

# A Note on Bartlett Correction Factor for Tests on Cointegrating Relations

Alessandra Canepa\*  
Brunel University London†

September 27, 2015

## Abstract

In this paper it is proposed to use a non-parametric bootstrap based Bartlett correction factor for the LR test for linear restrictions on the cointegrating vectors to reduce the finite sample size distortion problem of the test statistic.

**Keywords:** Cointegration, Bartlett correction factor, bootstrap method.

**JEL Classification:** C15, C22.

---

\*The author is grateful to Soren Johansen, Russell Davidson, James MacKinnon and the participants of several conferences for useful comments on a preliminary version of this paper. The author also would like to thank two anonymous referees for their excellent work.

†Email: Alessandra.Canepa@brunel.ac.uk. Department of Economics and Finance, Brunel University London, Uxbridge, Middlesex UB8 3PH, United Kingdom.

# 1 Introduction

Johansen's (2000) Bartlett corrected likelihood ratio ( $LR$ ) test for linear restrictions of cointegrating vectors relies on Gaussian innovations, but this strong assumption clashes with reality where there is rarely ground for presuming normality. There is therefore a need to examine small sample inference procedures that work well under weaker assumptions about the innovations.

In this paper it is proposed to use the bootstrap to approximate the finite sample expectation of the  $LR$  test in place of the analytical Bartlett correction. This procedure, which is in the spirit of Rocke (1989), involves calculating a number of bootstrap values for the quasi- $LR$  test statistic and estimating the expected value of the test statistic by the average value of the bootstrapped  $LR$  statistics. We believe that this non-parametric bootstrap based Bartlett correction can successfully be applied in the context under consideration to generate a test statistic that does not depend on the Gaussian assumption, the feature on which Johansen's analytical Bartlett correction crucially depends. If such an application were to be successful it would have significant practical implications, for several reasons. First, the non-parametric bootstrap does not require a choice of the error distribution, and this feature may be appealing to the applied researcher. Second, simulation results in Johansen (2000) indicate that the analytical correction factor is useful for some parameter values but does not work well for others. As Johansen points out "the influence of the parameters is crucial [...], there are parameters points close to the boundary where the order of integration or the number of cointegrating relations change where the correction does not work well" (cf. Johansen (2000) p.741). The dependency on the parameter values may be reduced by computing the Bartlett adjustment using the non-parametric bootstrap. Third, because the non-parametric bootstrap is used to approximate the first moment of a distribution, the resulting procedure is computationally less demanding than using the same method to approximate the tails of the distribution such as in the ordinary  $p$ -value bootstrap test.

The structure of the paper is as follows. Next section introduces the  $LR$  test for linear restrictions on cointegrated space, the Bartlett correction of Johansen (2000), the proposed bootstrap based inference procedure and the  $p$ -value bootstrap test. In Section 3, the consistency of the bootstrap based tests is considered. In Section 4, the design of the Monte Carlo experiment is explained, and the simulation results are reported. Finally, Section 5 contains some concluding remarks.

## 2 Model and Tests

Consider the  $p$ -dimensional  $VAR$  model

$$\Delta Y_t = \alpha (\beta' Y_{t-1} + \rho' D_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \phi d_t + \varepsilon_t, \quad t = 1, \dots, T \quad (1)$$

where  $Y_t$  and  $\varepsilon_t$  are  $(p \times 1)$  vectors with  $\varepsilon_t \sim N(0, \Omega)$ ;  $\Delta Y_t = Y_t - Y_{t-1}$ ;  $\alpha$  and  $\beta$  are  $(p \times r)$ ;  $\phi$  is  $(p \times p_d)$ ;  $\rho$  is  $(p_D \times r)$ ;  $\Gamma_1, \dots, \Gamma_{k-1}$  are  $(p \times p)$ ;  $d_t$  ( $p_d \times 1$ ) and  $D_t$  ( $p_D \times 1$ ) are deterministic terms. We focus on the hypothesis  $\mathcal{H}_0 : \beta = H\varphi$ , where  $H$  ( $p \times s$ ) (for  $r \leq s \leq p$ ) is a known matrix that specifies that the same restrictions are imposed on all cointegrating vectors ( $r$ ),  $s$  is the number of unrestricted parameters, and  $\varphi$  is a  $(s \times r)$  matrix; see Johansen

(1996) for a discussion of tests for other hypotheses. The test statistic for  $\mathcal{H}_0$  is given by

$$\Lambda = -T \sum_{i=1}^r \log \left[ \frac{(1 - \hat{\lambda}_i)}{(1 - \tilde{\lambda}_i)} \right], \quad (2)$$

where  $\hat{\lambda}_i$  and  $\tilde{\lambda}_i$  are the usual eigenvalues implied by the maximum likelihood estimation of the restricted and unrestricted models, respectively. The Bartlett adjustment is given by

$$\begin{aligned} \vartheta &= \frac{E_{\theta}(\Lambda)}{q} = 1 + \frac{1}{T} \left[ \frac{1}{2} (p + s - r + 1 + 2p_D) + p_d + kp \right] \\ &+ \frac{1}{Tr} [(2p + s - 3r - 1 + 2p_D) v(\alpha) + 2(c(\alpha) + c_d(\alpha))], \end{aligned} \quad (3)$$

where  $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\Omega})$ ,  $q = r(p - s)$ ,  $v(\alpha) = \text{tr} \left\{ (\alpha' \Omega^{-1} \alpha)^{-1} \sum_{\beta\beta}^{-1} \right\}$  with  $\sum_{\beta\beta} = \text{Var}(\beta' Y_t | \Delta Y_t, \dots, \Delta Y_{t-k+2})$ ,  $c_d(\alpha) = p_d v(\alpha)$ , and the constant  $c(\alpha)$  is given in Johansen (2000). Thus,  $\Lambda_B = \vartheta^{-1} \Lambda$  is the Bartlett corrected  $LR$  statistic.

