
Working Paper Series

08/21

SMALL SAMPLE ADJUSTMENT FOR HYPOTHESES TESTING ON COINTEGRATING VECTORS

ALESSANDRA CANEPA

Small Sample Adjustment for Hypotheses Testing on Cointegrating Vectors

Abstract

Johansen's (2000) Bartlett correction factor for the LR test of linear restrictions on cointegrated vectors is derived under the i.i.d. Gaussian assumption for the innovation terms. However, the distribution of most data relating to financial variables are fat-tailed and often skewed, there is therefore a need to examine small sample inference procedures that require weaker assumptions for the innovation term. This paper suggests that using the non-parametric bootstrap to approximate a Bartlett-type correction provides a statistic that does not require specification of the innovation distribution and can be used by applied econometricians to perform a small sample inference procedure that is less computationally demanding than its analytical counterpart. The procedure involves calculating a number of bootstrap values of the LR test statistic and estimating the expected value of the test statistic by the average value of the bootstrapped LR statistic. Simulation results suggest that the inference procedure has good finite sample property and is less dependent on the parameter space of the data generating process.

JEL Code: C12, C15, C22.

Keywords: Cointegration, LR test, Bootstrap, Bartlett Correction.

Word Count: 8168.

Alessandra Canepa,

Department of Economics and Economics and Statistics Cogneetti De Martiis

University of Turin,

E-mail: Alessandra.Canepa@unito.it

1 Introduction

The procedure for estimating and testing cointegrating relationships described in Johansen (2006) is available in virtually all econometric software packages and is widely used in applied research. Briefly this method involves maximizing the Gaussian likelihood function and analysing the eigenvalues and eigenvectors found using the reduced rank regression method. Once that the number of cointegrating vectors has been determined, hypotheses on the structural economic relationships underlying the long-run model can be tested using the likelihood ratio (LR) test.

Although the LR test of linear restrictions of cointegrating vectors has the correct size asymptotically, many studies contain reports that the approximation of the χ^2 distribution to the finite sample distribution of the LR test can be seriously inaccurate see, for example, Haug (2002), or Gredenhoff and Jacobson (2001). Broadly speaking, the problem can be described as one of lacking coherence between the test statistic and its reference distribution. One way of addressing this problem is to correct the test statistic so that the finite sample distribution is closer to the asymptotic distribution. In this respect, early attempts of correcting the test statistic were made in Podivinsky (1992) and Psaradakis (1994), where small sample corrections based on degrees of freedom were suggested. More recently, Johansen (2002) proposed a Bartlett type correction factor for LR statistic and analytically derives the asymptotic expansions needed to calculate the expectation of the test statistic. Multiplying the unadjusted statistic by a factor derived from an asymptotic expansion of the expectation of the test provides a closer approximation of the resulting adjusted statistic to the χ^2 distribution, thus reducing the size distortion problem. Simulation results presented by Johansen (2000) suggests 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. One way of overcoming such calculations is to use a numerical approximation in place of the analytical Bartlett correction. The first paper that suggested to calculate such approximation in time series context was the work by Canepa and Godfrey (2007). In their article the authors proposed 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. The Bartlett corrected quasi- LR test suggested in Canepa and Godfrey (2007) can be used in ARMA models as a misspecification test or alternatively as model specification procedure. The former case involves testing for the adequacy of an ARMA model by an overfitting diagnostic check, whereas in the latter case the inference procedure can be used to test the validity of simplifications of an ARMA model. In

the multivariate context, Canepa (2016) proposed to use the bootstrap to approximate the finite sample expectation of the LR test in place of the analytical Bartlett correction in the context of cointegrated VAR models (see also Hunter, Burke and Canepa, 2017). The author also proved the consistency of the suggested test under the assumption that in the innovations are independent and identically distributed (i.i.d.). The results in Canepa (2016) are promising since it was found that the performance of the bootstrap Bartlett test was 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 with respect to the Johansen (2000) Bartlett corrected LR test. A possible shortcoming of the inference procedure suggested in Canepa (2016) is that the suggested test statistic relies on the i.i.d. assumption. Evidence of violation of this strong assumption appear in many macroeconomic series, such as aggregate consumption and income, in interest rate data and in nominal and real price variables; see Sensier and van Dijk (2004). Similarly, it is a well-known stylized fact that GARCH-type models fit well to stock market returns (see Boswijk *et al.*, 2016; Harvey *et al.*, 2016; Engle and Rangel, 2008; among others).

Against this background, in this paper we built on Canepa (2016) and investigate if the bootstrap Bartlett corrected LR test can be used to reduce the size distortion problem in situations where an analytical solution is difficult or does not work well. If such an application were to be successful it would have significant practical implications, for several reasons. The bootstrap Bartlett corrected LR test does not rely on the Gaussian assumption of the innovations, and this feature may be appealing to the applied worker. Moreover, simulation results indicate that the 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, and where the correction does not work well" (cf. Johansen, 2000 p.741). Simulation results in Cavaliere *et al.* (2015) confirmed that the performance of test heavily depends on the parameter space (see also Canepa, 2006; Cavaliere *et al.*, 2020; Andrews and Guggenberger, 2009; Elliott *et al.*, 2015; Lu, 2016; McCloskey, 2017; Ketz, 2018).

We believe that the dependency on the parameter values may be reduced by computing the Bartlett adjustment using the non-parametric bootstrap. Because the bootstrap method involves replacing the unknown cumulative distribution function of the LR test statistic by the empirical distribution function of the bootstrap distribution of the same test, the resulting inference procedure may show less sensitivity to the values of the parameters of the data generating process (DGP) than a test based on the asymptotic critical values. Computing the bootstrap Bartlett correction factor is relatively straightforward. Roughly speaking, this procedure involves calculating a number of bootstrap values of the LR test statistic and estimating the expected value of the test

statistic by the average value of the bootstrapped LR statistic. The bootstrap Bartlett method was first proposed in Rocke (1989) where hypothesis testing in seemingly unrelated regression models was considered (see also Jacobson and Larsson, 1999). Rocke's simulation results showed that the Bartlett adjustment for the LR test determined using the non-parametric bootstrap was considerably more accurate than the Bartlett adjustment from the second-order asymptotic method of Rothenberg (1984). In the time series context the idea of using the bootstrap to tackle the problem of finite sample biased estimation is also considered in Engsted and Pedersen (2014) where a bootstrap based bias-correction method is considered as a simple approach to bias-adjustment in VAR models. The authors show that the bootstrap bias-correction approach yield a large reduction in bias compared to OLS estimate and the performance is similar to the analytical bias formula.

The contribution to the literature of this paper is twofold. First, it provides a "feasible" small sample correction factor that can easily be used to calculate LR test for linear restriction on the cointegrating space in situations where the analytical Bartlett correction factor would fail. Second, the consistency of the suggested procedure is considered and it is shown that the bootstrap Bartlett statistic converges weakly in probability to the correct asymptotic distribution. Establishing the conditions that ensure asymptotic refinements will be the subject of future research. However, 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 analytical Bartlett adjustment produces an error in rejecting probability 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 statistic, hence the resulting inference procedure should have accuracy of the same order.

The performance of the suggested procedure and the analytical Bartlett correction under non-normality assumption of the innovations are evaluated by Monte Carlo evaluation. Innovation structures typically found in financial data are considered such as fat tailed and conditionally heteroskedastic (i.e., *ARCH* and *GARCH*) innovations. Performance is assessed in terms of the size and power of the inference procedures under consideration. The performance of the suggested procedure is compared with the bootstrap p -value test. In general, unlike the Bartlett (1937) idea that involves replacing the original statistic with a corrected statistic which is closer to the reference distribution, the bootstrap p -value test involves making the reference distribution closer to the finite sample distribution. In the literature it has been shown that in many cases the bootstrap can be considered as a numerical approximation to analytical calculations of one-term Edgeworth expansion (see Hall,1992) where the critical values of the limit distribution are replaced with transformations of critical values obtained from the Edgeworth expansions of the distribution

function.

This section will close with a brief presentation of the Bartlett correction. The next section introduces the LR test for linear restrictions on cointegrating space, the Bartlett correction of Johansen (2000), and the two bootstrap inference procedures. In Section 3, the design of the Monte Carlo experiment is explained, and in Section 4, the simulation results are reported. In Section 5 some robustness checks are undertaken. An empirical application is considered in Section 6 and Section 7 contains some concluding remarks.

The Bartlett Correction

The Bartlett correction is based on a simple idea, but can be very effective in reducing the finite sample size distortion problem of the LR tests. This method takes the form of a correction to the mean of the LR statistic for a given parameter point θ under the null hypothesis. In regular cases, the asymptotic distribution of the LR statistic is given by $\Lambda = -2\log(LR) \sim \chi^2(q)$ where q is the dimension of the constraints, and the asymptotic mean of the LR statistic ought to be approximately equal to q . The Bartlett correction is intended to make the mean exactly equal to q by replacing the above equation by $\Lambda_B = \frac{q\Lambda}{E_\theta(\Lambda)}$ and then referring the resulting statistic to a $\chi^2(q)$. Typically, given the complicated form of the LR test, it is difficult to find an exact expression for $E_\theta(\Lambda)$ and one can instead find an approximation of the form

$$E_\theta(\Lambda) = q \left(1 + \frac{b(\theta)}{T} \right) + O(T^{-3/2}). \quad (1)$$

Thus, the quantity

$$\frac{\Lambda}{1 + \frac{b(\hat{\theta})}{T}}$$

has an expectation $q + O(T^{-3/2})$ which is closer to the limit distribution.

In the independent and identically distributed (i.i.d.) setup, the Bartlett correction has been widely studied in the literature since the pioneering work by Lawley (1956). In the paper the author showed that the finite-sample distribution of a Bartlett-corrected likelihood ratio test statistic was closer to the χ^2 -distribution than the original LR statistic; see Cribari-Neto and Cordeiro (1996) for a review. However, the Bartlett correction is relatively less explored for dependent data. In the time series context, Giersbergen (2009) derived a Bartlett correction factor for testing hypotheses about the autoregressive parameter in the stable AR(1) model with trend and intercept and showed that the Bartlett corrections are useful in controlling the size of the likelihood ratio statistic in small samples (see also Omtzigt, 2003). Similarly, Taniguchi (1991) calculated the corrections for the AR(1) and MA(1) models without disturbance parameters, whereas Van Garderen (1999) used differential geometry and suggested how to compute the Bartlett correction factor geometrically,

for the case of no disturbance parameter. In the context of ARMA (autoregressive moving average), Lagos and Morettin (2004) derived a Bartlett correction for the likelihood ratio statistic used to test hypotheses about parameters of a Gaussian stationary and invertible model. Ravishanker *et al.*(1990) considered the differential geometry of autoregressive fractionally integrated moving average processes and use the properties of Toeplitz forms associated with the spectral density functions of these long memory processes to compute the geometric quantities. The authors investigated the role of geometric quantities on the Bartlett correction to the likelihood ratio test statistics for the fractional difference parameter. Similarly, Chen and Cui (2006, 2007) proved that the empirical likelihood with moment restrictions is Bartlett correctable even in the presence of a nuisance parameter.

