{1}{3}\left(\frac{h}{nTa}\right)(T-1)tr(2W^3 + 3W^2W')\zeta\right)^2 > 0, \quad (4.13)$$

the transformation is monotonic, and we have

Corollary 4. *Let model (1.1) and Assumptions 1–6 hold. Under H_0 in (4.1), as $n \to \infty$*

$$P\left(F\left(\left(\frac{nT}{h}\right)^{1/2} a^{1/2}\hat{\lambda}\right) \le z_\alpha\right) = 1 - \alpha + o\left(\left(\frac{h}{nT}\right)^{1/2}\right).$$

Hence, the test that rejects (4.1) against (4.2) when

$$F((nT/h)^{1/2}a^{1/2}\hat{\lambda}) > z_\alpha \quad (4.14)$$

is expected to be more accurate than (4.11).

As with our corrected interval estimates, our corrected tests involve closed form functions of $\hat{\lambda}$ and given quantities, and can be rapidly computed.

5. Monte Carlo study of finite-sample performance

We report a small Monte Carlo exercise to investigate the finite sample performance of our Edgeworth-corrected cdf, confidence intervals and tests. For each of 1000 replications $\epsilon_i, i = 1, \ldots, nT$, are independently generated from $\mathcal{N}(0, 1)$, i.e. according to Assumption 1 with $\sigma^2 = 1$, and each component of the $(n \times 1)$ vector c_i is independently generated from a uniform distribution with support $[-1, 1]$. We choose a circulant structure for W, i.e.

$$W = \frac{1}{\|\Psi\|}\Psi, \quad (5.1)$$

where

$$\Psi = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 0 & \ldots & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & \ldots & 0 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 0\ldots & 0 & 1 & 1 & 1 \\ \ldots & & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & 1 & 1 & 1 & 1 & 0 & \ldots & 0 & 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}. \quad (5.2)$$

With the latter specification for W, $h = \|\Psi\|$ and is fixed as n increases.

Figs. 1–3 display the plots of the standard normal cdf against the (simulated) exact cdf of $(nT/h)^{1/2}a^{1/2}(\hat{\lambda} - \lambda_0)$, along with our Edgeworth-corrected cdf, respectively indicated in the figures as "normal", "exact", and "Edgeworth", where the latter is computed according to (2.15) for $x = a^{-1/2}\zeta$, i.e.

$$\Phi(\zeta) + \left(\frac{h}{nT}\right)^{1/2} f(\zeta)\phi(\zeta), \quad (5.3)$$

and $\lambda_0 = -0.9, 0, 0.9$. For this very small sample, $(n, T) = (12, 3)$, "Edgeworth" appears to be a very good approximation of the "exact" cdf for all values of λ_0 considered, while the standard normal does not offer a satisfactory approximation even for $\lambda_0 = 0$.

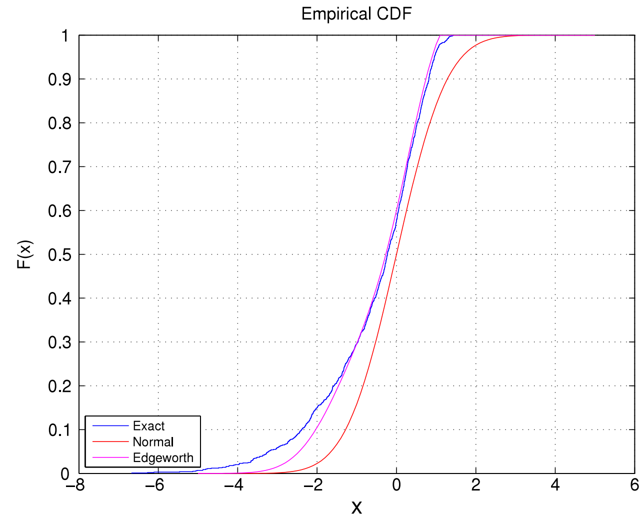

Fig. 1. Plots of the standard normal cdf and the exact and Edgeworth-corrected cdfs of $(nT/h)^{1/2}a^{1/2}(\hat{\lambda} - \lambda_0)$ for $\lambda_0 = 0.9$. $(n, T) = (12, 3)$.

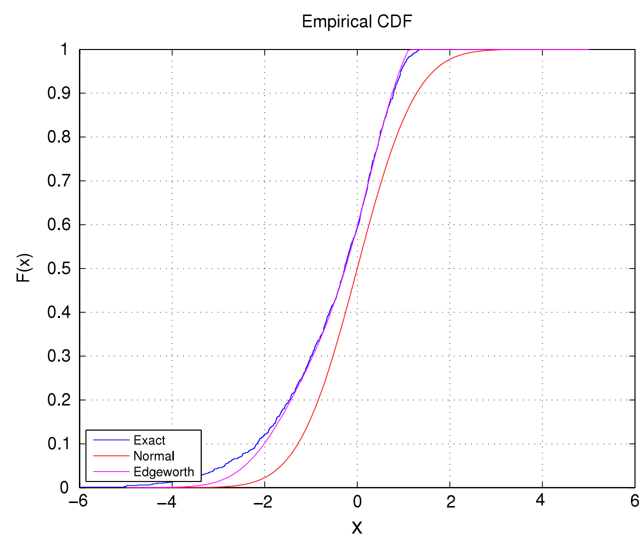

Fig. 2. Plots of the standard normal cdf and the exact and Edgeworth-corrected cdfs of $(nT/h)^{1/2}a^{1/2}(\hat{\lambda} - \lambda_0)$ for $\lambda_0 = 0$. $(n, T) = (12, 3)$.