The correction in (3) is derived under the assumption that the innovations are  $\varepsilon_t \sim N(0, \Omega)$ . Simulation results presented by Johansen (2000) suggest that applying this type of correction to the  $LR$  test statistic dramatically reduces the finite sample size distortion problem. However, the Bartlett correction factor is predicated under the assumption of Gaussian innovations. When the innovations are non-normal, the correction factor needs to be modified in order to account for skewness and kurtosis of the innovations. Because of the complicated formula of the  $LR$  statistic, deriving the asymptotic expansions needed to calculate the expectation of the test statistic can be demanding. One way of overcoming such calculations is to use a numerical approximation in place of the analytical Bartlett correction. By using the empirical distribution function in place of some specific parametric distribution, the non-parametric bootstrap does not require a choice of the error distribution, and this feature may be appealing to the applied researcher. Canepa and Godfrey (2007) propose computing the Bartlett adjustment for a quasi- $LR$  test using non-parametric bootstrapping as a simple method to generate a non-normality robust small sample inference procedure in the context of ARMA models. An alternative procedure is a straightforward application of the bootstrap  $p$ -value approach. The steps used to implement the two bootstrap based inference procedures are presented below.

### Algorithm 1: Bootstrap Bartlett Corrected Test

*Step (1):* Estimate the model in (2) and compute  $\Lambda$  and the restricted residuals

$$\hat{\varepsilon}_t = \Delta Y_t - \hat{\alpha} (\hat{\varphi}' H' Y_{t-1} + \hat{\rho}' D_t) - \sum_{i=1}^{k-1} \hat{\Gamma}_i \Delta Y_{t-i} - \hat{\phi} d_t.$$

*Step (2):* Resample the residuals from  $(\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_T)$  independently with replacement to obtain a bootstrap sample  $(\varepsilon_1^*, \dots, \varepsilon_T^*)$ . Generate the bootstrap sample

$$\Delta Y_t^* = \hat{\alpha} (\hat{\varphi}' H' Y_{t-1}^* + \hat{\rho}' D_t) + \sum_{i=1}^{k-1} \hat{\Gamma}_i \Delta Y_{t-i}^* + \hat{\phi} d_t + \varepsilon_t^*,$$

recursively from  $(\varepsilon_1^*, \dots, \varepsilon_T^*)$  using the estimated restricted model given in (2).

*Step (3):* Compute  $\Lambda_j^*$  using the data of Step (2) and repeat  $B$  times.

*Step (4):* To get the bootstrap-Bartlett correction factor, say  $\bar{\Lambda}^*$ , calculate  $\bar{\Lambda}^* = B^{-1} \sum_{j=1}^B \Lambda_j^*$ . A Bartlett-type corrected statistic is therefore

$$\Lambda_B^* = \frac{q\Lambda}{\bar{\Lambda}^*}.$$

The corrected statistic is then referred to a  $\chi^2(q)$  distribution (with  $q = r(p - s)$ ).

### Algorithm 2: Bootstrap p-value test

As far as the non-parametric bootstrap  $p$ -value test is concerned, the bootstrap algorithm adopted is similar to the procedure proposed by Gredenhoff and Jacobson (2001). This involves repeating Step (1)-(3) and then following Step (5) below.

*Step (5)* Compute the bootstrap  $p$ -value function of the observed value  $\Lambda$  by calculating

$$\hat{P}^*(\Lambda) = B^{-1} \sum_{j=1}^B I(\Lambda_j^* \geq \Lambda),$$

where  $I(\cdot)$  is an indicator function that equals one if the inequality is satisfied and zero otherwise. The bootstrap  $p$ -value test,  $\Lambda^*$ , is carried out by comparing  $\hat{P}^*(\Lambda)$  with the desired critical level,  $\gamma$ , and rejecting the null hypothesis if  $\hat{P}^*(\Lambda) \leq \gamma$ . Note that the subscript “\*” will be used to indicate the bootstrap analog throughout the paper.

## 3 Asymptotic Results

We now consider the asymptotic distribution of the bootstrap based procedures introduced in the previous section and we show that  $\Lambda_B^*$  and  $\Lambda^*$  converge weakly in probability to the  $\chi_q^2$  distribution. The proof relies heavily on the fact that the mean of the squared residuals converges in probability toward  $\Omega$  and that the estimators are consistent.

In the following  $\xrightarrow{w}$  denotes weak convergence,  $\xrightarrow{P}$  convergence in probability,  $\xrightarrow{w_p}$  weak convergence in probability as defined by Gine and Zinn (1990),  $P^*$  denotes the bootstrap probability and  $E^*$  relates to the expectation under  $P^*$ . Moreover, for any square matrix  $A$ ,  $|A|$  is used to indicate the determinant of  $A$ , the matrix  $A_\perp$  satisfies  $A'_\perp A = 0$ , and the norm  $\|A\|$  is  $\|A\| = [tr(A'A)]^{1/2}$ . For any vector  $a$ ,  $\|a\|$  denotes the Euclidean distance norm,  $\|a\| = (a'a)^{1/2}$ . Moreover, we make the following assumption:

#### Assumption 1:

*i)* Define the characteristic polynomial,

$$A(z) = (1 - z)I_p - \Pi z - \Gamma_1(1 - z)z - \dots - \Gamma_{k-1}(1 - z)z^{k-1},$$

where  $\Pi = \alpha\beta'$ . Assume that the roots of  $\det[A(z)] = 0$  are located outside the complex unit circle or at 1. Also assume that the matrices  $\alpha$  and  $\beta$  have full rank  $r$  and that  $\alpha'_\perp \Gamma \beta_\perp$  has full rank  $p - r$ , where  $\Gamma = I_p - \Gamma_1 - \dots - \Gamma_{k-1}$ .

ii) The innovations,  $\varepsilon_t$ , are independent and identically distributed with expectation 0 and covariance matrix  $\Omega$ .

Assumption 1 i) ensures that the observations are  $I(1)$  variables and rules out the possibility that some of the zeros of the estimated characteristic equation are outside the unit circle, in which case the bootstrap sample will become explosive. Under Assumption 1 i) Theorem 4.2 in Johansen (1996) holds, so that the process  $Y_t$  has the following representation

$$Y_t = C \sum_{i=1}^t (\varepsilon_i + \rho' D_i) + C(L)(\varepsilon_t + \alpha \rho' D_i + \phi d_t) + A_0,$$

where  $C = \beta_{\perp} \left( \alpha'_{\perp} \left( I - \sum_{i=1}^k \Gamma_i \right) \beta_{\perp} \right)^{-1} \alpha'_{\perp}$  and  $A_0$  is a term that depends only on the initial values and  $\beta' A_0 = 0$ .