Recently, Bartlett-type corrections in unstable autoregressive models have also attracted much attention. For example, Chan and Liu (2020) proved that the Bartlett correction can be used for Gaussian short-memory time series. Chan *et al.* (2014) extended Chan and Liu (2010) to Gaussian long-memory time series. Similarly, Chen *et al.* (2016) examine the Bartlett correction factor for the frequency domain empirical likelihood of linear time series models and showed that the factor can be used for non-Gaussian short-memory time series; see also Chan *et al.*(2010), (for earlier studies see for example Bravo, 1999; Nielsen, 1997; Larsson, 1998).

2 Model and Tests

Consider the p -dimensional VAR model

$$\Delta Y_t = \alpha (\beta' Y_{t-1} + \rho' D_t) + \sum_{i=1}^{k-1} \Gamma_i \Delta Y_{t-i} + \phi d_t + \varepsilon_t, \quad t = 1, \dots, T \quad (2)$$

where $Y_t, \varepsilon_t \sim (0, \Omega)$ are $(p \times 1)$ vectors with $E(\varepsilon_t \varepsilon_s) = 0$ (for $t \neq s$) and $\Delta Y_t = Y_t - Y_{t-1}$. The matrices of coefficients have the following dimensions: α and β are $(p \times r)$; ϕ is $(p \times p_d)$; ρ is $(p_d \times r)$; and $\Gamma_1, \dots, \Gamma_{k-1}$ are $(p \times p)$. Also, d_t ($p_d \times 1$) and D_t ($p_D \times 1$) are deterministic terms in (2). Once the cointegrating rank has been established linear restrictions on cointegrating space can be tested for. 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 φ is an $(s \times r)$ matrix; see Johansen (1996) for a discussion of tests for other hypotheses. The LR test statistic for \mathcal{H}_0 can be obtained from the concentrated likelihood function and is given by

$$\Lambda = -T \sum_{i=1}^r \log \left[\frac{(1 - \hat{\lambda}_i)}{(1 - \tilde{\lambda}_i)} \right], \quad (3)$$

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.

For the null hypothesis $\mathcal{H}_0 : \beta = H\varphi$ an approximation to the order T^{-1} for the Bartlett adjustment is given by

$$\begin{aligned} \vartheta = \frac{E_{\hat{\theta}}(\Lambda)}{q} = 1 + \frac{1}{T} \left[\frac{1}{2} (p + s - r + 1 + 2p_D) + p_d + kp \right] \\ + \frac{1}{Tr} [(2p + s - 3r - 1 + 2p_D)v(\alpha) + 2(c(\alpha) + c_d(\alpha))], \end{aligned} \quad (4)$$

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 likelihood ratio test in (3) and the correction in (4) are derived under the assumption that the innovations are $\varepsilon_t \sim N(0, \Omega)$. However, the Gaussian hypothesis is often too restrictive for the type of data used in economic applications. The fact that the distribution of most data relating to financial variables, for example (but certainly not exclusively), are fat tailed and often skewed has been extensively documented in the finance literature. Although, under weak conditions relaxing the Gaussian hypothesis does not affect the asymptotic distribution of Λ , one may expect the finite sample error in rejecting probability to be larger. Moreover, when innovations are non-Gaussian, the second terms of the asymptotic expansions of the mean and the variance of Λ depend on the skewness and kurtosis of their distribution. This means that in order to use the analytical Bartlett's correction factor it is necessary to estimate the skewness and kurtosis of the true distribution and accordingly modify the Bartlett's adjustment. Rather than undertaking these tedious calculations, it is proposed below that the non-parametric bootstrap be used to approximate the finite sample expectation of Λ . By using the empirical distribution function in place of some specific parametric distribution, the non-parametric bootstrap method does not require a choice of error distribution be made; this feature is desirable with many type of data. The proposed inference procedure involves undertaking a simulation study using the constrained estimates of β obtained by solving the eigenvalue problem, conditional on the initial values Y_0 and ΔY_0 , as the true values. Given these estimates and any required starting values, bootstrap data can be generated recursively after resampling residuals. From each generated sample, one obtains a bootstrap value of the LR statistic, say Λ_j^* , whose average estimates the mean of Λ under the null hypothesis. An alternative procedure is a straightforward application of the bootstrap p -value approach, where the significance level assigned to Λ is the fraction of the Λ_j^* greater than Λ . (Note that the subscript “*” is used to indicate the bootstrap analog throughout the paper).

2.1 Bootstrap Algorithms

Bootstrap methods rely on simulations to approximate the finite-sample distribution of the test statistic under consideration. In order to achieve accurate inference procedures the bootstrap *DGP* used for drawing bootstrap samples has to mimic the features of the underlying *DGP*. In this section, we describe the two bootstrap algorithms used to calculate the bootstrap Bartlett corrected test (Λ_B^*) and the bootstrap p -value test (Λ^*). The former is suitable for the model in (2) when the innovations are i.i.d., whereas the latter is used when innovations are independent but not identically distributed.

2.1.1 Algorithm 1

The steps used to implement the bootstrap algorithm for calculating the bootstrap Bartlett corrected LR test can be summarized as follows:

Step (1): Estimate the model in (2) and compute Λ and the estimated restricted residuals as

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

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

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

recursively from $(\varepsilon_1^*, \dots, \varepsilon_T^*)$ using the estimated restricted model given in (6.2).

Step (3): Compute Λ_j^* using the data of step (2) and repeat B times.

Step (4): Average the observed values $\Lambda_1^*, \dots, \Lambda_B^*$ to get an estimate, $\bar{\Lambda}^*$, of the average value of Λ . 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 $q = r(n-s)$).

As far as the test Λ^* is concerned, when innovations are independent and identically distributed with common variance, it is possible to obtain an accurate inference by simply resampling the residuals of the estimated restricted model in (2) without the need to make a particular parametric assumption about the distribution of the innovations. Swensen (2006) considers a recursive bootstrap algorithm for testing the rank of $\Pi = \alpha\beta'$ in (2) and shows that, under a variety of regularity conditions, the non-parametric bootstrap based test is consistent in the sense that the bootstrap statistic converges weakly in probability to the correct asymptotic distribution. This involves repeating Step (1)-(3) and then following Step (5) below.

Step (5) Compute the bootstrap p -value function of the observed value Λ 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, Λ^* , is carried out by comparing $\hat{P}^*(\Lambda)$ with the desired critical level, γ , and rejecting the null hypothesis if $\hat{P}^*(\Lambda) \leq \gamma$. Note that the resampling and testing in algorithm 1 is done once that the cointegrating rank has been established. Therefore, for a given cointegrating rank all unit roots have been eliminated.

2.1.2 Algorithm 2

When the innovations show conditional heteroskedasticity simply resampling from the residual fails to mimic essential features of the *DGP* that initially generated the data. A suitable modification of the residual based bootstrap procedure is the wild bootstrap, which is designed to accommodate the possibility of independent but not identically distributed innovations. The wild bootstrap method was developed by Liu (1988) based on a suggestion presented in Wu (1986). Regarding time series, Gonçalves and Kilian (2003) proposed a recursive-design implementation of the wild bootstrap for the autoregression model with conditionally heteroskedastic innovations. For cointegrated VAR models, noteworthy are the recent papers by Cavaliere, Rahbek and Taylor (2010a) and Cavaliere, Rahbek and Taylor (2010b).

The wild bootstrap *DGP* is given by

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

where $v_t^* = \hat{\varepsilon}_t Z_t$ and Z_t is specified as a two-point distribution

$$\begin{aligned} Z_t &= \frac{-(\sqrt{5}-1)}{2} && \text{with probability } \frac{(\sqrt{5}+1)}{2\sqrt{5}} \\ &= \frac{(\sqrt{5}+1)}{2} && \text{with probability } \frac{(\sqrt{5}-1)}{2\sqrt{5}} \end{aligned}$$

so that Z_t terms are mutually independent drawings from a distribution which is independent of the original data and has the properties that $E(Z_t) = 0$, $E(Z_t^2) = 1$, and $E(Z_t^3) = 1$. Given the bootstrap data, the associated value of the test statistic Λ_t^* can be calculated; repeat B times and follows *Step (4)* to calculate Λ_B^* and *Step (5)* to calculate Λ^* .

Using the fact that $\theta = f(\alpha, \beta, \Gamma_i, \Omega)$ is consistently estimated in the presence of conditional heteroskedastic innovations, we show below that Λ^* and Λ_B^* converge weakly in probability to the first order asymptotic null-distribution of Λ .

REMARK 2.1: The procedure outlined in Algorithm 2 is suitable when the innovations are serially uncorrelated. Many alternative procedures could be used for generating the bootstrap *DGP*, such as the block bootstrap for example. Establishing which bootstrap scheme is the best to calculate the Bartlett correction factor under different assumptions on the innovation process is outside the scope of this paper. In this work the wild bootstrap was preferred to the block bootstrap for the following reasons. First, the wild bootstrap method is easier to implement than the block bootstrap as it does not involve the problem of determining block length as the latter bootstrap method does. Second, under Assumption 1 below, the innovations form an uncorrelated martingale difference sequence and using the block bootstrap procedure when innovations are uncorrelated may result in a loss of efficiency. Finally, the consistency of the wild bootstrap in the present context can be proved using available tools for independent random variables. However, when innovations admit serial correlation using Algorithm 2 would fail to replicate the correlation structure of the residuals, therefore the procedure is no longer valid.

REMARK 2.2: Note that the bootstrap Bartlett correction could easily be extended to other inference procedure, such as the Wald test or may find good applications in structural VAR models in cases where inference on the slope parameters is needed. However, recent theoretical results in Brüggemann, Jentsch and Trenkler (2016) show that for inference in VAR statistics that depend both on the VAR slope and the variance parameters (e.g. in structural impulse response functions) For these statistics Algorithm 2 would fail in the presence of conditional heteroskedasticity because the bootstrap algorithm does not correctly replicate the relevant fourth moments' structure of the error terms (see also Jentsch and Lunsford, 2019). In contrast, the residual-based moving block bootstrap results in asymptotically valid inference. Investigating the usefulness of the suggested Bartlett test in the cases where the block bootstrap is needed will be the subject of future research.

2.2 Some Asymptotic Results

We now consider the Λ_B^* statistic and show that the distributions of the bootstrap based inference procedure coincides with the corresponding asymptotic counterpart. In this paper we focus on the pseudo-data generated by Algorithm 2 since the consistency of the bootstrap procedure proposed in Algorithm 1 is derived in Canepa (2016). Our approach builds on the important theoretical

results in Cavaliere, Rahbek and Taylor (2010a). The authors demonstrate that the limiting null distributions of the rank statistics remain valid in the less restrictive case of globally stationary, conditionally heteroskedastic shocks satisfying certain moment conditions. Cavaliere *et al.* (2010a) also show that the pseudo maximum likelihood estimator of the error correction model that assumes Gaussian i.i.d. disturbances remains consistent under these weaker conditions. Building on these results below we show that the functional central limit theorem for the stochastic process built from the sequence of partial sums for the bootstrap analog holds and the expected value of the test is consistency estimated.

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 $A_{\perp}A = 0$ (where (A, A_{\perp}) is a full rank matrix), and the norm $\|A\|$ is $\|A\| = [\text{tr}(A'A)]^{1/2}$. For any vector a , $\|a\|$ denotes the Euclidean distance norm, $\|a\| = (a'a)^{1/2}$.