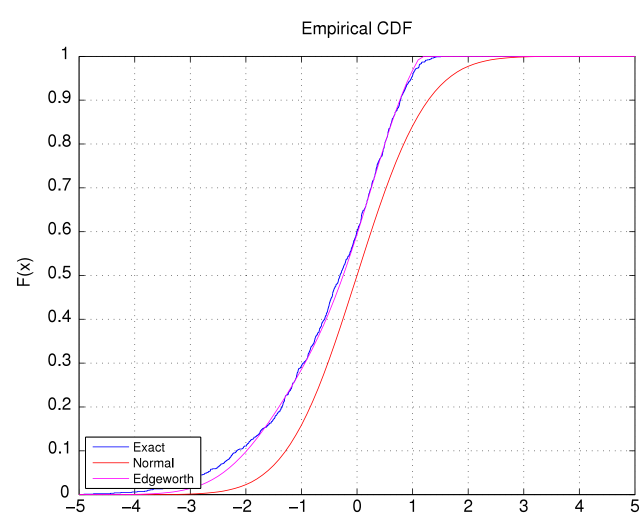

Fig. 3. Plots of the standard normal cdf and the exact and Edgeworth-corrected cdfs of $(nT/h)^{1/2}a^{1/2}(\hat{\lambda} - \lambda_0)$ for $\lambda_0 = -0.9$. $(n, T) = (12, 3)$.

Table 1
Empirical coverage probabilities of $I^{\mathcal{N}}$ and $\hat{I}^{Ed}$ in (3.7) and (3.9). $T = 3, \alpha = 5\%$.

n	:	12	15	20	40
I^N	λ_0				
	−0.5	0.980	0.972	0.964	0.965
	0	0.921	0.930	0.931	0.931
	0.5	0.907	0.918	0.929	0.923
	0.9	0.913	0.923	0.924	0.932
$\hat{I}^{Ed}$	λ_0				
	−0.5	0.952	0.952	0.948	0.947
	0	0.954	0.948	0.942	0.949
	0.5	0.956	0.951	0.948	0.953
	0.9	0.963	0.939	0.949	0.942

Table 2
Empirical sizes of one-sided tests of (4.1). $T = 3, \alpha = 5\%$.

n	12	15	20	40
A	0	0	0.005	0.011
ECV	0.062	0.046	0.048	0.046
ET	0.021	0.028	0.038	0.041

Table 3
Empirical powers of one-sided tests of (4.1) against (5.4) when $\bar{\lambda} = 0.1, 0.5$. $T = 3$. $\alpha = 5\%$.

n	$\bar{\lambda}$	12	15	20	40
A	0.1	0	0	0.005	0.045
	0.5	0.070	0.119	0.292	0.644
ECV	0.1	0.138	0.150	0.148	0.174
	0.5	0.594	0.601	0.626	0.805
ET	0.1	0.064	0.061	0.081	0.119
	0.5	0.412	0.440	0.531	0.778

For $\lambda_0 = -0.5, 0, 0.5, 0.9$, Table 1 compares the empirical coverage probabilities of the confidence sets based on the standard normal approximation in (3.7) with those of the Edgeworth-corrected one in (3.9), respectively indicated as "N" and "E" in the text. Table 2 instead shows empirical sizes of one-sided tests of H_0 in (4.1) based on asymptotic critical values, Edgeworth-corrected critical values and Edgeworth-transformed statistics, respectively displayed in (4.11), (4.10) and (4.14) and abbreviated in tables and text as "A", "ECV" and "ET". Consistent with our theoretical results of Sections 2–4 we increase n and keep T fixed, i.e. we compute empirical coverage probabilities and sizes for $(n, T) = (12, 3), (15, 3), (20, 3), (40, 3)$. In both Tables $\alpha = 5\%$.

In Table 1 empirical coverage probabilities of N appear to exceed the nominal 95% for $\lambda_0 = -0.5$, and to be considerably below 95% for non-negative values of λ_0. On the other hand, empirical coverage probabilities of E are very close to 95% even for very small n. For example, when $\lambda_0 = 0.5$, on average across sample sizes values for E are about 90% closer to 0.95 than N, with similar improvements for other λ_0.

Finite-sample corrections seem to be even more necessary in testing. From Table 2, A is severely under-sized for all n. Both ECV and ET instead offer an improvement over A, ECV outperforming ET throughout. On average across n, empirical sizes of ECV and ET are respectively 88% and 62% closer to 0.05 than A.

Table 3 displays empirical powers of the non-size-corrected tests A, ECV and ET of H_0 against

$$H_1 : \lambda_0 = \bar{\lambda} > 0, \tag{5.4}$$

for $\bar{\lambda} = 0.1, 0.5$. For $\bar{\lambda} = 0.1$ A offers very low power for all sample sizes considered and is drastically outperformed by both ECV and ET, with ECV giving the best performance. For $\bar{\lambda} = 0.5$ all tests display good power properties (with the exceptions of A for very small sample sizes), with again ECV offering superior performance compared to A and ET.

Acknowledgment

The first author research was supported by ESRC Grant ES/J007242/1.

