Dipartimento

ESt

Economia e Statistica
Cognetti de Martiis

Working Paper Series

BOOTSTRAP BARTLETT ADJUSTMENT FOR HYPOTHESES TESTING ON COINTEGRATING VECTORS

ALESSANDRA CANEPA

UNIVERSITÀ
DEGLI STUDI
DI TORINO

Bootstrap Bartlett 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 a 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 estimating the p-value of the observed statistic.

JEL Code: C12, C15, C22.

Keywords: Cointegration, Bootstrap, Bartlett Correction.

Word Count: 8168.

Alessandra Canepa,

Department of Economics and Economics and Statistics Cognetti 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 (1998)). In order to address this problem Johansen (2000) proposes a Bartlett adjustment 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 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) 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. One way of overcoming such calculations is to use a numerical approximation in place of the analytical Bartlett correction. 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.

The purpose of this paper is to 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 relay 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). We believe that the dependency on the parameter values may be reduced

by computing the Bartlett adjustment using the non-parametric bootstrap. Important theoretical and empirical results in the literature suggest that the size distortion of the LR test depends on the effects of the nuisance parameters on the small-sample distribution of the test statistic. 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).

This paper examines the performance of the LR test for linear restrictions on the cointegrating space, the Johansen (2000) Bartlett corrected LR test, the bootstrap p-value test and the proposed bootstrap Bartlett corrected LR statistic under non-normal assumption of the innovations. Innovation structures typically found in financial data are considered such as fat tailed and conditionally heteroskedastic (i.e., $ARCH$ and $GARCH$) innovations. Our objective is to decide which inference procedure performs better in finite samples when the Gaussian assumption on the innovations is relaxed. Performance is assessed in terms of the size and power of the inference procedures under consideration. 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. An empirical application is considered in Section 5 and Section 6 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

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

Thus, the quantity

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

has an expectation $q + O\left(T^{-3/2}\right)$ which is closer to the limit distribution. For a survey on this type of correction see for example Cribari-Neto and Cordeiro (1996).

2   Model and Tests

Consider the p-dimensional VAR model

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

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 $\Gamma_1, ..., \Gamma_{k-1}$ 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

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

where $\widehat{\lambda}_i$ 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 derived in Johansen (2000) and is given by

$$\vartheta \;=\; \frac{E_\theta\left(\Lambda\right)}{q} = 1 + \frac{1}{T}\left[\frac{1}{2}\left(p + s - r + 1 + 2p_D\right) + p_d + kp\right] \tag{4}$$

$$+ \frac{1}{Tr}\left[\left(2p + s - 3r - 1 + 2p_D\right)v\left(\alpha\right) + 2\left(c\left(\alpha\right) + c_d\left(\alpha\right)\right)\right]$$

where $q = r(p - s)$, $v\left(\alpha\right) = tr\left\{\left(\alpha'\Omega^{-1}\alpha\right)^{-1}\sum_{\beta\beta}^{-1}\right\}$ with $\sum_{\beta\beta} = Var(\beta'Y|\Delta Y_t, ..., \Delta Y_{t-k+2})$, $c_d = p_d v\left(\alpha\right)$, and the constant $c\left(\alpha\right)$ is given in Johansen. 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 $\varepsilon_t \backsim N(0, \Omega)$. 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).

In principle, one may expect the two bootstrap procedures to perform in a similar fashion given that in both cases we replace $F\left(\Lambda\right)$, the unknown cumulative distribution function of Λ, with $\hat{F}^*\left(\Lambda\right)$, the empirical distribution function of the bootstrap distribution of Λ_j^*. Let B be the number of bootstrap replications, as $B \to \infty$ then $\hat{F}^*\left(\Lambda\right) \to F^*\left(\Lambda\right)$, the true cumulative distribution function of Λ_j^*. Thus, the performance of the bootstrap procedures depend on how well

$F^*(\Lambda)$ approximates $F(\Lambda)$. When the bootstrap data generating process (DGP) fails to mimic the features of the true DGP it would be unwise to expect one bootstrap test to outperform the other. However, one may expect the bootstrap Bartlett corrected test to be less computationally intensive than the bootstrap p-value test.