In order to establish the validity of the wild bootstrap we need to impose some conditions on the innovations. More precisely, we make the following assumption:

Assumption 1

(i) Define the characteristic polynomial,

$$A(z) = (1 - z)I_p - \alpha\beta'z - \Gamma_1(1 - z)z - \dots - \Gamma_{k-1}(1 - z)z^{k-1}. \quad (5)$$

Assume that the roots of $[[A(z)]] = 0$ are located outside the complex unit circle or at 1. Also assume that the matrices α and β 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 $\{\varepsilon_t\}$ form martingale difference sequence with respect to the filtration \mathcal{F}_t , $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$, with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t\varepsilon_t') = \Omega < \infty$.

(iii) $E\|\varepsilon_t\|^{4+\varsigma} < \infty$, $\varsigma > 0$.

Assumption 1 replaces the usual Gaussian assumption on the innovations $\{\varepsilon_t\}$ by the less restrictive martingale sequence assumption. The innovations are not correlated, however ARCH and GARCH effects are now allowed in model (2) by Assumption 1-(ii). Finally, condition iii) requires the $4 + \varsigma$ moments to be uniformly finite.

Under Assumption 1, Theorem 1 in Rahbek, Hansen and Dennis (2002) implies that the process Y_t has the following representation

$$Y_t = C \sum_{i=1}^t (\varepsilon_i + \rho'D_i) + \sum_{i=0}^{\infty} C_i(\varepsilon_{t-i} + \phi d_{t-i}) + A_0, \quad (6)$$

where $C = \beta_{\perp} \left(\alpha'_{\perp} \left(I - \sum_{i=1}^{k-1} \Gamma_i \right) \beta_{\perp} \right)^{-1} \alpha'_{\perp}$, the coefficients C_i decrease exponentially, and A_0 is a term that depends only on the initial values and $\beta' A_0 = 0$. Moreover, Theorem 2.1 in Hansen (1992) implies the weak convergence of the stochastic integrals

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

$$T^{-2} \sum_{t=1}^T \left(\sum_{i=1}^t \varepsilon_i \right) \left(\sum_{i=1}^t \varepsilon_i \right)' \xrightarrow{w} \int_0^1 B(u) B(u)' du,$$

where $B = \Omega^{1/2} W$ is a p -dimensional Brownian motion with variance Ω and W a p -dimensional standard Brownian motion. Rahbek *et al.* (2002) use this result to derive the asymptotic distribution of the pseudo likelihood ratio test for cointegrating rank. They show that, under the assumption that innovations form a stationary and ergodic vector of martingale difference sequence, the limit distributions of the rank tests are invariant to heteroskedasticity (see also Seo, 2006). In the paper by Cavaliere, Rahbek and Taylor (2010a) it is shown that the limiting null distributions of the rank tests remain valid in the less restrictive case of global stationarity.

Turning now to the statistics constructed under the pseudo-data generated by Algorithm 2, the representation in (6) is still valid for each bootstrap replication. However, the remainder term, (A_0) , depends on the realization and needs careful consideration in the bootstrap context. Theorem 1 extends the validity of Lemma 1 in Swensen (2006) (derived under the assumption of i.i.d. innovations) to the case where innovations form an uncorrelated martingale sequence difference with finite fourth moments. In the following we set ρ, ϕ and the initial values of Y to zero, without loss of generality. As before, an asterisk (*) denotes the bootstrap analog.

Theorem 1: *Let the conditions of Assumption 1 hold. Then, under the null hypothesis, $\Lambda^* \xrightarrow{w_p} \Lambda$ as $T \rightarrow \infty$.*

Corollary 1: *Under the conditions of Theorem 1, $E^*(\Lambda^*) \xrightarrow{P} E(\Lambda)$ as $T \rightarrow \infty$.*

Proof of Theorem 1: By Lemma A.4 in Cavaliere *et al.* (2010a), under Assumption 1, the generated pseudo observations have the representation

$$Y_t^* = \hat{C} \sum_{i=1}^t v_i^* + T^{1/2} R_t^*, \quad (7)$$

where $\hat{C} = \hat{\beta}_{\perp} \left(\hat{\alpha}'_{\perp} \left(I - \sum_{i=1}^{k-1} \hat{\Gamma}_i \right) \hat{\beta}_{\perp} \right)^{-1} \hat{\alpha}'_{\perp}$ and, for all $\eta > 0$, $P^*(\max_{t=1, \dots, T} \|R_t^*\| > \eta) \rightarrow 0$ in probability as $T \rightarrow \infty$.

Using the results in (7) we can describe the asymptotic properties of the product moment matrices generated using the pseudo-observations, which are the basic properties of the test statistics. Following the standard notation, we define R_{0t} and R_{1t} as the residuals obtained by regressing $\tilde{Z}_{0t} = \Delta Y_t$ and $\tilde{Z}_{1t} = Y_{t-1}$, respectively on $\tilde{Z}_{2t} = [\Delta Y'_{t-1}, \Delta Y'_{t-2}, \dots, \Delta Y'_{t-k+1}]$. Moreover,

$$S_{i,j} = T^{-1} \sum_{t=1}^T R_{it} R'_{jt} = M_{ij} - M_{i2} M_{22}^{-1} M_{2j} \quad i, j = 0, 1.$$

and $M_{ij} = T^{-1} \sum_{t=1}^T \tilde{Z}_{it} \tilde{Z}'_{jt}$.

Let $\ddot{\Omega}_{\beta\beta} = p \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \beta' \tilde{Z}_{1t} \tilde{Z}'_{1t} \beta$, $\ddot{\Omega}_{\beta i} = p \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \beta' \tilde{Z}_{1t} \tilde{Z}'_{it}$ for $i = 0, 2$, and $\ddot{\Omega}_{ij} = p \lim_{T \rightarrow \infty} T^{-1} \sum_{t=1}^T \tilde{Z}_{it} \tilde{Z}'_{jt}$ for $i, j = 0, 2$. Under Assumption 1,

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

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

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

where $\Sigma_{ij} = \ddot{\Omega}_{ij} - \ddot{\Omega}_{i2} \ddot{\Omega}_{22}^{-1} \ddot{\Omega}_{2j}$ for $i, j = 0, 1, \beta$. Moreover,

$$T^{-1/2} \sum_{t=1}^{[Tu]} v_t^* \xrightarrow{w_p} B(u), \quad (11)$$

$$T^{-1} \hat{\beta}'_{\perp} S_{11}^* \hat{\beta}_{\perp} \xrightarrow{w_p} \int_0^1 F(u) F(u)' du, \quad (12)$$

$$\hat{\beta}'_{\perp} \left(S_{10}^* - S_{11}^* \hat{\beta} \alpha' \right) \alpha_{\perp} \xrightarrow{w_p} \int_0^1 F(u) dB' \alpha_{\perp}, \quad (13)$$

where $F(u) := \beta'_{\perp} C B(u)$ and $[Tu]$ is the integer value of uT .

The proof for (8)-(10) mimics the proof of Lemma A.7 in Cavaliere *et al.* (2010a). Similarly, Lemma A.5 in the same paper implies that the functional central limit theorem for the stochastic process built from the sequence of partial sums corresponding to the bootstrap resamples holds, so that

$$T^{-1/2} \sum_{t=1}^{[Tu]} v_t^* \xrightarrow{w_p} B(u).$$

Considering now, (11) and (12), as the reminder R_t^* in (7) vanishes Lemma 10 in Rahbek *et al.* (2002) holds and

$$T^{-1/2} \hat{\beta}'_{\perp} Y_{[Tu]}^* \xrightarrow{w_p} F(u)$$

such that the continuous mapping theorem gives

$$T^{-1}\hat{\beta}'_{\perp}S_{11}^*\hat{\beta}_{\perp} \xrightarrow{w_p} \int_0^1 F(u)F(u)'du.$$

Similarly, we have

$$\hat{\beta}'_{\perp} \left(S_{10}^* - S_{11}^*\hat{\beta}\hat{\alpha}' \right) \hat{\alpha}_{\perp} \xrightarrow{w_p} \int_0^1 F(u)dB'\alpha_{\perp}.$$

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 (7), then $\{H'Y_t^*\}$ also satisfies the same conditions. Moreover, the random process $\left\{ T^{-1/2} \left(\sum_{t=1}^{\lfloor Tu \rfloor} H'v_i^* \right) \right\}$ converges weakly toward a Brownian motion with covariance matrix $H'\hat{\Omega}H$ and the asymptotic distribution of the moment matrices is given by

$$T^{-1/2}\hat{\varphi}'_{\perp}H'Y_{[Tu]}^* \xrightarrow{w_p} H'\tilde{F}(u) \quad (14)$$

$$\hat{\varphi}'_{\perp}H'S_{10}^*\hat{\alpha}_{\perp} \xrightarrow{w_p} H' \int_0^1 \tilde{F}dB'\alpha_{\perp} \quad (15)$$

$$T^{-1}\hat{\varphi}'_{\perp}H'S_{11}^*H\hat{\varphi}_{\perp} \xrightarrow{w_p} H' \int_0^1 \tilde{F}(u)\tilde{F}(u)'duH. \quad (16)$$

where $\tilde{F}(u) := \varphi'_{\perp}CB(u)$. From Theorem 1 it follows that the $(p-r)$ smallest solutions of

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

converge to zero. Therefore, using (14)-(16) the asymptotic distribution of Λ^* can found by mimicking Theorem 13.9 in Johansen (1996). ■

Proof of Corollary 1:

Under Assumption 1, (11)-(12) imply weak convergence of the partial sums of stochastic integrals. Moreover, from (8)-(10) we have that $S_{ij}^* \rightarrow \Sigma_{ij}$ in probability and the estimators of the parameters are consistent. Under the conditions of Theorem 1, this trivially implies that $E^*(\Lambda^*) \rightarrow E(\Lambda)$ in probability as $T \rightarrow \infty$. ■

3 The Monte Carlo design

To what extent do deviations from the Gaussian assumption in model (2) affect the finite sample performance of the analytical Bartlett correction? In addition, can the non-parametric bootstrap based Bartlett adjustment introduced above deliver accurate small sample inference when the Gaussian assumption on the innovations is relaxed? Questions of this nature can best be settled by case and simulation studies. We now describe the Monte Carlo study that addresses these issues.