Appendix A. Proofs of theorems

Proof of Theorem 1. We begin by developing an expansion for $\hat{\lambda}-\lambda_0$, in terms of the objective function $l(\lambda)$ and its derivatives. We thence deduce an approximation to the cdf of $\hat{\lambda} - \lambda_0$, which we write as the cdf of a quadratic form in ϵ. After approximating the characteristic function of this quadratic form, we obtain the result by Fourier inversion.

For $i \geq 1$ let $\partial_i(\lambda) = \partial^i l(\lambda)/\partial\lambda^i$ where $l(\lambda)$ is defined in (2.1), and let $\partial_i = \partial_i(\lambda_0)$. Proceeding similarly to Taniguchi (1988), by the mean value theorem,

$$0 = \partial_1\left(\hat{\lambda}\right) = \partial_1 + \partial_2(\hat{\lambda} - \lambda_0) + \frac{1}{2}\partial_3(\hat{\lambda} - \lambda_0)^2 + \frac{1}{6}\partial_4\left(\bar{\lambda}\right)(\hat{\lambda} - \lambda_0)^3,$$

where $\bar{\lambda}$ is an intermediate point between $\hat{\lambda}$ and λ_0. Thus

$$\hat{\lambda} - \lambda_0 = (E(\partial_2))^{-1}\left(\partial_1 + (\partial_2 - E\partial_2)(\hat{\lambda} - \lambda_0) + \frac{1}{2}\partial_3(\hat{\lambda} - \lambda_0)^2 + \frac{1}{6}\partial_4\left(\bar{\lambda}\right)(\hat{\lambda} - \lambda_0)^3\right).$$

Defining

$$z_1 = \left(\frac{h}{nT}\right)^{1/2} E\partial_1, \qquad z_2 = \left(\frac{h}{nT}\right)^{1/2}(\partial_2 - E\partial_2),$$
$$z_3 = \left(\frac{h}{nT}\right)^{1/2}(\partial_3 - E(\partial_3)),$$
$$k = -\frac{h}{nT}E(\partial_2), \qquad j = \frac{h}{nT}E(\partial_3), \tag{A.1}$$

gives

$$\left(\frac{nT}{h}\right)^{1/2}(\hat{\lambda} - \lambda_0) = \frac{z_1}{k} + \frac{z_2}{k}(\hat{\lambda} - \lambda_0) + \frac{1}{2}\left(\frac{nT}{h}\right)^{1/2}\frac{j}{k}(\hat{\lambda} - \lambda_0)^2 + \frac{1}{2}\frac{z_3}{k}(\hat{\lambda} - \lambda_0)^2 + \frac{1}{6k}\left(\frac{h}{nT}\right)^{1/2}\partial_4\left(\bar{\lambda}\right)(\hat{\lambda} - \lambda_0)^3. \tag{A.2}$$

To investigate the quantities defined in (A.1), we introduce the notation

$$m(D) = \sum_{t=1}^{T}\tilde{Y}_t' D\tilde{Y}_t,$$

whence it is straightforward to show from (2.1) that

$$\partial_1(\lambda) = nT\frac{m\left(S(\lambda)'W\right)}{m(S(\lambda)'S(\lambda))} - Ttr(G(\lambda)),$$

$$\partial_2(\lambda) = -nT\frac{m\left(W'W\right)}{m(S(\lambda)'S(\lambda))} + 2nT\frac{m\left(S(\lambda)'W\right)^2}{m(S(\lambda)'S(\lambda))^2} - Ttr(G(\lambda)^2),$$

$$\partial_3(\lambda) = -6nT\frac{m\left(W'W\right)m\left(S(\lambda)'W\right)}{m(S(\lambda)'S(\lambda))^2} + 8nT\frac{m\left(S(\lambda)'W\right)^3}{m(S(\lambda)'S(\lambda))^3} - 2Ttr(G(\lambda)^3)$$

and

$$\partial_4(\lambda) = 6nT\frac{m\left(W'W\right)^2}{m\left(S(\lambda)'S(\lambda)\right)^2} - 36nT\frac{m\left(W'W\right)m\left(S(\lambda)'W\right)^2}{m\left(S(\lambda)'S(\lambda)\right)^3} + 48nT\frac{m\left(S(\lambda)'W\right)^4}{m\left(S(\lambda)'S(\lambda)\right)^4} - 6Ttr(G(\lambda)^4).$$

First, using (2.5)–(2.7) and results on moments of ratios of normal quadratic forms, given Assumption 1, and noting from (2.7) that $m(D) = r\left(S^{-1'}DS^{-1}\right), r(I_n) = \epsilon'\epsilon$,

$$\begin{aligned}
k &= h\frac{r\left(G'G\right)}{r(I_n)} - \frac{h}{2}\frac{r\left(G+G'\right)^2}{r(I_n)^2} + \frac{h}{n}tr(G^2)\\
&= \frac{h}{n}tr(G^2+G'G) - \frac{2h}{n^2}(tr(G))^2\left(1+\frac{2}{n(T-1)}\right)^{-1}\\
&\quad - \frac{2h}{n^2(T-1)}\left(tr(G^2+GG')\right)\left(1+\frac{2}{n(T-1)}\right)^{-1}\\
&= \frac{T}{T-1}a + O\left(\frac{1}{n(T-1)}\right),
\end{aligned} \tag{A.3}$$

which is finite and positive for sufficiently large n under Assumption 6. The first equality in (A.3) follows since both the ratios $r\left(G'G\right)/r(I_n)$ and $r\left(G+G'\right)/r(I_n)$ are independent of their own denominators and therefore have expectations equal to the ratio of the expectations (Pitman, 1937). Such properties are repeatedly used in the sequel, in particular we have