Assumption 1 ii) rules out the possibility that innovations are serially correlated or conditionally heteroskedastic. Under Assumption 1 ii), the partial sum of  $\{\varepsilon_t\}$  satisfies the functional central limit theorem so that

$$T^{-1/2} \sum_{t=1}^{\lfloor Tu \rfloor} \varepsilon_t \xrightarrow{w} B(u), \quad u \in [0, 1],$$

$$T^{-1} \sum_{t=1}^T \left( \sum_{i=1}^{t-1} \varepsilon_i \right) \varepsilon_t' \xrightarrow{w} \int_0^1 B(dB)',$$

where  $B = \Omega^{1/2} W$  is a  $p$ -dimensional Brownian motion with variance  $\Omega$  and  $W$  a  $p$ -dimensional standard Brownian motion. These results can be used to derive the asymptotic properties of  $\Lambda_B^*$  and  $\Lambda^*$ .

Following the conventional notation, we define  $R_{0t}$  and  $R_{1t}$  as the residuals obtained by regressing  $\Delta Y_t$  and  $(Y_{t-1}, D_t)'$  respectively on lagged differences and  $d_t$ . Moreover,

$$S_{ij} = T^{-1} \sum_{t=1}^T R_{it} R_{jt}', \quad i, j = 0, 1.$$

**Proposition 1:** *Let the conditions of Assumption 1 hold. Then, under the null hypothesis,  $\Lambda^* \xrightarrow{w_P} \Lambda$  and  $\bar{\Lambda}^* \xrightarrow{P} E(\Lambda)$  as  $T \rightarrow \infty$ .*

**Proof :** Under Assumption 1, Lemma 1 in Swensen (2006) implies that the generated pseudo observations have the representation<sup>1</sup>

$$Y_t^* = C \sum_{i=1}^t (\varepsilon_i^* + \hat{\rho}' D_i) + C(L)(\varepsilon_t^* + \hat{\alpha} \hat{\rho}' D_i + \hat{\phi} d_t) + T^{1/2} R_t^*,$$

where for all  $\eta > 0$ ,  $P^*(\max_{t=1, \dots, T} \|R_t^*\| > \eta) \xrightarrow{P} 0$  as  $T \rightarrow \infty$ . Moreover, Lemmas S1 and S2 in Swensen (2006)

<sup>1</sup>Note due to the presence of a trend in Swensen (2006) the details of the derivation are slightly different. However, ergodicity for the sequence  $\{\varepsilon_t\}$  is sufficient for the Granger representation to hold. Also, Assumption 1 implies the existence of second order moments of  $\{\varepsilon_t\}$ . Therefore, the convergence of appropriate sums of stochastic integrals can be derived in a similar way.

imply that

$$T^{-1/2} \sum_{t=1}^{[Tu]} \varepsilon_t^* \xrightarrow{w_P} B(u), \quad (5)$$

$$T^{-1} \hat{\beta}'_{\perp} S_{11}^* \hat{\beta}_{\perp} \xrightarrow{w_P} \int_0^1 F(u) F(u)' du, \quad (6)$$

$$\hat{\beta}'_{\perp} (S_{10}^* - S_{11}^* \hat{\beta}_{\perp}) \alpha_{\perp} \xrightarrow{w_P} \int_0^1 F dB' \alpha_{\perp}, \quad (7)$$

where  $F(u) := \beta'_{\perp} CB(u)$  and

$$P^* (\|S_{00}^* - \Sigma_{00}\| > \eta) \xrightarrow{P} 0, \quad (8)$$

$$P^* (\|\hat{\beta}' S_{11}^* \hat{\beta} - \Sigma_{\beta\beta}\| > \eta) \xrightarrow{P} 0, \quad (9)$$

$$P^* (\|\hat{\beta}' S_{01}^* - \Sigma_{0\beta}\| > \eta) \xrightarrow{P} 0, \quad (10)$$

where  $\Sigma_{00}$ ,  $\Sigma_{\beta\beta}$  and  $\Sigma_{0\beta}$  are the probability limits of  $S_{00}$ ,  $\beta' S_{11} \beta$  and  $S_{01} \beta$ , respectively. In the bootstrap algorithm the pseudo-observations are generated under the null hypothesis using the restricted innovations. When linear restrictions are imposed on the parameters  $\hat{\beta} = H\hat{\varphi}$ , a submodel is defined and the space spanned by the linear transformation  $z : \mathbb{R}^p \mapsto \mathbb{R}^s$  with matrix representation  $Y_t^* \mapsto H'Y_t^*$  forms a subspace such that  $sp(\hat{\beta}) \subset sp(H)$ . Given that linear transformations preserve linear combinations of vectors it follows that if  $\{Y_t^*\}$  satisfies Lemma 1 in Swensen (2006), then  $\{H'Y_t^*\}$  also satisfies the same conditions. Moreover, the random process  $\left\{ T^{-1/2} \left( \sum_{t=1}^{[Tu]} H' \varepsilon_t^* \right) \right\}$ , where  $[Tu]$  is the integer value of  $Tu$ , converges weakly toward a Brownian motion with covariance matrix  $H'\Omega H$ . This implies that the asymptotic distributions of the moment matrices are given by

$$T^{-1/2} \hat{\varphi}'_{\perp} H' Y_{[Tu]}^* \xrightarrow{w_P} H' \tilde{F}(u), \quad (11)$$

$$\hat{\varphi}'_{\perp} H' S_{10}^* \hat{\alpha}_{\perp} \xrightarrow{w_P} H' \int_0^1 \tilde{F} dB' \alpha_{\perp}, \quad (12)$$

$$T^{-1} \hat{\varphi}'_{\perp} H' S_{11}^* H \hat{\varphi}_{\perp} \xrightarrow{w_P} H' \int_0^1 \tilde{F}(u) \tilde{F}(u)' du H. \quad (13)$$

where  $\tilde{F}(u) := \varphi'_{\perp} CB(u)$ . From (8)-(10) it follows that the  $(p-r)$  smallest solution of

$$\left| \hat{\lambda} \hat{\varphi}' (H' S_{11}^* H - H' S_{10}^* S_{00}^{*-1} S_{01}^* H) \hat{\varphi} \right| = 0,$$

converges to zero. Therefore, using (11)-(13) the asymptotic  $\chi_q^2$  distribution of  $\Lambda^*$  can be found by mimicking Theorem 13.9 in Johansen (1996).