Monte Carlo experiments involving bootstrapping are computationally demanding as the number of loops involved is $N \times B$, where N is the number of Monte Carlo replications. Generally speaking, the number of bootstrap replications needed in a Monte Carlo experiment is directly related to the nature of the results desired. For instance, the more work that has to be done in the "thin" part of the distribution (e.g. the tails), the higher B is needed. This is because values of the statistics will occur in these areas much less frequently than they will in the "thicker" sections of the distribution. Thus, many more bootstrap trials are needed to approximate these sections. Because the bootstrap Bartlett corrected procedure uses bootstrapping to approximate the central moment of a distribution it may require a lower number of replications than the bootstrap p-value test, in which bootstrapping is used to approximate the tail of the distribution.

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

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.

The steps used to implement the bootstrap algorithm for calculating Λ_B^* can be summarized as follows:

Step (1): Estimate the model in (2) for a given cointegrating rank and obtain the unrestricted estimate $\hat{\Theta}$ and the residual $(\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T)$, the restricted estimates $\tilde{\Theta}$ and the residuals $(\tilde{\varepsilon}_1, ..., \tilde{\varepsilon}_T)$. Then, calculate Λ.

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

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

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

Step (3): Estimate VAR using Y_t^* the pseudodata; obtain the the unrestricted and restricted estimates of the coefficients and residuals, $\hat{\Theta}^*$ and $(\hat{\varepsilon}_1^*, ..., \hat{\varepsilon}_T^*)$, $\tilde{\Theta}^*$ and $(\tilde{\varepsilon}_1^*, ..., \tilde{\varepsilon}_T^*)$. Then, test for the hypothesis $\mathcal{H}_0 : \beta = \hat{\beta}^*$.

Step (4): Compute Λ_j^* using the data of step (2) and repeat B times. To get an estimate of the average value of Λ, say $\overline{\Lambda}^*$, average the observed values $\Lambda_1^*, ..., \Lambda_B^*$, . A Bartlett-type corrected statistic is therefore

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

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

As far as the bootstrap p-value test is concerned the bootstrap algorithm adopted is similar to the procedure proposed by Gredenhoff and Jacobson (2001) (see also Trenkler (2009)). This involves repeating step (1)-(3) and then following stwp (5) below.

where $I(\cdot)$ is the 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 $\hat{P}^*(\Lambda) \leq \gamma$.

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

$$\Delta Y_t^* = \hat{\alpha}\hat{\beta}' Y_{t-1}^* + \sum_{i=1}^{k-1} \hat{\Gamma}_i \Delta Y_{t-i}^* + \hat{\phi} d_t + v_t^*$$

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

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

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

Using the fact that $\theta = f\left(\alpha, \beta, \Gamma_i, \Omega\right)$ 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: There has been some discussion in the literature on the use of restricted rather than unrestricted residuals when implementing the bootstrap procedure. Using the restricted residuals implies estimating the constrained VAR model and generating the pseudo-data on the basis of the estimated constrained coefficients. Omtzigt and Fachin (2006) argue that if the null hypothesis is not true resampling from the unrestricted residuals greatly improve the power of the p-value bootstrap test and should be preferred. The authors also suggest that the Bartlett corrected test should be based on the unrestricted estimates. In this paper we have followed their recommendation.

REMARK 2.2: 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 schemes 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 because 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. In this case the block bootstrap or subsampling methods may be used. Investigating the usefulness of the suggested bootstrap Bartlett test in the case of correlated innovations will be the subject of future research.

2.1.3 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

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

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

$$Z_t = \frac{-\left(\sqrt{5}-1\right)}{2} \qquad \text{with probability } \frac{\left(\sqrt{5}+1\right)}{2\sqrt{5}}$$

$$= \frac{\left(\sqrt{5}+1\right)}{2} \qquad \text{with probability } \frac{\left(\sqrt{5}-1\right)}{2\sqrt{5}}$$

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

Using the fact that $\theta = f\left(\alpha, \beta, \Gamma_i, \Omega\right)$ 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 schemes 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. In this case the block bootstrap or subsampling methods may be used. Investigating the usefulness of the suggested bootstrap Bartlett test in the case of correlated innovations will be the subject of future research.

2.2   Some Asymptotic Results

We now consider the statistics Λ^* and Λ_B^* and we show that the distributions of the bootstrap tests coincide with the corresponding asymptotic counterparts. We focus on the pseudo-data generated by Algorithm 2 since the consistency of the bootstrap procedure proposed in Algorithm 1 can be

derived in a similar way. Our approach builds on the important theoretical results in Swensen (2006) and Cavaliere, Rahbek and Taylor (2010a).

In the following $\overset{w}{\to}$ denotes weak convergence, $\overset{P}{\to}$ convergence in probability, $\overset{w_p}{\to}$ 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, $\|a\| = (a'a)^{1/2}$.