$$\begin{aligned}
j &= -hE\left(\frac{3r\left(G'G\right)r\left(G+G'\right)}{r(I_n)^2}\right) - hE\left(\frac{4r\left(G+G'\right)^3}{r(I_n)^3}\right)\\
&\quad + 2\frac{h}{n}tr(G^3)E\\
&= -h\frac{3E\left(r\left(G'G\right)r\left(G+G'\right)\right)}{E\left(r(I_n)^2\right)} - h\frac{4E\left(r\left(G+G'\right)^3\right)}{E\left(r(I_n)^3\right)} + 2\frac{h}{n}tr(G^3)\\
&= O(1),
\end{aligned}$$

since, as $n \to \infty$, the first and second terms are respectively $O(1/h)$ and $O(1/h^2)$, while $htr(G^3)/n = O(1)$. Also, under Assumptions 1, 3, 4 and 6, $z_1 = O_e(1), z_2 = O_p(1)$ and $z_3 = O_p(1/h)$, as shown in Lemmas 1–3. Therefore as $n \to \infty$ the first term on the RHS of (A.2) is $O_e(1)$, where $O_e(.)$ denotes exact rate in probability.

To deal with the remainder term

$$\frac{1}{6k}\left(\frac{h}{nT}\right)^{1/2}\partial_4\left(\bar{\lambda}\right)(\hat{\lambda}-\lambda_0)^3$$

in (A.2), note that as indicated in Section 2, $\hat{\lambda}$ is consistent for λ_0. Thus with probability approaching 1 as $n \to \infty$, $\left|\bar{\lambda}-\lambda_0\right| \leq \left|\hat{\lambda}-\lambda_0\right| < \varepsilon$ for any $\varepsilon > 0$. Considering the denominators in $\partial_4\left(\bar{\lambda}\right)$, note that

$$\begin{aligned}
&(n(T-1))\, m\left(S(\bar{\lambda})'S(\bar{\lambda})\right)\\
&\quad \geq m\left(S'S\right) - \left|m\left(S(\bar{\lambda})'S(\bar{\lambda})\right) - m\left(S'S\right)\right|.
\end{aligned}$$

Now

$$\begin{aligned}
m\left(S'S\right) &= r(I_n) \geq E(r(I_n)) - |r(I_n) - E(r(I_n))|\\
&= \sigma_0^2 n(T-1) + O_p((n(T-1))^{1/2}),
\end{aligned}$$

whereas

$$\begin{aligned}
&m\left(S(\bar{\lambda})'S(\bar{\lambda})\right) - m\left(S'S\right) = r\left(S^{-1'}S(\bar{\lambda})'S(\bar{\lambda})S^{-1}\right) - r(I_n)\\
&= \left(\bar{\lambda}-\lambda_0\right)^2 r\left(G'G\right) - \left(\bar{\lambda}-\lambda_0\right) r\left(G+G'\right)\\
&= \varepsilon^2 O_p(nT) + \varepsilon O_p\left(nTh^{-1}\right)\\
&= O_p(\varepsilon nT),
\end{aligned}$$

whence it follows from arbitrariness of ε that

$$(n(T-1))^{-1} m\left(S(\bar{\lambda})'S(\bar{\lambda})\right) \geq \sigma_0^2 - O_p(\varepsilon) \geq \sigma_0^2/2 - o_p(1).$$

In view of these calculations it can also be seen that the numerators in $\partial_4\left(\bar{\lambda}\right)$ are $O_p\left(h^{-2}nT\right)$, while $tr(G(\bar{\lambda})^4) = O_p(n/h)$. Thus

$$\begin{aligned}
\frac{1}{6k}\left(\frac{h}{nT}\right)^{1/2}\partial_4\left(\bar{\lambda}\right)(\hat{\lambda}-\lambda_0)^3 &= O_p\left(\left(\frac{h}{nT}\right)^{1/2} nTh^{-2}\left|\hat{\lambda}-\lambda_0\right|^3\right)\\
&= O_p\left(\left(\frac{h}{nT}\right)^{1/2} nTh^{-2}\left(\frac{h}{nT}\right)^{3/2}\right)\\
&= O_p\left((nT)^{-1}\right),
\end{aligned}$$

using the fact that, as noted in Section 2, under our conditions $\hat{\lambda}-\lambda_0 = O_p\left((h/(nT))^{1/2}\right)$. The last fact also implies that (A.2) gives, more precisely, $\hat{\lambda}-\lambda_0 = (h/nT)^{1/2}\left(z_1/k + o_p(1)\right)$. Substituting $(h/nT)^{1/2}\left(z_1/k + o_p(1)\right)$ for $\hat{\lambda}-\lambda_0$ on the RHS of (A.2) gives

$$\left(\frac{nT}{h}\right)^{1/2}(\hat{\lambda}-\lambda_0) = \frac{z_1}{k} + \left(\frac{h}{nT}\right)^{1/2}\left(\frac{z_2z_1}{k^2} + \frac{1}{2}\frac{jz_1^2}{k^3}\right) + o_p\left(\left(\frac{h}{nT}\right)^{1/2}\right).$$