Under Assumption 1, (5)-(6) imply weak convergence of the partial sums of stochastic integrals. Moreover, from (8)-(10) we have that  $S_{ij}^* \xrightarrow{P} \Sigma_{ij}$  and the estimators of the parameters are consistent. This trivially implies that  $E^*(\Lambda^*) \xrightarrow{P} E(\Lambda)$  in probability as  $T \rightarrow \infty$ . ■

Before concluding this section, a caveat regarding the use of the bootstrap Bartlett correction is discussed. The bootstrap procedure used for approximating the finite sample expectation of the  $LR$  test is a general tool and can be readily applied to other  $LR$  statistics. In his seminal article, Beran (1988) concluded that for *asymptotically* pivotal statistics (i.e., statistics for which the limiting distribution does not depend on unknown nuisance parameters), the  $p$ -value bootstrap test accomplishes the analytical Bartlett adjustment automatically, in the sense that the error in rejecting probability of the two tests is of order  $O(T^{-3/2})$ . Approximating the finite sample expectation of the  $LR$  test using the bootstrap involves substituting a  $\sqrt{T}$  consistent estimate of  $b(\theta)$  in (1), hence the resulting test statistic should provide accuracy of the same order. However, Beran (1988) also pointed out that for pivotal statistics the error committed by using the  $p$ -value bootstrap test is of order  $O(T^{-2})$ , thus providing a further refinement with respect to the Bartlett corrected test. It follows that for pivotal statistics the reduction in the computational burden enjoyed by the bootstrap Bartlett corrected test does not pay off, since the error committed by the  $p$ -value bootstrap test is lower by an extra  $O(T^{-1/2})$  factor.

In this paper only the consistency of bootstrap based tests is considered. Establishing the conditions which ensure asymptotic refinements will be the subject of future research. However, given that under the i.i.d. assumption the  $LR$  test considered in this work is only asymptotically pivotal, we may expect the two Bartlett corrected and  $p$ -value bootstrap tests to be of an order  $O(T^{-1})$  smaller than the corresponding asymptotic  $LR$  test, but the error in rejecting probability may not be  $O(T^{-2})$ .

## 4 The Monte Carlo experiment

The  $DGP$  adopted is similar to Haug (2002) and it is given by

$$\begin{aligned} \Delta Y_{1t} &= \epsilon_{1t}, \\ \Delta Y_{2t} &= \epsilon_{2t}, \\ Y_{3t} &= Y_{4t} + u_{3t}, \quad \text{where } u_{3t} = \xi u_{3t-1} + \epsilon_{3t}, \\ Y_{4t} &= -Y_{3t} + u_{4t}, \quad u_{4t} = u_{4t-1} + \epsilon_{4t}, \end{aligned} \tag{14}$$

with

$$\begin{bmatrix} \epsilon_{jt} \\ \epsilon_{it} \end{bmatrix} \sim i.i.d. N \left[ \begin{pmatrix} 0 \\ (2 \times 1) \\ 0 \\ (2 \times 1) \end{pmatrix}, \begin{pmatrix} A & 0 \\ (2 \times 2) & (2 \times 2) \\ 0 & B \\ (2 \times 2) & (2 \times 2) \end{pmatrix} \right],$$

where  $\epsilon_{jt} = [\epsilon_{1t} \ \epsilon_{2t}]'$ ,  $\epsilon_{it} = [\epsilon_{3t} \ \epsilon_{4t}]'$ ,  $A = \sigma^2 I$ , and  $B = \begin{pmatrix} \sigma^2 & \sigma\eta \\ \sigma\eta & \sigma^2 \end{pmatrix}$ . The null hypothesis of interest is

$$H_0 : \beta = H\varphi = \begin{bmatrix} 0 \\ (1 \times 3) \\ I \\ (3 \times 3) \end{bmatrix} \varphi, \quad (3 \times 1)$$

where  $I$  is an identity matrix.

The parameter space experimented with is  $T \in (50, 100, 250)$ ;  $\xi \in (0.2, 0.5, 0.8)$ ;  $\eta \in (-0.5, 0.5)$ ;  $\sigma = 1$ .

Standardized forms of the  $\chi^2(c)$  and the  $t(c)$  have been used to provide evidence of the impact of skewness and kurtosis on the empirical sizes of the test statistics under consideration. In particular, two sets of simulation experiments have been undertaken. In the first exercise, the effect of skewness has been investigated by systematically increasing  $c$  from an extremely skewed  $\chi^2(c)$  distribution to a random variable approaching symmetry and normality by varying the degree of freedom ( $c$ ) in the  $\chi^2(c)$  distributed innovations from 3 to 30 in increments of 1. In a similar exercise, the effect of kurtosis has been investigated by monotonically increasing  $c$  in the  $t(c)$  distribution.

The Monte Carlo experiment has been based on  $N = 10,000$  replications for  $\Lambda$ ,  $\Lambda_B$  and on  $N = 1,000$  replications for  $\Lambda_B^*$  and  $\Lambda^*$ . All the bootstrap distributions have been generated by resampling and calculating the test statistic 800 times.

#### 4.1 The Monte Carlo results

Tables 1 and 2 report the simulation results on the performance of  $\Lambda$ ,  $\Lambda_B$ ,  $\Lambda_B^*$  and  $\Lambda^*$ . The finite sample significance levels are estimated against a nominal level of 5% and all estimates are given as percentages. In Table 1 the normal distribution serves as a benchmark and it also contains results relating to the sensitivity of the error in rejection probability to variations of key parameters in (14). For the case with  $\chi^2(c)$  and  $t(c)$  distributions, the Monte Carlo results are summarized in Table 2 using response surface regressions.