The *DGP* adopted is given by

$$Y_{1t} = Y_{2t} + u_{1t} \quad \text{where} \quad u_{1t} = \xi u_{1t-1} + \varepsilon_{1t} \quad (17)$$

$$Y_{2t} = -Y_{1t} + u_{2t} \quad u_{2t} = u_{2t-1} + \varepsilon_{2t}$$

$$\Delta Y_{3t} = \varepsilon_{3t}$$

$$\Delta Y_{4t} = \varepsilon_{4t}$$

with

$$\begin{bmatrix} \varepsilon_{s,t} \\ \varepsilon_{l,t} \end{bmatrix} \sim i.i.d. N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}_{(2 \times 1)}, \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}_{(2 \times 2)} \right]$$

where $\varepsilon_{s,t} = \begin{bmatrix} \varepsilon_{1t} & \varepsilon_{2t} \end{bmatrix}'$, $\varepsilon_{l,t} = \begin{bmatrix} \varepsilon_{3t} & \varepsilon_{4t} \end{bmatrix}'$, $A = \begin{pmatrix} \sigma^2 & \sigma\kappa \\ \sigma\kappa & \sigma^2 \end{pmatrix}$, and $B = \sigma^2 I$. The null

hypothesis

$$\mathcal{H}_0 : \beta = H\varphi = \begin{bmatrix} 1 & 0 \\ -1 & 0 \\ 0 & I \end{bmatrix} \begin{matrix} \varphi \\ \varphi \\ \varphi \end{matrix},$$

is tested against the alternative $\mathcal{H}_1 : \beta$ unrestricted. For easy of interpretation the *DGP* in (17) is also given in VECM form

$$\begin{bmatrix} \Delta Y_{1t} \\ \Delta Y_{2t} \\ \Delta Y_{3t} \\ \Delta Y_{4t} \end{bmatrix} = \begin{bmatrix} \Delta Y_{2t} \\ -\Delta Y_{1t} \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} (\xi - 1) \\ 0 \\ 0 \\ 0 \end{bmatrix} (Y_{1t-1} - Y_{2t-1}) + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix}$$

From (17) it is easy to see that under the null hypothesis the variables Y_{1t} , and Y_{2t} enter into the cointegrated relationship with coefficients proportional to $(1, -1)$. This restriction matches the hypothesis of proportional co-movements of the two random variables. The *DGP* in (17) is similar to that used in Gonzalo (1994). Among others, Gonzalo considers a simple two dimensional *VAR* in which cointegration holds between the $I(1)$ series Y_{1t} and Y_{2t} in (17). The *DGP* used in Gonzalo allows for high control over the many parameters affecting the size distortion of Λ such as the speed of adjustment (ξ), the correlation between the innovations (η), and the volatility parameter (σ). A possible shortcoming, however, is that bivariate cointegrated *VARs* are rarely encountered in empirical applications. The *DGP* in (17) maintains high control over the experimental design while also having greater practical relevance. The experimental parameter space is $T \in (50, 100, 250, 500)$, $\xi \in (0.2, 0.5, 0.8, 1)$, $\eta \in (-0.5, 0, 0.5)$, $\sigma = 1$. In addition, combinations of these parameters with alternative distributions of ε_t are considered.

Although non-normality is not a feature confined to financial data, it is the financial literature that has extensively documented substantial departures from the assumption of Gaussian innovations. For example, it is well established that the unconditional distributions of returns from financial market variables such as equity prices and interest rates are characterized by non-normality. Equity returns tend to be negatively skewed, whereas the patterns of skewness for bond market yields are more varied. Non-normality of the marginal distributions of returns does not necessarily imply the non-normality of the conditional distributions, but many empirical studies suggest that for financial data the Gaussian distribution is highly counterfactual. Given the widespread use of Johansen's procedure in financial applications, it seems appropriate to consider innovation distributions that better describe the behavior of financial markets.

To illustrate the problem of non-normality in financial market variables, we use the behavior of exchange rates. It is well known that exchange rate changes do not follow a Gaussian distribution. Potentially important sources of non-zero skewness and excess kurtosis are recurrent periods of a volatile and then quiet currency markets. To mimic the jump-like behavior caused by the volatility clustering of exchange rates, several researchers have allowed the innovations to be drawn from fat tailed distributions. Among others, Tucker and Pond (1988) provide evidence on the descriptive validity of the mixture of normal distribution as a statistical model for currency markets. Hull and

White (1998) give indications on the choice of parameters of the mixture of normals that match the higher-order moments of exchange rate changes for a number of major trading currencies. Building on these studies we allow the innovations to be drawn from these distributions.¹ The *DGP* in (5) has

$$\varepsilon_{s,t} \sim i.i.d. \omega_1 N(\mu_1, \sigma_{M1}^2) + \omega_2 N(\mu_2, \sigma_{M2}^2) \quad \text{and} \quad \varepsilon_{l,t} \sim N(0, 1) \quad (18)$$

with $\omega_1\mu_1 + \omega_2\mu_2 = 0$, $\omega_1, \omega_2 \geq 0$, and $\omega_1 + \omega_2 = 1$. Volatility clustering is introduced in (5) by ω_1 that causes occasional "jumps" in the innovation process of the cointegrated VAR(1).

When $\mu_1 = \mu_2 = 0$ the zero skewness assumption about $\varepsilon_{s,t}$ is preserved, being the means of the normal distributions mixed at zero. In this case, excess kurtosis has been introduced in (17) by choosing $\omega_1 < \omega_2$ and $\sigma_{M1}^2 > \sigma_{M2}^2$. Under this assumption the kurtosis in $\varepsilon_{s,t}$ is strictly increasing according to the quantity $\frac{\sigma_{M1}^2}{\sigma_{M2}^2}$. Consistent with these considerations, the following five distributions of $\varepsilon_{s,t}$ have been investigated

$$D_1 : \varepsilon_{s,t} \sim i.i.d. \quad 0.15N(0, 3.1329) + 0.85N(0, 0.6084),$$

$$D_2 : \varepsilon_{s,t} \sim i.i.d. \quad 0.52N(-2, 1.5876) + 0.48N(2.18, 0.3721),$$

$$D_3 : \varepsilon_{s,t} \sim i.i.d. \quad 0.7N(2, 1.4161) + 0.3N(-4.7, 0.0196),$$

$$D_4 : \varepsilon_{s,t} \sim i.i.d. \quad 0.32N(4, 1.9881) + 0.7N(-1.9, 0.5329).$$

Table 1 summarizes the descriptive statistics for D_1 , D_2 , D_3 and D_4 . Note that the skewness coefficient, (*Skew*), is computed as the third theoretical sample moment standardized by three halves power of the variance, whereas the kurtosis coefficient, (*Kurt*), is the fourth theoretical sample moment divided by the square of the variance. For a normal distribution *Skew* should be zero and *Kurt* should be equal to three.²

Table 1. Descriptive measures of D_1 , D_2 , D_3 , and D_4 .

¹ In a related paper Canepa (2016) considers a large range of distributions from heavily fat tailed (i.e. Student-t distributions with 3 to 30 degree of freedom) to highly skewed innovations (χ^2 distribution with 3 to 30 degrees of freedom), in order to disentangle the effect of skewness and kurtosis on the LR test. Simulation results in Canepa (2016) are obtained by response surface regression.

² Note: The second, third and fourth central moments of ε_{it} are calculated as $E(\varepsilon_{it}^2) = \sum_{b=1}^2 \omega_b [\sigma_b^2 + \mu_b^2]$, $E(\varepsilon_{it}^3) = \sum_{b=1}^2 \omega_b [3\mu_b\sigma_b^2 + \mu_b^3]$, and $E(\varepsilon_{it}^4) = \sum_{b=1}^2 \omega_b [3\sigma_b^4 + 6\mu_b^2\sigma_b^2 + \mu_b^4]$ (for $b = 1, 2$), respectively.

	D_1	D_2	D_3	D_4
<i>Mean</i>	0.00	0.00	0.00	0.00
<i>Var</i>	1.08	3.28	3.82	6.29
<i>Skew</i>	0.00	-0.50	-2.65	1.34
<i>Kurt</i>	4.61	4.49	12.83	4.16

As it emerges from Table 1, the innovations generated using mixture of normals cover a broad range of fat tailed and skewed distributions. Innovations generated under D_1 have mildly fat tails but are not skewed, whereas D_2 , D_3 and D_4 are fat tailed and skewed distributions.

Though mixture of normals introduces fat tails, it preserves the i.i.d. structure of the innovations. Among others, Bollerslev (1987) suggests that ARCH and GARCH models better fit exchange rate data measured over short time intervals (i.e. daily or weekly). Accordingly, simulations with conditional heteroskedastic innovations have been carried out with

$$\varepsilon_{1t} = \sqrt{h_t}\varsigma_{1t}, \quad \varepsilon_{2t} = \varkappa\varepsilon_{1t} + \sqrt{(1 - \varkappa^2)}\varsigma_{2t}, \quad \varepsilon_{3t} \text{ and } \varepsilon_{4t} \sim N(0, 1),$$

with $\varsigma_{i,t} \sim N(0, 1)$ (for $i = 1, 2$) and h_t denotes the conditional variance. Two specifications of the variance schemes are used: an ARCH(1) process given by

$$h_t = \frac{\sigma}{1 - \varrho} + \varrho\varepsilon_{1t-1}^2, \quad (18)$$

with $\sigma = 1$, and a GARCH(1, 1) process given by

$$h_t = \psi_0 + \psi_1\varepsilon_{1t-1}^2 + \psi_2h_{t-1}. \quad (19)$$

with $\psi_0 = 0.1$. As for the choice of the other parameters, the following values have been selected

D_5 : $\varepsilon_{s,t}$ as in (7) with $\varrho = 0.4$;

D_6 : $\varepsilon_{s,t}$ as in (7) with $\varrho = 0.8$;

D_7 : $\varepsilon_{s,t}$ as in (8) with $\psi_1 = 0.570$ and $\psi_2 = 0.921$;

D_8 : $\varepsilon_{s,t}$ as in (8) with $\psi_1 = 0.095$ and $\psi_2 = 0.881$.

The parameter values in D_7 and D_8 are those estimated for the exchange rate markets in Bollerslev (1987).

Estimates of the rejection probabilities have been obtained using pseudo-random numbers with programs written in GAUSS.³ The Monte Carlo experiment was based on $N = 10,000$ replications

³ The code used for the Monte Carlo simulation experiment will be available upon request.

for Λ , Λ_B and on $N = 1,000$ replications for Λ_B^* and Λ^* . All bootstrap distributions have been generated by resampling and then calculating the test statistic 800 times. The random number generator was restarted for each T value with the initial value set equal to zero. The VAR(1) model was fitted with an unrestricted constant. Moreover, note that in the Johansen procedure, the maximum likelihood estimator of β in equation (2) is calculated as the set of eigenvectors corresponding to the s largest eigenvalues of $S'_{0k}S_{00}^{-1}S_{0k}$ with respect to S_{kk} , where S_{00} , S_{kk} and S_{0k} are the moment matrices formed from the residuals Δy_t and y_{t-k} , respectively, onto the Δy_{t-j} . In this paper in place of the conventional algorithm for cointegration analysis (i.e. the algorithm for maximum likelihood estimation that uses the second moment matrices), all simulation results reported have been obtained using an algorithm based on QR decomposition; see Doornik and O'Brien (2002). This yields simulation results that are more numerically stable.

4 The Monte Carlo results

Tables 2-5 report the simulation results on the performance of Λ , Λ_B , Λ_B^* , and Λ^* . The finite sample significance levels are estimated for nominal levels of 5% and all estimates are given as percentages. In Table 2, the normal distribution serves as a benchmark, whereas Table 3 shows the results for the case of innovations drawn from a mixture of two normal distributions. Table 3 also contains results relating to the sensitivity of the error in rejection probability to variations of key parameters of the DGP . Finally, Table 4 reports the rejection frequencies for the case of ARCH and GARCH innovations.