We deduce that for any real x,

$$\begin{aligned}
&P\left(\left(\frac{nT}{h}\right)^{1/2}(\hat{\lambda}-\lambda_0) \leq x\right)\\
&= P\left(\frac{z_1}{k} + \left(\frac{h}{nT}\right)^{1/2}\left(\frac{z_2z_1}{k^2} + \frac{1}{2}\frac{jz_1^2}{k^3}\right) + o_p\left(\left(\frac{h}{nT}\right)^{1/2}\right) \leq x\right)\\
&= P\Bigg(\left(\frac{h}{nT}\right)^{1/2} r(A) + \left(\frac{h}{nT}\right)^{1/2}\\
&\quad \times \left(\frac{z_2z_1}{k} + \frac{1}{2}\frac{jz_1^2}{k^2}\right)\frac{r(I_n)}{nT} - x\frac{kq(I_n)}{nT} \leq 0\Bigg) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)\\
&= P\left(\epsilon'C\epsilon + q \leq 0\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right),
\end{aligned}$$

where the second equality is obtained by substituting for z_1 and rearrangement,

$$C = \frac{1}{2}\left(\frac{h}{nT}\right)^{1/2}\left(I_{T-1}\otimes(A+A')\right) - x\frac{k}{nT}I_{n(T-1)},$$

with A defined in (2.10) and (2.12), and

$$q = \left(\frac{h}{nT}\right)^{1/2}\left(\frac{z_2z_1}{k} + \frac{1}{2}\frac{jz_1^2}{k^2}\right)\frac{r(I_n)}{nT}.$$

We approximate the characteristic function of $\epsilon' C\epsilon + q$ by $1+\psi$, where

$$\psi = itE(\epsilon' C\epsilon + q) + \frac{1}{2}(it)^2 E((\epsilon' C\epsilon + q)^2) + \frac{1}{6}(it)^3 E((\epsilon' C\epsilon + q)^3),$$

and thus approximate its cumulant generating function by

$$\log(1+\psi) = \sum_{s=1}^{\infty}(-1)^{s+1}\frac{\psi^s}{s}.$$

Let κ_s be the sth cumulant of ψ. To calculate the κ_s note that q involves ratios of quadratic forms $r(.)$ in ϵ, in particular $q = (h/nT)^{1/2}(q_1+q_2+q_3+q_4)$, with

$$q_1 = -\frac{h}{k}\frac{r(A)\,r(B)}{r(I_n)}, \qquad q_2 = \frac{2h}{k}\frac{r(A)\,r(B)^2}{r(I_n)^2},$$

$$q_3 = -\frac{2h}{k}\left(\frac{(T-1)(trG)^2 + tr(G'G)}{(T-1)n^2+2n}\right) r(A),$$

$$q_4 = \frac{jh}{2k^2}\frac{r(A)^2}{r(I_n)},$$

where

$$B = G'G - \frac{tr(G'G)}{n}I_n.$$

We deduce that

$$\kappa_1 = E(\epsilon' C\epsilon + q) = -\frac{T-1}{T}\sigma_0^2 ax + \left(\frac{h}{nT}\right)^{1/2}\frac{\sigma_0^2 h}{an} \times \left(\frac{trGtr(4G^2+3G'G)}{n} - \frac{4(trG)^3}{n^2} - tr(G^3+2G^2G')\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right), \quad (A.4)$$

$$\kappa_2 = -E(r(I_n))^2 + E(r(I_n)^2) + 2E(q\epsilon' C\epsilon) - 2E(\epsilon' C\epsilon)E(q) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right) = \sigma_0^4\frac{(T-1)}{T}a + o\left(\left(\frac{h}{nT}\right)^{1/2}\right) \quad (A.5)$$

and

$$\kappa_3 = 2\left(E\left(\epsilon' C\epsilon\right)\right)^3 + E\left(\left(\epsilon' C\epsilon\right)^3\right) + 3E(((\epsilon' C\epsilon)^2 q) - 3E(\epsilon' C\epsilon)E((\epsilon' C\epsilon)^2)) - 6E(\epsilon' C\epsilon)E(\epsilon' C\epsilon q) - 3E(q)E\left(\left(\epsilon' C\epsilon\right)^2\right) + 6\left(E\left(\epsilon' C\epsilon\right)\right)^2 E(q) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$
$$= 8\sigma_0^6 tr(C^3) + 3\sigma_0^4 E(q)((trC)^2 - 2tr(C^2)) + 3E\left(q\left(\epsilon' C\epsilon\right)^2\right) - 6E(\epsilon' C\epsilon)E(\epsilon' C\epsilon q) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$
$$= -2\sigma_0^6\left(\frac{h}{nT}\right)^{1/2}\frac{h(T-1)}{nT}\left(tr(2G^3+3G'^2) - \frac{3trG(tr(2G^2+G'G))}{n} + \frac{4(trG)^3}{n^2}\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right). \quad (A.6)$$

The cumulant generating function of the standardized version of $\epsilon' C\epsilon + q$, i.e. $(\epsilon' C\epsilon + q - \kappa_1)/\kappa_2^{1/2}$, can be written as