As far as  $\Lambda$  is concerned, Table 1 mainly confirms previous findings that inference based on first order asymptotic critical values is markedly inaccurate with excessively high rejection frequencies. Correcting  $\Lambda$  using the analytical Bartlett factor improves the behavior of the test statistic. However, Table 1 indicates that the performance of  $\Lambda_B$  is highly dependent on the parameter values of the *DGP*. When  $\xi$  is large (i.e. the speed of adjustment to the cointegrated equilibrium is low), the correction does not work well. The magnitude of the correlation between noises,  $\eta$ , also affects the performance of the test statistic, but to a much lesser extent. By contrast, using the bootstrap to approximate the Bartlett adjustment factor produces estimated levels that show less variation over the grid of parameters taken under consideration.

Table 1. Empirical sizes for the 5% critical value (in percent). Case with  $N(0, 1)$  innovations.

		$\xi = 0.8$		$\eta = 0.5$	
		$\eta = -0.5$	$\eta = 0.5$	$\xi = 0.2$	$\xi = 0.5$
$T = 50$	$\Lambda$	26.0	26.3	9.51	12.5
	$\Lambda_B$	14.3	14.8	4.64	9.2
	$\Lambda_B^*$	7.3	8.9	4.8	6.2
	$\Lambda^*$	8.0	9.1	5.3	6.3
$T = 100$	$\Lambda$	12.9	13.1	6.77	7.74
	$\Lambda_B$	8.51	8.51	4.76	4.94
	$\Lambda_B^*$	7.7	5.8	5.8	5.8
	$\Lambda^*$	7.8	6.0	5.7	6.0
$T = 250$	$\Lambda$	7.31	7.60	5.87	6.05
	$\Lambda_B$	6.03	6.12	4.73	4.96
	$\Lambda_B^*$	4.5	4.8	5.0	5.1
	$\Lambda^*$	4.5	5.0	5.1	5.1

Note: The estimated rejection probabilities of  $\Lambda_B^*$  and  $\Lambda^*$  have been calculated using Algorithm 1 and 2 in Section 2. For  $\Lambda$  and  $\Lambda_B$  the number of replications is  $N=10,000$ , for  $\Lambda_B^*$  and  $\Lambda^*$   $N=1,000$  and  $B=800$ . A 95% confidence interval around the nominal level of 5% is given by (3.6, 6.4). The asymptotic distribution is  $\chi^2(1)$ .



To evaluate the sensitivity of  $\Lambda_B^*$  and  $\Lambda_B$  to departures from the Gaussian assumption, a simple Monte Carlo experiment has been conducted with  $\varepsilon_{it}$  drawn from the  $\chi^2(c)$  and the  $t(c)$  distributions. This experiment has adopted a factorial design covering a large range of distributions from heavily fat tailed to highly skewed innovations, but  $T$  has been fixed at 50 to control for the effect of the sample size. For each  $c$  (for  $c = 3, \dots, 30$ ), a Monte Carlo experiment assessing the size distortion of each test statistic has been undertaken. Repeating this exercise for  $GDPs$  with  $\chi^2(c)$  and  $t(c)$  distributed innovations in turn, gives us a total of  $3 \times 28 \times 2 = 168$  Monte Carlo experiments. The response surface analysis has been carried out using the error in rejecting probability,  $ERP$ , (calculated as the estimated level minus the nominal level of the statistic) as the dependent variable in a multiple regression model with a set of dummy variables as the independent variables (each dummy takes on the value 1 for a given test type and the value 0 otherwise),  $\log c$ , and, if significant, the interactions (i.e.  $\log c * \Lambda_B$ ,  $\log c * \Lambda_B^*$ ,  $\log c * \Lambda^*$ ) with  $\Lambda$  as the reference category. More complicated forms of model specification such as a Hendry (1984) style power-series expansion failed to be significant.

Table 2 displays the results of the response surface analysis. The second and third columns report the estimated coefficients for the  $\chi^2(c)$  and  $t(c)$  innovations, respectively. Robust standard errors are also reported.

From Table 2 it appears that the estimated coefficients for the dummies of  $\Lambda_B$ ,  $\Lambda_B^*$ , and  $\Lambda^*$  are significantly different from zero and have the correct signs, as we expect these tests to outperform the asymptotic test in the base line. Overall, the ability of the regression model to fit the data is good, given the high  $R^2$  values reported at the bottom of Table 2. As far as the estimated coefficients are concerned, three points should be noted. First, from the constant term and the magnitude of the estimated coefficients for  $\log c$  in the two regressions, it appears that  $\Lambda$  is more sensitive to excess kurtosis than to skewness. Of course, this implies that all the other estimated coefficients in the second column of Table 2 are smaller in modulus than the corresponding coefficients reported in the third column. Similar experiments with  $T = 75, 100, 125, \dots, 250$  have shown that excess kurtosis still affects the performance of  $\Lambda$  and  $\Lambda_B$  at  $T = 100$ , but this effect vanishes at  $T = 150$ . Secondly, as one may expect, increasing the degree of kurtosis affects the performance of  $\Lambda_B$  to a much greater extent than it affects the two non-parametric bootstrap tests. Third, given that the estimated parameter for  $\log c$  in the second column of Table 2 is not significant, the difference in the error in rejection probability between  $\Lambda_B$  and  $\Lambda_B^*$  must be attributed to the magnitude of the parameters  $\eta$  and  $\xi$  in (14), that is to the fact that  $\Lambda_B$  is more sensitive to the value of the parameters in the  $DGP$  than its bootstrap counterpart.

Table 2. Response surface analysis for  $\chi^2(c)$  and  $t(c)$  innovations.