Before looking at other specifics of the simulation results it is noteworthy to consider the benchmark case in which $\varepsilon_t \sim N(0,1)$, (note that in this case the distribution of $\varepsilon_{s,t} = \varepsilon_{t,t}$ in (17)). As far as Λ is concerned, Table 2 mainly confirms previous findings that inference based on first order asymptotic critical values is markedly inaccurate with excessively high rejection frequencies. Correcting Λ using the analytical Bartlett factor improves the behavior of the test statistic. However, Table 2 indicates that the performance of Λ_B is highly dependent on the autoregressive coefficient of the error correction mechanism, ξ . When ξ is large (i.e. the speed of adjustment to the cointegrated equilibrium is low), the correction does not work well. Using the bootstrap to approximate the Bartlett adjustment factor produces estimated levels that are less sensitive to the value of ξ parameter. The performance of the p -value bootstrap test is also less dependent on the value of the speed of adjustment parameter. Looking at the simulation results in Table 2 it appears that when $T = 100$ and $\eta \neq 0$, Λ^* and Λ_B^* work well for $\xi \leq 0.8$, whereas the empirical levels of Λ_B are within the 95% confidence interval for $\xi \leq 0.5$, say. When $\xi = 1$ the process Y_t is a pure $I(1)$ process that does not cointegrate. In this case, we do not expect the

resampling schemes presented in Section 2 to work, since the roots of the characteristic polynomial of the model in (2) are located inside the unit circle, and the process $Y_t^* - E(Y_t^*)$ is not stationary. The size distortion of Λ_B^* and Λ^* is still quite moderate, but there is no reason to believe that the test statistics would have adequate power. (Note that for the near unit-root model the bootstrap becomes inconsistent just as the exact unit root case). Coming to η , the estimated sizes reported in columns 3-6 show that the error in rejection probability increases when $\eta \rightarrow 0$. However, no matter the value of η , bootstrap based inference outperforms Λ_B .

Table 2. Estimated rejection probabilities, for the 5% critical value (in percent). Case with $N(0, 1)$ innovations.

		$\xi = 0.8$			$\eta = 0.5$		
	<i>Test</i>	$\eta = -0.5$	$\eta = 0$	$\eta = 0.5$	$\xi = 0.2$	$\xi = 0.5$	$\xi = 1$
$T = 50$	Λ	25.7	29.5	26.1	9.2	11.9	41.9
	Λ_B	14.5 (1.559)	17.3 (1.563)	14.4 (1.565)	4.5 (1.407)	5.6 (1.473)	25.3 (1.597)
	Λ_B^*	7.4	7.9	7.3	5.1	5.3	11.2
	Λ^*	8.1	8.3	8.6	5.1	5.7	13.1
$T = 100$	Λ	12.8	16.1	12.8	6.8	7.9	40.8
	Λ_B	8.4 (1.290)	11.0 (1.281)	8.7 (1.280)	4.8 (1.207)	5.1 (1.233)	32.1 (1.298)
	Λ_B^*	5.5	6.0	5.2	4.9	5.1	11.1
	Λ^*	5.4	5.9	5.5	4.8	5.3	13.0
$T = 250$	Λ	7.7	8.7	7.9	5.6	5.8	40.6
	Λ_B	6.2 (1.090)	7.0 (1.112)	6.1 (1.116)	5.0 (1.079)	4.7 (1.097)	36.4 (1.120)
	Λ_B^*	4.6	5.1	5.0	5.3	5.0	11.0
	Λ^*	4.8	5.1	5.2	5.4	5.2	12.7
$T = 500$	Λ	5.7	6.7	6.9	5.4	5.3	25.7
	Λ_B	5.2 (1.003)	6.4 (1.092)	5.2 (1.036)	5.0 (1.054)	4.8 (1.031)	20.3 (1.110)
	Λ_B^*	5.2	5.2	5.0	5.4	5.0	8.00
	Λ^*	5.1	5.0	5.1	5.4	5.1	9.02

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

Turning to the question of assessing how good is the bootstrap Bartlett correction when the innovations are fat-tailed, Table 3 suggests that the answer depends in a complicated way on ξ ,

η , T and the distribution of $\varepsilon_{s,t}$. Looking at the estimated levels of Λ over the range D_1, \dots, D_4 in the first place, a match with the excess kurtosis and skewness coefficients in Table 2 reveals that, in general, the error in the rejection probability of the test increases with $|Kurt|$ and $|Skew|$, with the highest size distortion for the case of D_3 and D_4 . Furthermore, comparing the estimated sizes from the top and the bottom panel in Table 3 it appears that the effect of non-Gaussian innovations on the estimated level of the test is, once again, 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 latter is fast (i.e., $\xi = 0.2$). Bewley and Orden (1994) report that Johansen's estimator β produces outliers when the speed of adjustment is slow, while Phillips (1994) provides a theoretical analysis showing that the finite sample distribution of β is leptokurtic. The simulations in Bewley and Orden and the theoretical results in Phillips explain why Λ behaves so poorly when the combinations of $\xi = 0.8$ and the non-Gaussian distributions in Table 3 are selected: 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, Λ_B can only be partially successful because the second terms of the asymptotic expansions of the mean of Λ 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 Λ_B^* the Gaussian distribution is replaced with the empirical density function of the innovations. This strongly mitigates the effects of skewness and kurtosis on the finite sample mean of the test and makes the finite sample distribution of Λ_B^* closer to the asymptotic distribution.

Table 3. Empirical sizes for the 5% (in percent) critical value. Case with mixture of normal innovations.

		$\xi = 0.8$							
		$\eta = -0.5$				$\eta = 0.5$			
$\varepsilon_{s,t}$		Λ	Λ_B	Λ_B^*	Λ^*	Λ	Λ_B	Λ_B^*	Λ^*
$T = 50$	D_1	31.0	18.3 (1.563)	8.6	9.3	31.6	18.4 (1.566)	8.5	9.7
	D_2	28.0	15.8 (1.564)	8.2	8.8	28.9	16.3 (1.561)	8.4	9.1
	D_3	32.6	19.1 (1.562)	8.3	9.0	33.2	19.9 (1.559)	8.8	10.5
	D_4	34.8	20.9 (1.562)	9.3	10.9	34.5	20.3 (1.556)	8.6	10.4
$T = 100$	D_1	18.5	13.2 (1.276)	5.4	6.1	19.0	13.2 (1.281)	6.2	6.7
	D_2	15.4	10.7 (1.281)	5.2	5.4	15.8	10.7 (1.283)	6.0	6.1
	D_3	21.0	14.8 (1.283)	6.5	6.7	21.7	15.4 (1.279)	5.8	7.0
	D_4	23.8	17.7 (1.281)	7.2	8.0	23.4	17.0 (1.281)	5.8	6.1
$T = 250$	D_1	8.6	7.1 (1.107)	4.8	4.6	8.8	7.3 (1.114)	4.7	4.7
	D_2	7.7	6.0 (1.118)	5.2	5.1	8.0	6.4 (1.115)	4.5	4.5
	D_3	9.6	8.0 (1.109)	5.7	5.3	9.6	8.0 (1.114)	5.2	5.4
	D_4	11.1	9.0 (1.113)	4.4	4.2	10.6	8.9 (1.116)	4.1	4.3
$T = 500$	D_1	5.3	5.2 (1.055)	4.9	4.9	5.5	5.3 (1.011)	4.9	4.8
	D_2	5.5	5.0 (1.061)	5.1	5.4	5.1	5.0 (1.017)	4.9	4.8
	D_3	5.4	4.9 (1.009)	5.5	5.2	5.4	5.1 (1.011)	5.2	5.0
	D_4	4.8	6.0 (1.019)	4.9	4.7	5.7	5.9 (1.014)	5.1	5.3

Table 3. (Continue)

		$\eta = 0.5$							
		$\xi = 0.2$				$\xi = 0.5$			
$\varepsilon_{s,t}$		Λ	Λ_B	Λ_B^*	Λ^*	Λ	Λ_B	Λ_B^*	Λ^*
$T = 50$	D_1	10.1	5.3 (1.394)	4.8	4.6	15.2	7.9 (1.471)	6.3	6.9
	D_2	9.5	5.0 (1.399)	4.4	5.6	13.0	6.4 (1.473)	6.2	5.8
	D_3	11.0	5.1 (1.396)	5.9	5.8	17.3	9.5 (1.474)	6.0	6.3
	D_4	12.3	6.2 (1.397)	5.9	5.8	18.7	10.1 (1.473)	6.1	6.6
$T = 100$	D_1	7.8	5.4 (1.203)	4.0	4.2	9.2	6.1 (1.239)	5.9	4.6
	D_2	7.2	4.9 (1.200)	4.7	5.2	8.4	5.4 (1.238)	5.0	5.4
	D_3	7.5	5.1 (1.202)	5.1	5.0	9.6	6.4 (1.235)	5.4	5.6
	D_4	8.1	5.6 (1.113)	4.6	4.7	10.5	7.3 (1.235)	5.4	4.8
$T = 250$	D_1	5.8	5.0 (1.079)	4.9	4.7	6.5	5.4 (1.094)	5.0	4.5
	D_2	5.7	4.9 (1.077)	4.7	5.0	6.2	5.0 (1.096)	5.1	4.6
	D_3	6.1	5.2 (1.079)	4.7	4.4	6.7	5.5 (1.095)	4.5	4.7
	D_4	6.1	5.2 (1.078)	4.8	4.7	6.9	5.7 (1.096)	4.5	4.6
$T = 500$	D_1	5.2	5.5 (1.017)	5.1	5.0	5.7	5.2 (1.006)	4.9	5.2
	D_2	5.1	5.2 (1.031)	5.0	5.2	5.6	5.0 (1.001)	5.0	4.9
	D_3	5.4	5.0 (1.019)	4.9	4.7	5.2	5.1 (1.031)	4.9	4.9
	D_4	5.5	5.1 (1.012)	4.9	5.0	5.3	5.4 (1.001)	4.9	5.0

Note: The estimated rejection probabilities of Λ_B^* and Λ^* have been calculated using algorithm 1 in Section 2. For Λ and Λ_B the number of replications is $N=10,000$, for Λ_B^* and Λ^* $N=1,000$ and $B=800$. The Bartlett corrections are given in parenthesis.

The final set of simulation experiments relates the ARCH and GARCH innovations. Table 4 presents the empirical sizes for the inference procedure under consideration when different values of ϱ , ψ_1 and ψ_2 are considered. As for the other cases, the error in rejection probability of Λ and Λ_B heavily depend on the distribution of $\varepsilon_{s,t}$. In contrast, Λ_B^* and Λ^* behave quite well leaving open the possibility of extending the bootstrap algorithm presented for the Bartlett correction in Section 2 to other cases in which the ordinary residual based bootstrap procedure would fail.

Table 4. Empirical sizes (in percent) for the 5% critical value. Case with ARCH and GARCH innovations.