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

where $\kappa_s^c = \kappa_s/\kappa_2^{s/2}$/ Thus the characteristic function of $\epsilon' C\epsilon + q$ is

$$e^{-\frac{1}{2}t^2}\exp\left(\sum_{s=3}^{\infty}\frac{\kappa_s^c (it)^s}{s!}\right) = e^{-\frac{1}{2}t^2}\left(1 + \sum_{s=3}^{\infty}\frac{\kappa_s^c (it)^s}{s!} + \frac{1}{2!}\left(\sum_{s=3}^{\infty}\frac{\kappa_s^c (it)^s}{s!}\right)^2 + \frac{1}{3!}\left(\sum_{s=3}^{\infty}\frac{\kappa_s^c (it)^s}{s!}\right)^3 + \cdots\right)$$
$$= e^{-\frac{1}{2}t^2}\left(1 + \frac{\kappa_3^c (it)^3}{3!} + \frac{\kappa_4^c (it)^4}{4!} + \frac{\kappa_5^c (it)^5}{5!} + \left(\frac{\kappa_6^c}{6!} + \frac{(\kappa_3^c)^2}{(3!)^2}\right)(it)^6 + \cdots\right).$$

Thus, by Assumption 1 and Fourier inversion,

$$P\left(\frac{\epsilon' C\epsilon + q - \kappa_1}{\kappa_2^{1/2}} \le z\right) = \int_{-\infty}^{z}\phi(z)dz + \frac{\kappa_3^c}{3!}\int_{-\infty}^{z}H_3(z)\phi(z)dz + \frac{\kappa_4^c}{4!}\int_{-\infty}^{z}H_4(z)\phi(z)dz + \cdots,$$

where $H_j(.)$ is the jth Hermite polynomial. Collecting the results derived above,

$$P\left(\left(\frac{nT}{h}\right)^{1/2}(\hat{\lambda} - \lambda_0) \le x\right) = P\left(\epsilon' C\epsilon + q \le 0\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right) = P\left(\frac{\epsilon' C\epsilon + q - \kappa_1}{\kappa_2^{1/2}} \le -\kappa_1^c\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$
$$= \Phi(-\kappa_1^c) - \frac{\kappa_3^c}{3!}\Phi^{(3)}(-\kappa_1^c) + \frac{\kappa_4^c}{4!}\Phi^{(4)}(-\kappa_1^c) + \cdots, \quad (A.7)$$

where $\Phi^{(i)}$ denotes the ith derivative of Φ.

Now from (A.4) and (A.5),

$$\kappa_1^c = -a^{1/2}x + \left(\frac{h}{nT}\right)^{1/2} a^{-3/2}\frac{h(T-1)}{nT} \times \left(\frac{trGtr(4G^2+3G'G)}{n} - \frac{4(trG)^3}{n^2} - tr(G^3+2G^2G')\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right), \quad (A.8)$$

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

$$\kappa_3^c = -2\left(\frac{h}{nT}\right)^{1/2} a^{-3/2}\frac{h(T-1)}{nT} \times \left(tr(2G^3+3G^2G') - \frac{3trGtr(2G^2+G'G)}{n} + \frac{4(trG)^3}{n^2}\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right), \quad (A.9)$$

where a is defined in (2.11) and (2.12). By Taylor expansion of $\Phi(-\kappa_1^c)$ in (A.7) and using $\Phi^{(3)}(u) = u^2 - 1$,

$$P\left(\left(\frac{nT}{h}\right)^{1/2}(\hat{\lambda}-\lambda_0) \le x\right) = \Phi\left(a^{1/2}x\right) + \left(\frac{h}{nT}\right)^{1/2} \times a^{-3/2}\frac{h(T-1)}{3nT}\left(\frac{8(trG)^3}{n^2} - \frac{6trGtr(G^2+G'G)}{n} + tr(G^3+3G^2G')\right)\phi(a^{1/2}x) + \left(\frac{h}{nT}\right)^{1/2} a^{-1/2}\frac{h(T-1)}{3nT} \times \left(tr(2G^3+3G^2G') - \frac{3trGtr(2G^2+G'G)}{n} + \frac{4(trG)^3}{n^2}\right) \times x^2\phi(a^{1/2}x) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right), \quad (A.10)$$

whence the result follows from (2.13).

Proof of Theorem 2. We begin by developing an approximation to the cdf of a data-free scaling of $\hat{\lambda}-\lambda_0$, similar to that considered in Theorem 1, and an approximation to its probability density function. After thence obtaining a Taylor approximation to $\widehat{a}^{1/2}$ we approximate the characteristic function of our studentized statistic and complete the proof by Fourier inversion.

Define

$$U = \left(\frac{nT}{h}\right)^{1/2} a^{1/2}(\hat{\lambda}-\lambda_0),$$

$$u_1 = \frac{h(T-1)}{nTa^{3/2}}\left(\frac{trGtr(4G^2+3G'G)}{n} - \frac{4(trG)^3}{n^2} - tr(G^3+2G^2G')\right), \quad (A.11)$$

$$u_2 = \frac{h(T-1)}{3nTa^{3/2}}\left(tr(2G^3+3G^2G') - \frac{3trGtr(2G^2+G'G)}{n} + \frac{4(trG)^3}{n^2}\right), \quad (A.12)$$

so that for $x = a^{-1/2}\zeta$ with ζ being any real number, from (A.7)–(A.9) and after a Taylor expansion of $\Phi(-\kappa_1^c)$ and $\Phi^{(3)}(-\kappa_1^c)$,

$$P(U \le \zeta) = \Phi(\zeta) - \left(\frac{h}{nT}\right)^{1/2} u_1\phi(\zeta) + \left(\frac{h}{nT}\right)^{1/2} u_2\Phi^{(3)}(\zeta) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right). \quad (A.13)$$

From (A.13) we write the probability density function of U, denoted pdf_U, as

$$pdf_U(\zeta) = \phi(\zeta) - \left(\frac{h}{nT}\right)^{1/2} u_1\Phi^{(2)}(\zeta) + \left(\frac{h}{nT}\right)^{1/2} u_2\Phi^{(4)}(\zeta)$$
$$= \phi(\zeta) + \left(\frac{h}{nT}\right)^{1/2}\zeta(u_1+3u_2)\phi(\zeta) - \left(\frac{h}{nT}\right)^{1/2} u_2\zeta^3\phi(\zeta) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right),$$

where the last equality follows since $\Phi^{(2)}(\zeta) = -\zeta\phi(\zeta)$ and $\Phi^{(4)}(\zeta) = -(\zeta^3 - 3\zeta)\phi(\zeta)$.