Regressors and Statistics	$\chi^2(c)$	$t(c)$
$\Lambda_B$	-11.39* (0.09)	-12.55* (0.845)
$\Lambda_B^*$	-17.51* (0.159)	-22.48* (1.094)
$\Lambda^*$	-16.84* (0.161)	-21.38* (0.991)
$\log c$	-0.165 (0.106)	-1.527* (0.237)
$\log c * \Lambda_B$	-	0.30 (0.294)
$\log c * \Lambda_B^*$	-	1.33* (0.396)
$\log c * \Lambda^*$	-	1.24* (0.363)
<i>Constant</i>	20.36* (0.293)	25.59* (0.677)
$R^2$	0.9934	0.9891
$F(7, 104)$	6088.65	1933.1
[p-value]	[0.000]	[0.000]

(\*) Significant at 1%. Robust standard errors reported. *DGP* with  $\eta = 0.5$ ,  $\xi = 0.8$ , and  $T = 50$ .

The results of the response surface regression are perhaps more easily interpreted by plotting the actual and the predicted *ERP* for each test against  $c$ . The top panel of Figure 1 shows *ERP* for the case of  $\epsilon_{it} \sim \chi^2(c)$ , whereas in the bottom panel *ERP* related to  $\epsilon_{it} \sim t(c)$  innovations are plotted. (Note that for ease of interpretation  $c$  is presented in levels). If the difference between the rejection frequency under the null and the level of the test was small, then the error in rejecting probability should be close to zero and the *ERP* lines plotted in Figure 1 should be close to the  $x$ -axis. Moreover, if non Gaussian innovations affect the performance of the test statistic *ERP* ( $c$ ) should be a decreasing function of  $c$ . From Figure 1 it appears that skewness has little effect on  $\Lambda$ . On the other hand, excess kurtosis heavily affects the performance of the test in finite samples as the actual *ERP* is 25.3, 23.5 and 20.8 when  $c$  is equal to 3, 5, and 30, respectively. This implies that the empirical levels of  $\Lambda$  drop from approximately 30 when  $\epsilon_{it} \sim t(3)$  to approximately 26 when  $\epsilon_{it} \sim t(30)$  in the *DGP*. Looking at the simulation results for  $\Lambda_B$  it appears that the inference procedure is also not robust to excess kurtosis, as the actual *ERP* decreases from 12.7 when  $\epsilon_{it} \sim t(3)$  to 8.9 when  $\epsilon_{it} \sim t(30)$  indicating that approximately 40% of the size distortion is due to the innovation distribution.

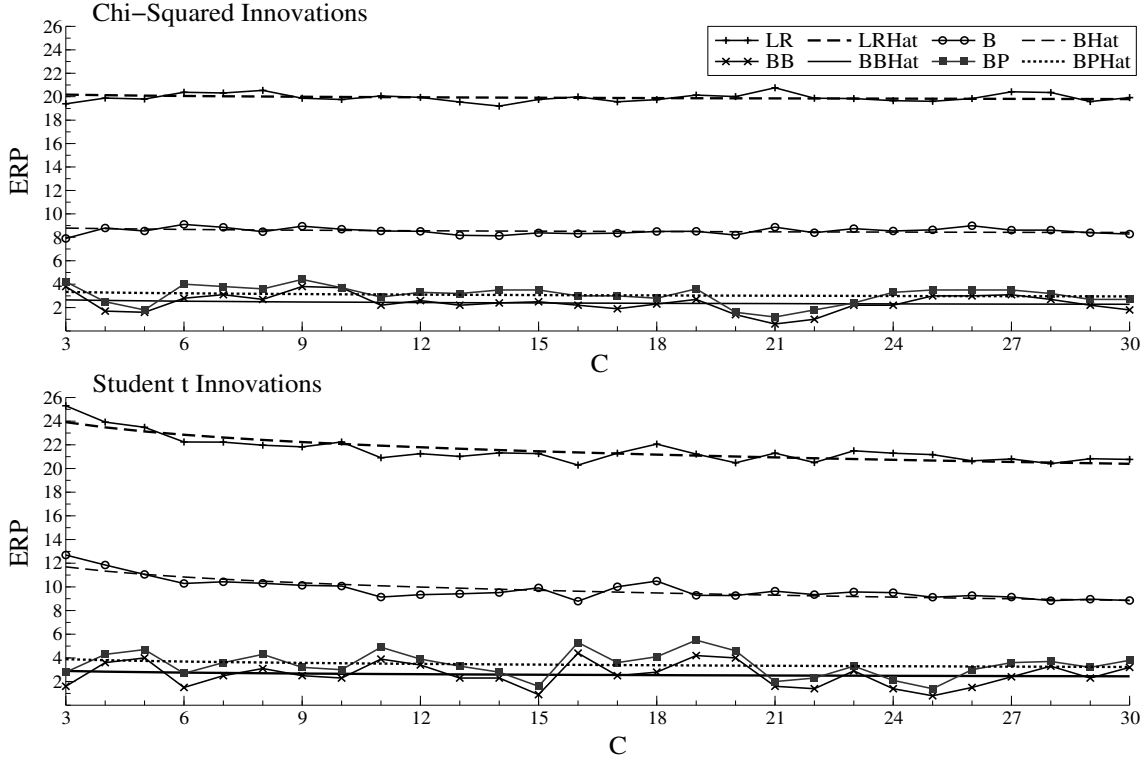


Figure 1: Response surface regression: actual and predicted ERP. Note:  $\Lambda$ ,  $\Lambda_B$ ,  $\Lambda_B^*$ , and  $\Lambda^*$  are labelled LR, B, BB, and BP respectively. The label "Hat" indicates the predicted ERP.