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,

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

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 $\Gamma = I_p - \Gamma_1 - ... - \Gamma_{k-1}$.

(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

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

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

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

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

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)). Recently, 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 reminder term depends on the realization and needs careful consideration in the bootstrap context. Lemma 1 extends the validity of Lemma 1 in Swensen (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, $\Lambda^* \overset{w_p}{\to} \Lambda$ as $T \to \infty$.*

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

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

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

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 $\eta > 0$, $P^* (\max_{t=1,...,T} \| R_t^* \| > \eta) \to 0$ in probability as $T \to \infty$.

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 $\tilde{Z}_{2t} = \left[\Delta Y'_{t-1}, \Delta Y'_{t-2}, ..., \Delta Y'_{t-k+1} \right]$. Moreover,

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

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

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

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

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

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

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

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

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

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

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

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

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

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

such that the continuous mapping theorem gives

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

Similarly, we have

$$\hat{\beta}'_\perp \left(S^*_{10} - S^*_{11}\hat{\beta}\hat{\alpha}' \right) \hat{\alpha}_\perp \overset{w_p}{\to} \int_0^1 F(u)d)B'\alpha_\perp.$$

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 $z : \mathbb{R}^p \longmapsto \mathbb{R}^s$ with matrix representation $Y^*_t \longmapsto H'Y^*_t$ forms a subspace such that $sp\left(\hat{\beta}\right) \subset sp\left(H\right)$. 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 $\left\{T^{-1/2}\left(\sum_{t=1}^{[Tu]} H'v^*_i\right)\right\}$ converges weakly toward a Brownian motion with covariance matrix $H'\hat{\Omega}H$ and the asymptotic distribution of the moment matrices is given by

$$T^{-1/2}\hat{\varphi}'_\perp H'Y^*_{[Tu]} \overset{w_p}{\to} H'\tilde{F}(u) \tag{14}$$

$$\hat{\varphi}'_\perp H'S^*_{10}\hat{\alpha}_\perp \overset{w_p}{\to} H'\int_0^1 \tilde{F}dB'\alpha_\perp \tag{15}$$

$$T^{-1}\hat{\varphi}'_\perp H'S^*_{11}H\hat{\varphi}_\perp \overset{w_p}{\to} H'\int_0^1 \tilde{F}(u)\tilde{F}(u)'duH. \tag{16}$$

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

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

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 $S^*_{ij} \to \Sigma_{ij}$ in probability and the estimators of the parameters are consistent. Under the conditions of Theorem 1, this trivially implies that $E^*\left(\Lambda^*\right) \to E\left(\Lambda\right)$ in probability as $T \to \infty$.

∎

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 here is given by

$$
\begin{aligned}
Y_{1t} &= Y_{2t} + u_{1t} & \text{where} \quad u_{1t} &= \xi u_{1t-1} + \epsilon_{1t} & (5) \\
Y_{2t} &= -Y_{1t} + u_{2t} & u_{2t} &= u_{2t-1} + \epsilon_{2t} \\
\Delta Y_{3t} &= \epsilon_{3t} \\
\Delta Y_{4t} &= \epsilon_{4t}
\end{aligned}
$$

with

$$
\begin{bmatrix} \epsilon_{it} \\ \epsilon_{jt} \end{bmatrix} \sim i.i.d.\ N \left[\begin{pmatrix} \underset{(2\times1)}{0} \\ \underset{(2\times1)}{0} \end{pmatrix}, \begin{pmatrix} \underset{(2\times2)}{A} & \underset{(2\times2)}{0} \\ \underset{(2\times2)}{0} & \underset{(2\times2)}{B} \end{pmatrix} \right]
$$

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

$$
\mathcal{H}_0 : \beta = H\varphi = \begin{bmatrix} 1 & \underset{(1\times2)}{0} \\ -1 & \underset{(1\times2)}{0} \\ \underset{(2\times1)}{0} & \underset{(2\times2)}{I} \end{bmatrix} \underset{(3\times1)}{\varphi},
$$

is tested against the alternative $\mathcal{H}_1 : \beta$ unrestricted. For easy of interpretation the DGP in (5) 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 (5) 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 (5) 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 (5). 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 (5) maintains high control over the experimental design while also having greater practical relevance. The experimental parameter space is $T \in (50, 100, 250)$, $\xi \in (0.2, 0.5, 0.8, 1)$, $\eta \in (-0.5, 0, 0.5)$, $\sigma = 1$. 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. The DGP in (5) has