Expanding $\widehat{a}^{1/2}$ around λ_0,

$$\left(\frac{nT}{h}\right)^{1/2}\widehat{a}^{1/2}(\hat{\lambda}-\lambda_0) = U + \left(\frac{h}{nT}\right)^{1/2} a^{-3/2}dU^2 + o_p\left(\left(\frac{h}{nT}\right)^{1/2}\right),$$

where d is defined in (3.2) and $d = O(1)$ as $n \to \infty$, so that the characteristic function of the LHS can be expanded as follows:

$$E\left(\exp\left(it\left(U + \left(\frac{h}{nT}\right)^{1/2} a^{-3/2}dU^2 + o_p\left(\left(\frac{h}{nT}\right)^{1/2}\right)\right)\right)\right)$$
$$= \frac{1}{(2\pi)^{1/2}}\int_{\Re}\left(e^{it\xi}\left(1 + it\left(\frac{h}{nT}\right)^{1/2} a^{-3/2}d\xi^2\right)e^{-\xi^2/2} \times \left(1 + \left(\frac{h}{nT}\right)^{1/2}\left((u_1+3u_2)\xi - u_2\xi^3\right)\right)\right)d\xi + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$
$$= \frac{e^{-t^2/2}}{(2\pi)^{1/2}}\int_{\Re}\left(e^{-(\xi-it)^2/2}\left(1 + it\left(\frac{h}{nT}\right)^{1/2} a^{-3/2}d\xi^2\right) \times \left(1 + \left(\frac{h}{nT}\right)^{1/2}\left((u_1+3u_2)\xi - u_2\xi^3\right)\right)\right)d\xi + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$
$$= e^{-t^2/2}\left(1 + it\left(\frac{h}{nT}\right)^{1/2} a^{-3/2}dE(X^2) + \left(\frac{h}{nT}\right)^{1/2}(u_1+3u_2)E(X) - u_2E(X^3)\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right), \quad (A.14)$$

where X is a complex normal variate with mean it and unit variance. Thus, by the same results on moments of normal variates as before, and by rearranging terms, (A.14) becomes

$$e^{-t^2/2}\left(1 + \left(\frac{h}{nT}\right)^{1/2} it(a^{-3/2}d + u_1 + 3u_2 - 3u_2) + \left(\frac{h}{nT}\right)^{1/2}(it)^3(a^{-3/2}d - u_2)\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$
$$= e^{-t^2/2}\left(1 + \left(\frac{h}{nT}\right)^{1/2} it(a^{-3/2}d + u_1) + \left(\frac{h}{nT}\right)^{1/2}(it)^3(a^{-3/2}d - u_2)\right) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right).$$

By Fourier inversion, formally,

$$P\left(\left(\frac{nT}{h}\right)^{1/2}\widehat{a}^{1/2}(\hat{\lambda}-\lambda_0) \le \zeta\right)$$
$$= \Phi(\zeta) - \left(\frac{h}{nT}\right)^{1/2}(a^{-3/2}d + u_1)\phi(\zeta) - \left(\frac{h}{nT}\right)^{1/2}(a^{-3/2}d - u_2)\Phi^{(3)}(\zeta) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$

$$= \Phi(\zeta) - \left(\frac{h}{nT}\right)^{1/2} (u_1 + u_2 + (a^{-3/2}d - u_2)\zeta^2)\phi(\zeta) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right)$$

$$= \Phi(\zeta) + \left(\frac{h}{nT}\right)^{1/2} \left(f(\zeta) - \frac{d}{\tilde{a}^{3/2}}\zeta^2\right)\phi(\zeta) + o\left(\left(\frac{h}{nT}\right)^{1/2}\right),$$

where the last equality follows by (2.13), (A.11) and (A.12) and rearrangement.

Appendix B. Technical lemmas

Lemma 1. *Under Assumptions 1, 3, 4 and 6, for $z_1 = O_e(1)$ as $n \to \infty$.*

Proof. We have

$$z_1 = \left(\frac{h}{nT}\right)^{1/2} \left(nT\frac{m(S'W)}{m(S'S)} - TtrG\right)$$

$$= (hnT)^{1/2} \left(nT\frac{r(S'W)}{r(S'S)} - TtrG\right)$$

$$= (hnT)^{1/2} \frac{r(A+A')}{2r(I_n)}.$$

Proceeding as before,

$$E\left({z_1}^2\right) = hnT\frac{\frac{1}{2}(T-1)tr((A+A')^2)}{n^2(T-1)^2 + 2n(T-1)} = \frac{T}{T-1}a + o(1),$$

which is finite and strictly positive in the limit under Assumption 6. Thus, by Markov's inequality, $z_1 = O_e(1)$ as $n \to \infty$.