		$\xi = 0.5$				$\xi = 0.8$			
$\varepsilon_{s,t}$		Λ	Λ_B	Λ_B^*	Λ^*	Λ	Λ_B	Λ_B^*	Λ^*
$T = 50$	D_5	16.7	8.8 (1.474)	5.3	5.6	33.0	19.6 (1.562)	7.7	8.4
	D_6	19.4	10.6 (1.472)	6.0	6.2	34.8	21.1 (1.560)	8.1	9.3
	D_7	16.9	9.0 (1.476)	5.7	5.5	31.2	18.6 (1.561)	8.8	9.8
	D_8	16.5	8.8 (1.468)	5.5	5.5	31.8	18.7 (1.562)	8.6	9.3
$T = 100$	D_5	8.5	5.6 (1.241)	5.2	5.5	19.3	13.5 (1.280)	6.9	7.4
	D_6	9.5	6.4 (1.235)	5.0	5.2	22.6	16.3 (1.279)	7.7	8.2
	D_7	9.6	6.4 (1.234)	5.4	5.1	20.9	14.8 (1.281)	6.8	6.9
	D_8	9.4	6.3 (1.238)	5.2	5.3	20.4	14.3 (1.283)	6.8	6.4
$T = 250$	D_5	6.4	5.2 (1.095)	4.4	4.6	9.3	7.6 (1.113)	4.9	5.4
	D_6	6.5	6.4 (1.094)	4.6	4.4	10.2	8.6 (1.111)	5.1	5.0
	D_7	6.4	5.6 (1.096)	5.2	5.1	9.6	7.9 (1.106)	4.5	4.7
	D_8	6.7	5.7 (1.096)	5.0	5.1	9.5	7.7 (1.110)	4.9	4.8
$T = 500$	D_5	5.5	5.0 (1.005)	5.1	4.9	5.3	5.1 (1.011)	5.1	5.2
	D_6	5.1	5.6 (1.094)	4.9	4.6	5.4	5.2 (1.010)	5.3	5.2
	D_7	5.0	5.9 (1.003)	5.3	5.2	6.3	5.5 (1.089)	5.0	4.9
	D_8	5.6	5.2 (1.008)	5.1	5.1	6.5	5.5 (1.080)	4.9	4.9

Note: The estimated rejection probabilities of Λ_B^* and Λ^* have been calculated using algorithm 2 in Section 2. DGP with $\eta = 0$. For Λ and Λ_B the number of replications is $N=10,000$, for Λ_B^* and Λ^* $N=1,000$ and $B=800$. The asymptotic size of the tests is 5%. The Bartlett corrections are given in parenthesis.

To wrap up the discussion, in Tables 2-4, Λ is greatly oversized in most instances. The error in rejection probability of the test statistic crucially depends on the parameter values of the DGP , and violations of the Gaussian assumption worsen the performance of the test for finite samples. Λ_B offers improvements over the uncorrected statistic but its behavior mimics the performance of Λ and thus, it is not entirely reliable. In contrast, the two bootstrap procedures are less sensitive to the parameter values of the DGP and appear to be relatively robust to both non-Gaussian and conditionally heteroskedastic innovations.

4.1 Results Under the Alternative Hypothesis

It is well known that the Bartlett correction factor is designed to bring the actual size of asymptotic tests close to their respective nominal size, but it may lead to a loss in power. Accordingly, the power properties of the proposed procedure are considered in this section.

For the experiments evaluating the power of the tests, data were generated under the alternative hypothesis

$$\mathcal{H}_1 : \beta = H\varphi = \begin{bmatrix} 1 & 0 \\ -g & 0 \\ 0 & I \end{bmatrix} \begin{matrix} (1 \times 2) \\ (1 \times 2) \\ (2 \times 2) \end{matrix} \varphi, \quad (3 \times 1)$$

where $g \in (1.2, 1.4, 1.6, 1.8, 2)$ with $\xi = 0.5, \eta = 0.5$, and $\sigma = 1$ in (17). The results of this set of experiments are reported in Table 5. Once again, the case of ε_t , i.i.d. $N(0, 1)$ (i.e the case with $\varepsilon_{s,t} = \varepsilon_{l,t}$ in (17)) serves as a benchmark, then $\varepsilon_{s,t} \sim D_1$ and $\varepsilon_{s,t} \sim D_5$ are considered. Experiments using the other distributions for the innovations considered in Table 3-4 produced similar power properties and results will be omitted in the interest of brevity. Note that simulation results were obtained using algorithm 1, for $\varepsilon_t \sim N(0, 1)$ and $\varepsilon_{s,t} \sim D_1$, whereas algorithm 2 was used for $\varepsilon_{s,t} \sim D_5$.

Simulation experiment results under the alternative hypotheses are presented in Table 5. The simulation results show that the sample size and the distance between the null and the alternative hypothesis play an important role in determining the power of the test statistics under consideration.⁴ Considering the asymptotic test first, it appears that the power of the Λ is badly affected by the choice of the distribution of the innovations: the test is relatively well behaved when the innovation are fat-tailed but i.i.d., whereas the performance of the test deteriorates when ARCH innovations are introduced in the *DGP*. Turning to the comparison of the power among the different procedures, overall it is found that in small samples (i.e. $T = 50$) correcting the test statistic for the size shifts the estimated power function down. There is evidence that Λ_B^* and Λ^* share similar power properties, with no test uniformly outperforming its competitor. The results for the sensitivity of the inference procedures to the parameters of the *DGP* are not reported in detail here but simulation experiment showed that a slow adjustment to the equilibrium worsens the rejection frequencies for Λ_B^* , Λ^* and Λ_B . On the other side, changing the correlation between the noises does not have an important impact on the power estimates.

⁴ Note that in Table 5 simulation results for the sample size $T=500$ are not reported since the 100% rejection frequency was achieved no matter the innovation distribution under consideration.

Table 5. Rejection frequencies under the alternative hypothesis (in percentage) for the 5% critical value.

ε_t		$g = 1.2$	$g = 1.4$	$g = 1.6$	$g = 1.8$	$g = 2$	
$T = 50$	N	Λ	71.7	89.8	93.8	95.2	95.7
		Λ_B	60.7	84.0	89.6	91.6	92.5
		Λ_B^*	60.4	83.6	88.7	90.5	90.8
		Λ^*	60.6	83.0	89.2	90.6	91.0
	D_1	Λ	50.0	72.9	80.7	83.9	85.5
		Λ_B	37.0	61.6	72.0	75.7	77.5
		Λ_B^*	29.3	53.3	59.7	63.9	65.0
		Λ^*	30.2	54.5	61.5	64.8	66.4
	D_5	Λ	25.3	41.0	52.6	58.8	62.7
		Λ_B	15.7	28.3	39.3	46.5	50.3
		Λ_B^*	11.4	21.5	29.1	33.2	36.6
		Λ^*	11.1	24.4	29.5	33.9	37.3
$T = 100$	N	Λ	98.2	99.9	100	100	100
		Λ_B	97.5	99.8	99.9	100	100
		Λ_B^*	96.8	99.0	100	100	100
		Λ^*	96.8	99.7	100	100	100
	D_1	Λ	86.1	98.1	99.4	99.7	99.8
		Λ_B	82.1	97.3	99.0	99.5	99.7
		Λ_B^*	80.1	96.7	99.0	99.3	99.4
		Λ^*	79.6	96.7	98.9	99.3	99.4
	D_5	Λ	44.5	77.2	89.1	93.6	95.3
		Λ_B	37.7	72.1	85.9	91.4	93.6
		Λ_B^*	32.7	66.7	80.9	87.2	89.5
		Λ^*	32.5	66.5	80.9	87.6	89.5
$T = 250$	N	Λ	99.9	100	100	100	100
		Λ_B	99.9	100	100	100	100
		Λ_B^*	99.8	100	100	100	100
		Λ^*	99.8	100	100	100	100
	D_1	Λ	96.3	100	100	100	100
		Λ_B	95.8	100	100	100	100
		Λ_B^*	95.7	100	100	100	100
		Λ^*	96.7	100	100	100	100
	D_5	Λ	93.0	99.8	100	100	100
		Λ_B	92.0	99.8	100	100	100
		Λ_B^*	91.9	99.7	100	100	100
		Λ^*	91.9	99.7	100	100	100

Note: The estimated rejection probabilities of Λ_B^* and Λ^* have been calculated using algorithm 1 and 2 in

Section 2 using $N=1000$ and $B=800$. DGP with $\eta = 0.5$, $\xi = 0.5$.

5 Robustness Check

Previous simulation results revealed several insights on the performance of the Bootstrap Bartlett corrected test. A possible shortcoming is that *DGPs* with only one cointegration vector and a limited number of lags were considered. Increasing the number of nuisance parameters in the *DGP* is expected to increase the finite sample size distortion of the asymptotic LR test (see for example Podivinsky, 1992). To investigate the effect of the nuisance parameters on the finite sample performance of the inference procedures a further set of simulation experiments was undertaken. In this case the simulation design involved: *i*) considering the effect of increasing the number of lags in the *DGP*; *ii*) increasing the cointegrating rank from $r = 1$ to $r = 2$. To complete the picture, the effect of including both greater dynamic and increasing the cointegration rank was also considered.

In particular, point *i*) was addressed by undertaking simulation experiments using the *DGP* as in Eq. (17) with $\eta = 0.5$, $\xi = 0.5$, and $N(0, 1)$ innovations. To investigate the effect of the nuisance parameters the number of lags was progressively increased from $k = 1$ to $k = 3$. To address point *ii*) a second *DGP* with $r = 2$ was considered. The second *DGP*, labelled as *DGP2*, is a 4-dimensional VAR given by

$$\begin{bmatrix} \Delta X_{1t} \\ \Delta X_{2t} \\ \Delta X_{3t} \\ \Delta X_{4t} \end{bmatrix} = \begin{bmatrix} \alpha_{11} & 0 \\ 0 & \alpha_{22} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_{11} & \beta_{21} & \beta_{31} & 0 \\ 0 & \beta_{22} & \beta_{32} & 0 \end{bmatrix} \begin{bmatrix} X_{1t-1} \\ X_{2t-1} \\ X_{3t-1} \\ X_{4t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \end{bmatrix}$$

where $\alpha_{11}, \alpha_{22} = 1$, $\beta_{11} = 0.4, \beta_{21} = 1.7, \beta_{31} = 0.1, \beta_{22} = 0.3, \beta_{32} = 0.6$ and the $\varepsilon_{i,t}$ as in *GDP1*. In this case the null hypothesis of interest is

$$\mathcal{H}_0 : \beta = H\varphi = \begin{bmatrix} I \\ (2 \times 2) \\ 0 \\ (2 \times 2) \end{bmatrix} \begin{matrix} \varphi \\ \\ \\ \end{matrix},$$

where I is an identity matrix. Once again, under alternative hypothesis β is unrestricted.

Table 6 reports the simulation results for the empirical sizes of Λ , Λ_B , Λ_B^* and Λ^* . For ease of interpretation the empirical sizes in Table 2 for $k = 1$ are also reported.

Table 6. Estimated rejection probabilities, for the 5% critical value (in percent).