$$\varepsilon_{it} \sim i.i.d. \quad \omega_1 N\left(\mu_1, \sigma_{M1}^2\right) + \omega_2 N\left(\mu_2, \sigma_{M2}^2\right) \quad \text{and } \varepsilon_{jt} \sim N(0,1) \tag{6}$$

with $\omega_1\mu_1 + \omega_2\mu_2 = 0$, ω_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 ε_{it} is preserved, being the means of the normal distributions mixed at zero. In this case, excess kurtosis has been introduced in (5) by choosing $\omega_1 < \omega_2$ and $\sigma_{M1}^2 > \sigma_{M2}^2$. Under this assumption the kurtosis in ε_{it} is strictly increasing according to the quantity $\frac{\sigma_{M1}^2}{\sigma_{M2}^2}$. Consistent with these considerations, the following five distributions of ε_{it} have been investigated

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

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

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

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

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 bero zero and $Kurt$ should be equal to three.[1]

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

	D_1	D_2	D_3	D_4
$Mean$	.0	.0	.0	.0
Var	1.08	3.28	3.82	6.29
$Skew$	.0	-0.50	-2.65	1.34
$Kurt$	4.610	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

---

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

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

with $\varsigma_{it} \sim N\left(0,1\right)$ (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

$$h_t = \frac{\sigma}{1-\varrho} + \varrho\varepsilon_{1t-1}^2, \tag{7}$$

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

$$h_t = \psi_0 + \psi_1\varepsilon_{1t-1}^2 + \psi_2 h_{t-1}. \tag{8}$$

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

$$D_5 \quad : \quad \varepsilon_{it} \text{ as in (7) with } \varrho = 0.4;$$

$$D_6 \quad : \quad \varepsilon_{it} \text{ as in (7) with } \varrho = 0.8;$$

$$D_7 \quad : \quad \varepsilon_{it} \text{ as in (8) with } \psi_1 = 0.570 \text{ and } \psi_2 = 0.921;$$

$$D_8 \quad : \quad \varepsilon_{it} \text{ as in (8) with } \psi_1 = 0.095 \text{ and } \psi_2 = 0.881.$$

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. The Monte Carlo experiment was based on $N = 10,000$ replications 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

Table 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. For the case with $\chi^2(c)$ and $t(c)$ distributions, the Monte Carlo results are summarized in Table 4 using response surface regressions. Finally, Table 5 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 $\varepsilon_{it} \sim N(0,1)$. 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 $\eta \to 0$. 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.

	Test	ξ = 0.8			η = 0.5		
		$\eta = -0.5$	$\eta = 0$	$\eta = 0.5$	$\xi = 0.2$	$\xi = 0.5$	$\xi = 1$
$T = 50$	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

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 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 ε_{it}. Looking at the estimated levels of Λ over the range $D_1, ..., D_4$ 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 $\hat{\beta}$ 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.

| | | $\xi = 0.8$ | | | | | | | | |
| | | $\eta = -0.5$ | | | | | $\eta = 0.5$ | | | |
	ε_{it}	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

Table 3. (Continue)

	ε_{it}		$\eta = 0.5$							
			$\xi = 0.2$					$\xi = 0.5$		
		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 (1.079)	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

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 ε_{it}. 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.

	ε_{it}	$\xi = 0.5$				$\xi = 0.8$			
		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

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

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

where $g \in (1.2, 1.4, 1.6, 1.8.2)$ with $\xi = 0.5, \eta = 0.5$, and $\sigma = 1$ in (5). The results of this set of experiments are reported in Table 5. Once again, the case of ϵ_{it} i.i.d. $N \sim (0, 1)$ serves as a benchmark, then $\epsilon_{it} \sim D_1$ and $\epsilon_{it} \sim D_5$ 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.

In Table 5 power estimates 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. 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 deteriorate 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 sample (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.

24

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
$T = 250$		Λ	99.9	100	100	100	100
	N	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 $\eta = 0.5$, $\xi = 0.5$.

Before concluding this section, an important point regarding the bootstrap algorithm used to calculate the approximation of the finite sample expectation of the LR test is discussed. In algorithm 1 and 2 the residuals were resampled with replacement from the restricted model. In alternative the resampling could be implemented using the unrestricted residuals. This implies estimating the unconstrained VAR model and generating the pseudo-data on the basis of the estimated unconstrained coefficients. Omtzigt and Fachin (2006) argue that if the null hypothesis is not true resampling from the unrestricted residuals greatly improve the power of the test statistic and should be preferred. The authors also suggest that the p-value bootstrap test should be based on the unrestricted estimates. However, Omtzigt and Fachin (2006) consider the analytical Bartlett correction factor proposed by Johansen (2000) and it is not clear if their recommendations should be followed in this context.