Lemma 2. *Under Assumptions 1, 3 and 4, $z_2 = O_p(1)$ as $n \to \infty$.*

Proof. By standard algebra

$$z_2 = -(hnT)^{1/2} \left(\frac{r(G'G - tr(G'G)I_n/n)}{r(I_n)} - 2\left(\left(\frac{r(G)}{r(I_n)}\right)^2 - E\left(\frac{r(G)}{r(I_n)}\right)^2\right)\right).$$

By the c_r-inequality,

$$E(z_2^2) \le 2hnTE\left(\left(\frac{r(G'G - tr(G'G)I_n/n)}{r(I_n)}\right)^2 + 2\left(\left(\frac{r(G)}{r(I_n)}\right)^2 - E\left(\frac{r(G)}{r(I_n)}\right)^2\right)^2\right). \qquad (B.1)$$

Proceeding as before, the first term on the RHS of (B.1) is $O(1)$ as $n \to \infty$, since it equals

$$\frac{4h}{n}\frac{T}{T-1}\left(tr((G'G)^2) - \frac{(tr(G'G))^2}{n}\right)\left(1 + \frac{2}{n(T-1)}\right)^{-1}.$$

Similarly

$$hnTE\left(\left(\frac{r(G)}{r(I_n)}\right)^2 - E\left(\frac{r(G)}{r(I_n)}\right)^2\right)^2 = O\left(\frac{T}{(T-1)h}\right)$$

as $n \to \infty$. Thus from (B.1) and Markov's inequality, $z_2 = O_p(1)$ as $n \to \infty$. Note that though we are not attempting to provide an exact rate, we cannot omit the term $E(r(G)/r(I_n))^2$ from the bound (B.1) as this would neglect relevant terms.

Lemma 3. *Under Assumptions 1, 3 and 4, $z_3 = O_p(1/h)$ as $n \to \infty$*

Proof. By the c_r-inequality,

$$E(z_3^2) \le KhnT\left(E\left(\frac{r(G)r(G'G)}{r(I_n)^2} - E\left(\frac{r(G)r(G'G)}{r(I_n)^2}\right)\right)^2 + E\left(\frac{r(G)^3}{r(I_n)^3} - E\left(\frac{r(G)^3}{r(I_n)^3}\right)\right)^2\right). \qquad (B.2)$$

The estimation of the RHS is not reported here, but it can be shown that

$$E\left(\frac{r(G)r(G'G)}{r(I_n)^2}\right)^2 = \frac{(tr(G'G))^2(trG)^2}{n^4} + O\left(\frac{1}{n(T-1)h^3}\right)$$

and

$$\left(E\left(\frac{r(G)r(G'G)}{r(I_n)^2}\right)\right)^2 = \frac{(tr(G'G))^2(trG)^2}{n^4} + O\left(\frac{1}{n(T-1)h^3}\right),$$

proceeding as before, so the first term on the RHS of (B.2) is $O(1/h^2)$ as $n \to \infty$. Similarly,

$$E\left(\frac{r(G)^6}{r(I_n)^6}\right) = \frac{(trG)^6}{n^6} + O\left(\frac{1}{n(T-1)h^5}\right),$$

$$\left(E\left(\frac{r(G)^3}{r(I_n)^3}\right)\right)^2 = \frac{(trG)^6}{n^6} + O\left(\frac{1}{n(T-1)h^5}\right),$$

so the second term on the RHS of (B.2) is $O(1/h^4)$ as $n \to \infty$. Therefore, whether $h \to \infty$ or $h = O(1)$ as $n \to \infty$, $E(z_3)^2 = O(1/h^2)$ irrespective of whether $h \to \infty$ or $h = O(1)$, it follows that $z_3 = O_p(1/h)$.

References

Bao, Y., Ullah, A., 2007. Finite sample properties of maximum likelihood estimator in spatial models. J. Econometrics 137, 396–413.

Bhattacharya, R.N, Ghosh, J.K., 1978. On the validity of the formal Edgeworth expansion. Ann. Statist. 6, 434–451.

Case, A.C., 1991. Spatial patterns in household demand. Econometrica 59, 953–965.

Hall, P., 1992. The Bootstrap and Edgeworth Expansion. Springer-Verlag, New York.

Kelejian, H.H., Prucha, I.R., 1998. A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. J. Real Estate Finance Econ. 17, 99–121.

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

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

Lee, L.F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory 26, 187–230.

Lee, L.F., Yu, J., 2010. Estimation of spatial autoregressive panel data models with fixed effects. J. Econometrics 154, 165–185.

Pitman, E.G.J., 1937. Significance tests which may be applied to samples from any population. Suppl. J. R. Stat. Soc. 4, 119–130.

Robinson, P.M., Rossi, F., 2014a. Refined tests for spatial correlation. Econometric Theory 31, 1–32.

Robinson, P.M., Rossi, F., 2014b. Improved Lagrange multiplier tests in spatial autoregressions. Econom. J. 17, 139–154.

Taniguchi, M., 1988. Asymptotic expansion of the distributions of some test statistics for Gaussian ARMA processes. J. Multivariate Anal. 27, 494–511.

Yanagihara, H., Yuan, K., 2005. Four improved statistics for contrasting means by correcting skewness and kurtosis. Br. J. Math. Stat. Psychol. 58, 209–237.