Table 2 reveals that, in general, the error in the rejection probability of the test statistics is affected by the non-normality of the innovations. Furthermore, from Table 1 it appears that the effect of Gaussian innovations on the estimated level of the test is highly dependent on the parameter values of the *DGP*: it is pronounced when the speed of adjustment is slow and it is relatively mild when the speed of adjustment is fast (e.g.,  $\xi = 0.2$ ). Bewley and Orden (1994) report that Johansen's estimator  $\beta$  produces outliers when the speed of adjustment is slow, while Phillips (1994) provides a theoretical analysis showing that the finite sample distribution of  $\hat{\beta}$  is leptokurtic. The simulations in Bewley and Orden (1994) and the theoretical results in Phillips (1994) explain why  $\Lambda$  behaves so poorly when the combinations of  $\xi = 0.8$  and the non-Gaussian distributions in Table 2 are selected:<sup>2</sup> excess kurtosis in the innovations magnifies the effect of the slow speed of adjustment increasing the mismatch between the finite sample and the asymptotic reference distribution of the test statistic by moving the distribution to the left. In this situation,  $\Lambda_B$  can only be partially successful because the second terms of the asymptotic expansion of the mean of  $\Lambda$  depend on the skewness and kurtosis of its distribution, and the conditions under which this dependence vanishes have not yet been established. In contrast, when using  $\Lambda_B^*$  the Gaussian distribution is replaced with the empirical density function of the innovations. This strongly mitigates the effect of skewness and kurtosis on the finite sample

<sup>2</sup>Note: a companion paper contains extensive simulation results including the case where innovations are heteroskedastic. For further results see Canepa (2012).

mean of the test and makes the finite sample distribution of  $\Lambda_B^*$  closer to the asymptotic distribution. Table 3 reports the simulated average values of the row and the adjusted  $LR$  statistic, labelled  $\bar{\Lambda}$ ,  $\bar{\Lambda}_B$  and  $\bar{\Lambda}_B^*$  respectively. If the tests were of the correct size, then the average values of the simulated  $\bar{\Lambda}$ ,  $\bar{\Lambda}_B$  and  $\bar{\Lambda}_B^*$  should be near  $q$ . Once again,  $\varepsilon_{it} \sim N(0, 1)$  is the benchmark case and the average values of the test statistics are reported in the first row. An examination of column  $\bar{\Lambda}_B$  and  $\bar{\Lambda}_B^*$  of Table 3 shows that the average values of  $\bar{\Lambda}_B^*$  are closer to  $q = 1$  than those of  $\bar{\Lambda}_B$ , explaining why  $\Lambda_B^*$  works better in finite samples.

Table 3. Average values of the simulated  $\Lambda$ ,  $\Lambda_B$  and  $\Lambda_B^*$ . Case with  $N(0, 1)$ ,  $\chi^2(3)$  and  $t(3)$  innovations.

$\varepsilon_{it}$	$T = 50$			$T = 100$			$T = 250$		
	$\bar{\Lambda}$	$\bar{\Lambda}_B$	$\bar{\Lambda}_B^*$	$\bar{\Lambda}$	$\bar{\Lambda}_B$	$\bar{\Lambda}_B^*$	$\bar{\Lambda}$	$\bar{\Lambda}_B$	$\bar{\Lambda}_B^*$
$N(0, 1)$	2.197	1.674	1.215	1.397	1.273	1.181	1.125	1.101	1.034
$\chi^2(3)$	3.351	2.232	1.221	1.992	1.395	1.218	1.244	1.161	1.067
$t(3)$	3.610	2.386	1.350	2.071	1.542	1.242	1.443	1.251	1.060

Note:  $DGP$  with  $\eta = 0.5$ ,  $\xi = 0.8$ . The average values of the test statistics are given for  $\bar{\Lambda}$ ,  $\bar{\Lambda}_B$  and  $\bar{\Lambda}_B^*$  only, as  $\Lambda^*$  does not yield an adjusted  $LR$  test.  $\bar{\Lambda}$  and  $\bar{\Lambda}_B$  are calculated on  $N=10,000$ , whereas  $\bar{\Lambda}_B^*$  is calculated on  $B=800$  and  $N=1000$ .

Before concluding this section a few points need to be made concerning the performance of the bootstrap procedures. Clearly,  $\Lambda^*$  and  $\Lambda_B^*$  perform better in terms of  $ERP$  with respect to  $\Lambda_B$ . However, from Figure 1 it appears the  $\Lambda_B^*$  slightly, but consistently outperforms  $\Lambda^*$  to some extent. One possible explanation is that the Monte Carlo design uses the same number of bootstrap replications for both bootstrap procedures. Using the delta method Rocke (1989) proves that the bootstrap Bartlett method needs fewer bootstrap replications than the bootstrap  $p$ -value method to obtain the same order of accuracy. Rocke (1989) quantifies the computational gains to be of a factor of 2-10, depending on the dimension of the constraints and how far out in the tails the observed value is. Generally speaking, using the bootstrap to approximate the first moment of a distribution is computationally less demanding than using the same method to approximate the tails of the distribution. The reason being that in the simulation experiments the value of the statistic will occur in the tails less frequently than in the "thicker" sections of the distribution. Thus, many more bootstrap trials are needed to approximate these sections. For this reason, the bootstrap Bartlett correction method may be preferred to the bootstrap  $p$ -value approach.

## 4.2 On the number of bootstrap replications

The Monte Carlo simulations presented in Tables 1 and 2 are computationally intensive. Recall that 800 bootstrap replications were performed for every 1000 Monte Carlo replications, so producing the results presented in Table 2, for example, entailed 800,000 replications for each case considered. As mentioned in Section 4.1,  $\Lambda_B^*$  may require fewer replications than  $\Lambda^*$  to obtain the same order of accuracy, depending on the constraints and the desired rejection probability under the null hypothesis. When performing a Monte Carlo evaluation of the bootstrap tests, the penalty for using a smaller number of bootstrap replications is the loss of power. According to Davidson and MacKinnon (2006)  $B = 399$  is the minimum number of replications for the  $p$ -value bootstrap test at a 5% level that does not entail loss of power. In order to investigate the claim in Section 4.1, the size and power of  $\Lambda_B^*$  and  $\Lambda^*$  have been compared when  $B = 1000$  and when  $B$  is reduced to the lowest acceptable number. For the experiments evaluating the power of the tests, data has been generated under the alternative

$$H_1 : \beta = H\varphi = \begin{bmatrix} C \\ (1 \times 3) \\ I \\ (3 \times 3) \end{bmatrix} \varphi_{(3 \times 1)},$$

where  $C = [g \ 0 \ 0]$  for  $g \in (0.5, 1, 1.5)$ . The results of this set of experiments are reported in Table 4. Note that in Table 4 only the results relating  $\varepsilon_t$  selected from  $t(3)$  and  $\chi^2(3)$  are reported. Simulation results for  $c = 4, \dots, 30$  (not reported) reveal that the rejection frequencies are in line with the results presented below.