			Λ	Λ_B	Λ_B^*	Λ^*
$T = 50$	<i>GDP1</i>	$k = 1$	11.9	5.6 (1.473)	5.3	5.7
		$k = 2$	24.8	11.9 (1.671)	10.2	10.8
		$k = 3$	39.2	15.2 (1.674)	14.3	14.4
	<i>GDP2</i>	$k = 1$	11.3	6.2 (1.544)	6.1	6.2
		$k = 2$	20.6	11.7 (1.573)	10.4	10.5
		$k = 3$	32.2	12.7 (1.515)	12.3	12.6
$T = 100$	<i>GDP1</i>	$k = 1$	7.9	5.1 (1.233)	5.1	5.3
		$k = 2$	11.5	7.9 (1.333)	6.7	7.1
		$k = 3$	15.9	8.3 (1.359)	7.3	7.4
	<i>GDP2</i>	$k = 1$	7.7	5.2 (1.423)	5.5	5.6
		$k = 2$	11.2	7.8 8 (1.231)	7.1	7.4
		$k = 3$	14.9	10.0 (1.451)	10.1	9.8
$T = 250$	<i>GDP1</i>	$k = 1$	5.8	4.7 (1.097)	5.0	5.2
		$k = 2$	7.0	5.7 (1.008)	5.0	5.3
		$k = 3$	10.1	6.2 (1.124)	6.6	5.9
	<i>GDP2</i>	$k = 1$	6.1	5.1 (1.002)	6.0	5.6
		$k = 2$	8.1	6.2 (1.182)	6.4	6.2
		$k = 3$	10.1	6.8 (1.183)	7.1	7.2
$T = 500$	<i>GDP1</i>	$k = 1$	5.3	4.8 (1.031)	5.0	5.1
		$k = 2$	5.6	5.2 (1.034)	5.1	5.3
		$k = 3$	5.4	5.3 (1.035)	5.3	5.2
	<i>GDP2</i>	$k = 1$	5.5	5.2 (1.031)	5.3	5.0
		$k = 2$	5.7	5.3 (1.032)	5.5	5.5
		$k = 3$	5.6	5.2 (1.031)	5.6	5.4

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

As far as the simulation results are concerned the first thing to note in Table 6 is that inference based on first order asymptotic critical values is, once again, markedly inaccurate with excessively high rejection rates. Increasing the number of lags, k , dramatically increases the deviation from the nominal levels. By contrast, allowing more cointegrating vectors, r , in the system slightly reduces the size distortion. Turning to empirical sizes for Λ_B , and Λ_B^* , we can see that they are much closer to the nominal sizes than the first order asymptotic critical values. Overall, the results in Table 6 show that correcting the LR test statistic is worthwhile, since all the empirical sizes reported for the corrected test are closer to the nominal 5% level than the unadjusted test statistic. However, introducing many nuisance parameters in the model affects the size accuracy of all test statistics under consideration.

Note that in Table 6 only results with normal innovation are reported. In line with the results in Table 3, departures from the normal distribution of the innovations increased the size distortion of the Λ and Λ_B procedures, whereas Λ^* and Λ_B^* presented empirical sizes closer to the nominal size (results not reported but available upon request).

6 An Empirical Application

As an illustration, the bootstrap Bartlett procedure discussed in Section 2 has been applied to investigate purchasing power parity (PPP) relationship. According to economic theory, once converted to a common currency, national price levels should be equal. In other words,

$$P = \ddot{P} + E$$

where P is the log of the domestic price level, \ddot{P} is the log of the foreign price level, and E denotes the log of the spot exchange rate (home currency price of a unit of foreign currency). Therefore, departures from PPP relationship at time t can be defined as

$$PPP_t = P_t - \ddot{P}_t - E_t. \tag{20}$$

Equation (20) implies that if the PPP mechanism is functioning, one should observe the tendency of the two markets to adjust toward the long-run equilibrium level of exchange rates, meaning that PPP_t should be a stationary stochastic process. However, using conventional unit root tests a number of studies examining the empirical validity of the PPP relationship for the period of floating exchanges rates have failed to reject the null hypothesis of non-stationarity for PPP_t leading to what Obstfeld and Rogoff (2001) define as the "PPP puzzle" (see also Rogoff, 1996). The progress in econometrics over the past thirty years added hundreds of papers to the PPP's

literature. Failure to find empirical evidence of the PPP relationship in the empirical literature has identified the PPP puzzle as one of the six major puzzles in international economics (see for example Razzak, 2018; Chen and Engel, 2005; Frydman *et al.*, 2009; Imbs *et al.*, 2005; Ford and Horioka, 2017; Horioka and Ford, 2017; Falahati, 2019; Johansen, and Juselius, 1992).

Reviewing the existing empirical works on PPP a large consensus on two facts emerges. First, consensus estimates suggest that the marginal distributions of prices and exchange rates exhibit excess kurtosis and nonzero skewness such that a Gaussian conditional distribution for the innovations is typically counterfactual (see for example Tucker and Pond (1988), Fujihara and Park (1990), Engel and West, 2005). Second, there is fairly persuasive evidence that it takes long time before PPP returns to its steady-state value, meaning that speed of adjustment toward PPP equilibrium is very slow (see for example Alves *et al.*, 2001; Masih and Masih, 2004; Costa and Crato, 2001). Because Λ_B^* is less sensitive to parameter values of the *DGP* (the empirical levels of Λ_B^* reported in Table 2 showed much less variation over the grid of parameters considered in the Monte Carlo experiment) and better able to cope with deviations from the Gaussian assumption, this test statistic may be appropriate when using Johansen’s procedure for testing PPP hypotheses.

As an application we test whether or not the PPP relationship holds for the real exchange rate between the US dollar and the currency of a number of countries (economic regions) using quarterly data from 2000:Q1 to 2020:Q1. Namely, the countries included in the sample are Canada, UK, China, Australia and the EU area.⁵ For the h -country ($h = 1, \dots, 5$) let $E_{h,t}$ be the nominal dollar exchange rate, $P_{h,t}$ the domestic consumer price index, and the US consumer price index ($\ddot{P}_{USA,t}$).

As for the estimation results, preliminary analysis on the unrestricted VAR(2) models ruled out serial correlation however, evidence of non normality and heteroskedasticity of the ARCH type was detected for the series $(P_{h,t} - \ddot{P}_{USA,t})$ in all models under consideration. Under these circumstances, the test Λ_B no longer constitutes a valid inference procedure. For this reason Λ_B was discarded from the analysis.

In Table 7, the empirical p -values for Λ , Λ_B^* and Λ^* are reported. The null hypothesis under consideration is that $(P_{h,t} - \ddot{P}_{USA,t}) - E_{h,t}$ is stationary, or equivalently, that the vector $(1, -1)' \in sp(\beta)$. This can be formulated as the hypothesis $\mathcal{H}_0 : \beta = H\hat{\varphi}$ with

$$H' = [1 \quad -1]$$

versus the alternative $\mathcal{H}_1 : \beta$ unrestricted. Empirical levels for Λ_B^* and Λ^* in the third and fourth columns were obtained using algorithm 2 in Section 2 with $B = 5,000$. The p -values for Λ were calculated by taking the 95% percentile from the $\chi^2(1)$ and calculating the actual p -value as the

⁵ For further examples on the application of the Bootstrap Bartlett corrected LR test see Canepa 2020.

frequency of rejection.

Table 7. Actual rejection probabilities for the PPP relationship between the U.S. and other countries.

	Λ	Λ_B^*	Λ^*
<i>UK</i>	0.002	0.072	0.061
<i>AUSTRALIA</i>	0.030	0.151	0.132
<i>CANADA</i>	0.436	0.567	0.439
<i>EU AREA</i>	0.055	0.132	0.101
<i>CHINA</i>	0.809	0.901	0.911

From Table 7 it appears in finite samples the asymptotic LR inference procedure does not work well as it rejects the null hypothesis in favor of the alternative in three cases (i.e. UK, Australia and the EU area), whereas the corrected *LR* statistic confirms that the PPP relationship holds between the U.S. dollar and the other currencies for all countries (economic regions). The Λ^* also seems to work well thus providing an alternative test to the corrected LR test.

7 Concluding remarks

Johansen's (2000) Bartlett corrected *LR* test relies on Gaussian innovations. However, in empirical applications, there is limited information on the distributional form of the innovations. Therefore there is a need to investigate procedures that do not rest on the Gaussian assumption (or on any other specific distribution).

This paper considers a non-parametric bootstrap Bartlett *LR* test, and finds that the bootstrap Bartlett correction serves two purposes at once. First, it is able to control for the size distortion generated by a slow speed of adjustment to the cointegrated equilibrium as well as other crucial parameters of the data generating process. Second, it is robust to violations of the Gaussian assumption. No matter the distribution of innovations under consideration, (i.e., mixture of normals, ARCH or GARCH) there is little evidence that the size of the bootstrap Bartlett statistic depends in any important way on the form of innovations. Together, these results constitute an important improvement with respect to the analytical Bartlett correction, particularly in light of the fact that in empirical applications, the true underlying data generating process is not known.

References

- Alves, D. S. C. O., Cati, R. C. and Fava, V. L. (2001). "Purchasing Power Parity in Brazil". *Applied Economics*, 33, 1175–85.

- Andrews, D.W.K. (2001). "Testing when a Parameter is on the Boundary of the Maintained Hypothesis". *Econometrica*, 69, 721-762.
- Bartlett, M.S. (1937). "Properties of Sufficiency and Statistical tests". *Proceedings of the Royal Society*, 160, 263-268.
- Beran, R. (1988). "Prepivoting Test Statistic: A Bootstrap View of Asymptotic Refinements". *Journal of the American Statistical Association*, 83, 687-697.
- Bewley, R. and D. Orden (1994). "Alternative Methods of Estimating Long-Run Responses with Application to Australian Import Demand". *Econometric Review*, 13, 179-204.
- Bravo, F. (1999). "A Correction Factor for Unit Root Test Statistics". *Econometric Theory*, 15, 218-227.
- Brüggemann, R., C. Jentsch, and C. Trenkler, (2016). "Inference in VARs with Conditional Heteroskedasticity of Unknown Form". *Journal of Econometrics*, 191, 69-85.
- Bollerslev, T. (1987). "A Conditionally Heteroskedastic Time Series Model for speculative Prices and Rates of Returns", *The Review of Economics and Statistics*, 3, 542-547.
- Boswijk, H. P., G. Cavaliere, A. Rahbek and A. M. R. Taylor (2016). "Inference on Co-Integration Parameters in Heteroskedastic Vector Autoregressions". *Journal of Econometrics*, 192, 64-85.
- Canepa A., (2006). "Small Sample Corrections for Linear Restrictions on Cointegrating Vectors: a Monte Carlo Comparison". *Economics Letters*, 91, 330-336.
- 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.
- Canepa, A. (2016). "A note on Bartlett Correction Factor for Tests on Cointegrating Relations". *Statistics and Probability Letters*, 110, 295-304.
- Canepa, A. (2020) "Bootstrap Bartlett Adjustment for Hypotheses Testing on Cointegrating Vectors". *EST Working Papers 06/2020*, University of Turin.
- Chan, N. H., K. Chen, C. Y. Yau (2014). "On the Bartlett Correction of Empirical Likelihood for Gaussian Long-Memory Time Series". *Electronic Journal of Statistics*, 8, 1460-1490.
- Chan, N. H., L. Liu. (2010). "Bartlett Correctability of Empirical Likelihood in Time Series". *Journal of the Japanese Statistical Society*, 40, 1-5.
- Chen, K., N.H. Chan, C.Y. Yau (2016). "Bartlett Correction of Empirical Likelihood for non-Gaussian Short-Memory time series". *Journal of Time Series Analysis*, 37, 624-649.
- Chen, S. X., H. Cui (2006). "On Bartlett Correction of Empirical Likelihood in the Presence of Nuisance Parameters". *Biometrika*, 93, 215-220.
- Chen, S. X., H. Cui (2007). "On the Second Order Properties of Empirical Likelihood with Moment Restrictions". *Journal of Econometrics*, 141, 492-516.
- Chen, K., N.H. Chan, Yau, C. Y. (2020). "Bartlett Correction of Frequency Domain Empirical Likelihood for Time Series with Unknown Innovation Variance". *Annals of the Institute of Statistical Mathematics*, 72, 1159-1173.

- Chen, S.S., and C. Engel (2005). "Does Aggregation Biase Explain the PPP Puzzle". *Pacific Economic Review*, 10, 49-72.
- Cavaliere, G., A. Rahbek, and A.M.R. Taylor, (2010a). "Cointegration Rank testing Under Conditional Heteroskedasticity", *Econometric Theory*, 26, 1719-1760.
- Cavaliere, G., A. Rahbek, and A.M.R. Taylor, (2010b). "Testing for Co-integration in Vector Autoregressions with Non-stationary Volatility". *Journal of Econometrics*, 158, 7-24.
- Cavaliere, G., H.B. Nielsen, A. Rahbek, (2015). "Bootstrap Testing of Hypotheses on Co-Integration Relations in Vector Autoregressive Models". *Econometrica*, 83, 813-831.
- Cavaliere, G., H. B. Nielsen, R. S. Pedersen, A. Rahbek (2020). "Bootstrap Inference on the Boundary of the Parameter Space, with Application to Conditional Volatility Models". *Journal of Econometrics*, <https://doi.org/10.1016/j.jeconom.2020.05.006>.
- Costa, A. A. and N. Crato (2001). "Long-run versus shortrun behaviour of the real exchange rates", *Applied Economics*, 33, 683-88.
- Cribari-Neto, F. and G. Cordeiro (1996). "On Bartlett and Bartlett Type Corrections". *Econometric Review*, 15, 339-401.
- Davidson, R. and J. G. MacKinnon (2006). "The power of Bootstrap Tests and Asymptotic Test", *Journal of Econometrics*, 133, 421-441.
- Doornik, J. A. and R. J. O'Brien (2002). "Numerically Stable Cointegration Analysis", *Computational Statistics and Data Analysis*, 41, 185-193.
- Elliott, G., U., Müller, M., Watson (2015). "Nearly Optimal Tests When a Nuisance Parameter Is Present Under the Null Hypothesis". *Econometrica*, 83, 771-811.
- Engel, C. and K. D. West (2005). "Exchange Rates and Fundamentals". *Journal of Political Economy*, 113, 485-517.
- Engle, R. F. and J. G. Rangel (2008). "The Spline-GARCH Model for Low-Frequency Volatility and Its Global Macroeconomic Causes". *The Review of Financial Studies*, 21, 1187-1222.
- Engsted, T. and Pedersen, T. Q. (2014). "Bias-Correction in Vector Autoregressive Models: A Simulation Study", *Econometrics*, 2, 45-71.
- Falahati, K. (2019). "Time, Arbitrage, and the Law of One Price: The Case for a Paradigm Shift". *Journal of Economic Issues*, 53, 115-154.
- Ford, N., and C. Y. Horioka (2017). "The 'Real' Explanation of the Feldstein-Horioka Puzzle". *Applied Economics Letters*, 24, 325-328.
- Fujihara, R., and K. Park (1990). "The Probability Distribution Of Futures Prices In The Foreign Exchange Market: A Comparison Of Candidate Processes". *Journal of Futures Markets*, 623-641.
- Frydman, R, M. D. Goldberg, S. Johansen, K. Juselius (2009) "A Resolution of the Purchasing Power Parity Puzzle: Imperfect Knowledge and Long Swings". *SSRN Electronic Journal*, DOI: 10.2139/ssrn.1315233.
- Gine, E. and J. Zinn, (1990). "Bootstrapping General Empirical Measures". *The Annals of Probability*, 18, 851-869.

- Gonçalves, S. and L. Kilian, (2004). "Bootstrapping Autoregressions with Conditional Heteroskedasticity of Unknown Form". *Journal of Econometrics*, 123, 89-120.
- Gonzalo, J. (1994). "Comparison of Five Alternative Methods of Estimating Long-run Equilibrium Relationships". *Journal of Econometrics*, 60, 203-233.
- Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.
- Hansen, B. E. (1992). "Convergence of Stochastic Integrals for Dependent Heterogenous Processes". *Econometric Theory*, 8, 489-500.
- Harvey D., S. J. Leybourne, R. Sollis and A. M. R. Taylor (2016), "Tests for Explosive Financial Bubbles in the Presence of Non-Stationary Volatility". *Journal of Empirical Finance*, 38, 548–574.
- 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.
- Horioka, C.Y., N. Ford. (2017). "A Possible Explanation of the 'Exchange Rate Disconnect Puzzle': a Common Solution to Three Macroeconomic Puzzles?". *Applied Economics Letters*, 24, 918-922.
- Gredenhoff, M. and T. Jacobson (2001). "Bootstrap Testing Linear Restrictions on Cointegrating Vectors". *Journal of Business and Economics Statistics*, 19, 63-72.
- 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.
- Horowitz, J.L. and N.E. Savin (2000). "Empirically Relevant Critical Values for Hypothesis Tests: a Bootstrap Approach", *Journal of Econometrics*, 95, 375-389.
- Hull, J. and White, A. (1998). "Value at Risk When Daily Changes in Market Are Not Normally Distributed", *Journal of Derivative*, 5, 9-16.
- Hunter, J., Burke, S., and A. Canepa. (2017). *Modelling Multivariate non Stationary Economic Time Series*. 2nd Edition. Palgrave Text in Econometrics, Palgrave MacMillan, London.
- Imbs, J., H. Mumtaz, M. O. Ravan, and H. Rey, (2005). "PPP Strikes Back: Aggregation and The Real Exchange Rate". *Quarterly Journal of Economics*, 120, 1-43.
- Jacobson, T. and T. Larsson, (1999). "Bartlett Correction in Cointegration Testing", *Computational Statistics and Data Analysis*, 31, 203-225.
- Jentsch, C. and K. G. Lunsford (2019). "The Dynamic Effects of Personal and Corporate Income Tax Changes in the United States: Comment". *American Economic Review*, 109, 2655-78.
- Johansen, S. (1988). "Statistical Analysis of Cointegrated Vectors", *Journal of Economic Dynamic and Control*, 12, 231-254.
- Johansen, S. and K. Juselius, (1992). "Testing for Structural Hypotheses in a Multivariate Cointegration Analysis of PPP and UIP for UK", *Journal of Econometrics*, 53, 211-244.
- Johansen, S. (1996). *Likelihood Inference in Cointegrated Vector Auto-Regressive Models*. Oxford University Press, Oxford.

- Johansen, S. (2000). "A Bartlett Correction Factor for Tests of on the Cointegrating Relations". *Econometric Theory*, 16, 740-778.
- Lagos, B.M. and P.A. Morettin (2004). "Improvement of the Likelihood Ratio Test Statistic in ARMA Models". *Journal of Time Series Analysis*, 25, 83-101.
- Larsson, R. (1998). "Bartlett Corrections for Unit Root Test Statistics". *Journal of Time Series Analysis*, 19, 425-438.
- Lawley, D.N. (1956). "A General Method for Approximating to the Distribution of Likelihood Ratio Criteria". *Biometrika*, 43, 295-303.
- Liu, R. Y. (1988). "Bootstrap Procedures Under Some non-i.i.d. Models", *Annals of Statistics*, 16, 1696-1708.
- Lu, Z. (2016). "Extended MaxT Tests of One-Sided Hypotheses". *Journal of American Statistical Association*, 111, 423-437.
- McCloskey, A. (2017). "Bonferroni-based Size-Correction for Nonstandard Testing Problems". *Journal of Econometrics*, 200, 17-35.
- Masih, A. M. M. and R. Masih (2004). "Fractional Cointegration, Low Frequency Dynamics and Long-Run Purchasing Power Parity: an Analysis of the Australian Dollar Over its Recent Float". *Applied Economics*, 36, 593-605.
- Nielsen, B. (1997). "Bartlett Correction of the Unit Root Test in Autoregressive Models". *Biometrika*, 84, 500-504.
- Obstfeld, M. and K. Rogoff (2001). "The Six Puzzles in International Macroeconomics: Is there a Common Cause?" In *NBER Macroeconomics Annual 2000*, (eds.) Ben S. Bernanke and Kenneth Rogoff, 339-412.
- Omtzigt, P.H. (2003). "Notes on the Bartlett Correction of AR(1) Processes". Manuscript, European University Institute, Florence.
- 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.
- Podivinsky, J. M. (1992). "Small sample properties of tests of linear restrictions on cointegrating vectors and their weights", *Economics Letters*, 39, 13-18.
- Psaradakis, Z. (1994). "A comparison of tests of linear hypotheses in cointegrated vector autoregressive model", *Economics Letters* 45, 137-144.
- Razzak, W., (2018). "The Purchasing Power Parity Puzzle: An Update," *Journal of Economics and Econometrics*, 61, 77-104.
- Ravishanker, N., E. L. Melnick, and C. TSAI (1990). "Differential Geometry of ARMA Models". *Journal of Time Series Analysis*, 11, 259-74.
- Rahbek, A., E. Hansen, and J. Dennis (2002). "ARCH Innovations and Their Impact on Cointegration Rank Testing", Department of Theoretical Statistics and Operation Research, University of Copenhagen, Working Paper 22.

- Rocke, D.M. (1989). "Bootstrap Bartlett Adjustment in Seemingly Unrelated Regression", *Journal of the American Statistical Association*, 84, 598-601.
- Rogoff, K. (1996). "The Purchasing Power Parity Puzzle", *Journal of Economic Literature*, 2, 647-668.
- Rothemberg, T.J. (1984). "Hypothesis Testing in Linear Models When the Error Covariance Matrix is Non-scalar". *Econometrica*, 52, 827-842.
- Sensier, M. and D. van Dijk (2004). "Testing for Volatility Changes in U.S. Macroeconomic Time Series". *The Review of Economics and Statistics*, 86, 833-839.
- Seo, B. (2006). "Asymptotic Distribution of the Cointegrating Vector Estimator in Error Correction Models with Conditional Heteroskedasticity", *Journal of Econometrics*, 137, 68-111.
- Swensen, A.R. (2006). "Bootstrap Algorithms for Testing and Determining the Cointegration Rank in VAR Models", *Econometrica*, 6, 1699-1714.
- Taniguchi, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*. Springer-Verlag.
- Trenkler, C. (2009). "Bootstrapping Systems Cointegration Tests with a Prior Adjustment for Deterministic Terms", *Econometric Theory*, 25, 243-269.
- Tucker, A. L. and L. Pond (1988). "The Probability Distribution of Foreign exchange price Changes: Tests of Candidate Processes", *The Review of Economics and Statistics*, 70, 368-647.
- Van Garderen, K. J. (1999). "Exact Geometry of Autoregressive Models". *Journal of Time Series Analysis*, 20, 1-21.
- Wu, C.F.J. (1986). "Jackknife and Other Resampling Methods in Regression Analysis" *Annals of Statistics*, 14, 1261-1343.