Table 4. Comparing size and power of  $\Lambda_B^*$  and  $\Lambda^*$  under different number of bootstrap replications.

		<i>Size</i>			<i>Power</i>					
		$\varepsilon_t$	$\Lambda_B^*$	$\Lambda^*$	$g = 0.5$		$g = 1.0$		$g = 1.5$	
$B$					$\Lambda_B^*$	$\Lambda^*$	$\Lambda_B^*$	$\Lambda^*$	$\Lambda_B^*$	$\Lambda^*$
T=50	400	$t(3)$	10.1	12.4	27.1	28.9	30.4	32.3	31.2	34.4
	400	$\chi^2(3)$	8.4	9.1	39.0	41.9	44.3	47.3	45.0	38.7
	1000	$t(3)$	9.5	11.0	28.0	30.2	31.4	34.3	32.8	45.4
	1000	$\chi^2(3)$	7.2	8.3	42.8	45.2	47.6	51.0	49.3	52.7
T=100	400	$t(3)$	7.1	7.7	59.3	61.9	67.2	70.5	69.9	72.8
	400	$\chi^2(3)$	7.4	7.6	81.6	84.2	85.6	87.6	87.3	89.4
	1000	$t(3)$	7.0	8.1	62.6	65.1	70.1	72.4	73.2	75.4
	1000	$\chi^2(3)$	5.8	6.7	83.3	82.3	88.3	90.1	89.7	90.9
T=250	400	$t(3)$	4.7	5.0	99.3	99.8	100	100	100	100
	400	$\chi^2(3)$	5.1	5.5	100	100	100	100	100	100
	1000	$t(3)$	5.1	4.9	99.4	99.9	100	100	100	100
	1000	$\chi^2(3)$	6.1	5.5	100	100	100	100	100	100

Note: The estimated rejection probabilities of  $\Lambda_B^*$  and  $\Lambda^*$  have been calculated using algorithm 1 and 2 in Section 2.  $\overline{DGP}$  with  $\eta = 0.5$ ,  $\xi = 0.8$ .

Results from Table 4 show that the sample size and the distance between the null and the alternative hypothesis play an important role in determining the power of both bootstrap tests. However, comparing the power of  $\Lambda_B^*$  calculated using  $B = 400$  with the power of  $\Lambda^*$  estimated using 1000 bootstrap replications, it is clear that reducing the number of trials to the minimum produces only marginal differences in the rejection frequencies under the alternative.

To conclude this section, the results presented above show that  $\Lambda_B$  works well under specific assumptions about the innovations and caution has to be taken when violations of these assumptions occur. The failure of this test to match the performance of  $\Lambda_B^*$  implies that it would be inappropriate to compare rejection estimates derived from experiments in which the null hypothesis is false. Thus, there is a limited scope in extending the simulation exercise undertaken in Table 4 to compare the power of  $\Lambda$ ,  $\Lambda_B$ ,  $\Lambda_B^*$  and  $\Lambda^*$  given that differences in finite sample null rejection probabilities invalidate power comparisons.

## 5 Concluding remarks

In this paper we compare the non-parametric bootstrap Bartlett and the Bartlett corrected  $LR$  test proposed by Johansen (2000) and find that the performance of the former is less dependent on the values of the parameters of the

data generating process and better able to cope with violations of the Gaussian assumption about the innovations. The bootstrap  $p$ -value test is also found to perform well.

## References

- [1] Beran, R. (1988). "Prepivoting Test Statistic: A Bootstrap View of Asymptotic Refinements". *Journal of the American Statistical Association*, 83, 687-697.
- [2] Bewley, R. and D. Orden (1994). "Alternative Methods of Estimating Long-Run Responses with Application to Australian Import Demand". *Econometric Review*, 13, 179-204.
- [3] Canepa, A. and L. G. Godfrey (2007). "Improvement of the Quasi-Likelihood Ratio Test in ARMA Models: Some Results for Bootstrap Methods". *Journal of Time Series Analysis*, 28, 434-453.
- [4] Canepa A. (2012). "Robust Bartlett Adjustment for Hypotheses Testing on Cointegrating Vectors". *Economics and Finance Working Paper Series, Working Paper No. 12-10*.
- [5] Davidson, R. and J. G. MacKinnon (2006). "The Power of Bootstrap Tests and Asymptotic Test", *Journal of Econometrics*, 133, 421-441.
- [6] Gine, E. and J. Zinn, (1990). "Bootstrapping General Empirical Measures". *The Annals of Probability*, 18, 851-869.
- [7] Gredenhoff, M. and T. Jacobson (2001). "Bootstrap Testing Linear Restrictions on Cointegrating Vectors". *Journal of Business and Economics Statistics*, 19, 63-72.
- [8] Haug, A. (2002). "Testing Linear Restrictions on Cointegrating Vectors: Sizes and Powers of Wald and Likelihood Ratio Tests in Finite Samples". *Econometric Theory*, 18, 505-524.
- [9] Hendry, D. F. (1984). "Monte Carlo Experimentation in Econometrics". In Z. Griliches and M.D. Intriligator (Eds.), *Handbook of Econometrics (Vol 2)*. Amsterdam, Netherlands: Elsevier.
- [10] Johansen, S. (1996). *Likelihood Inference in Cointegrated Vector Auto-Regressive Models*. Oxford University Press, Oxford.
- [11] Johansen, S. (2000). "A Bartlett Correction Factor for Tests of on the Cointegrating Relations". *Econometric Theory*, 16, 740-778.
- [12] Phillips, P.C.B. (1994). "Some Exact Distribution Theory for Maximum Likelihood Estimators of Cointegrating Coefficients in Error Vector Correction Models", *Econometrica*, 62, 73-93.
- [13] Roche, D.M. (1989). "Bootstrap Bartlett Adjustment in Seemingly Unrelated Regression". *Journal of the American Statistical Association*, 84, 598-601.
- [14] Swensen, A.R. (2006). "Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models", *Econometrica*, 74, 1699-1714.