In order to check if the power properties of the bootstrap Bartlett corrected test was affected by the choice of the resampling procedure, a simulation experiment was undertaken comparing the rejection frequencies for the size and power of the test generated using the restricted and the unrestricted residuals. The results of this simulation experiment are reported in Table 6. In Table 6 the test statistic generated using the unrestricted residuals is labelled as $\Lambda^*_{B_u}$. Once again, the estimated rejection probabilities of Λ^*_B and $\Lambda^*_{B_u}$ were calculated using the DGP in (5) with $\eta = 0.5$ and $\xi = 0.5$. The rejection frequencies for $\Lambda^*_{B_u}$ were obtained using the algorithm below.

4.1.1 Algorithm 3

Step (i): Estimate the model in (2) under the null hypothesis and calculate the unrestricted residuals

$$\hat\varepsilon_t = \Delta Y_t - \hat\alpha\beta\left(Y_{t-1} + \hat\rho' D_t\right) - \sum_{i=1}^{k-1}\hat\Gamma_i\Delta Y_{t-i} - \hat\phi d_t.$$

Step (ii): Compute Y_t^* recursively from

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

with sampled residuals $(\varepsilon_1^*, ..., \varepsilon_T^*)$ drown with replacement from $(\hat\varepsilon_1, ..., \hat\varepsilon_T)$.

Step (iii) Using the bootstrap sample compute the unrestricted and restricted estimates of the coefficients and calculate Λ_j^*.

Step (iv) Follow Step (4) to compute $\overline\Lambda_u^*$ and refer $\Lambda^*_{B_u} = \frac{q\Lambda}{\overline\Lambda_u^*}$ to a $\chi^2(q)$ distribution (with $q = r(p - s)$).

In the case were innovations are heteroskedastic the rejection probabilities were calculated by simply replacing the wild bootstrap DGP in step 1 and then following the other three steps as for algorithm 2.

For ease of interpretation the rejection frequencies for the power of Λ_B^* in Table 5 are also reported in Table 6.

Table 6. Rejection frequencies for Λ_B^* and $\Lambda_{B_u}^*$ for the 5% critical value.

			Size	Power				
		ε_t		$g = 1.2$	$g = 1.4$	$g = 1.6$	$g = 1.8$	$g = 2$
$T = 50$	Λ_B^*	N	5.3	60.4	83.6	88.7	90.5	90.8
	$\Lambda_{B_u}^*$		5.7	64.9	87.0	90.8	96.6	98.9
	Λ_B^*	D_1	6.3	29.3	53.3	59.7	63.9	65.0
	$\Lambda_{B_u}^*$		8.3	38.4	60.6	65.7	72.4	73.1
	Λ_B^*	D_2	6.2	28.2	44.0	52.6	57.4	60.5
	$\Lambda_{B_u}^*$		10.1	32.8	54.1	65.0	62.0	66.0
	Λ_B^*	D_5	5.3	11.4	21.5	29.1	33.2	36.6
	$\Lambda_{B_u}^*$		6.1	12.7	26.8	34.7	39.0	42.3
	Λ_B^*	D_8	5.5	9.4	18.6	31.0	37.1	39.9
	$\Lambda_{B_u}^*$		8.8	12.4	25.1	35.1	42.3	46.9
$T = 100$	Λ_B^*	N	5.1	96.8	99.0	100	100	100
	$\Lambda_{B_u}^*$		5.2	99.3	99.5	99.8	100	100
	Λ_B^*	D_1	5.9	80.1	96.7	99.0	99.3	99.6
	$\Lambda_{B_u}^*$		6.1	86.8	98.3	99.3	99.8	100
	Λ_B^*	D_2	5.0	58.9	87.6	95.1	96.8	98.0
	$\Lambda_{B_u}^*$		7.3	62.8	87.7	96.5	98.9	99.9
	Λ_B^*	D_5	5.2	32.7	66.7	80.9	87.2	89.5
	$\Lambda_{B_u}^*$		4.9	35.2	70.7	82.3	92.3	97.1
	Λ_B^*	D_8	5.2	22.1	56.8	75.5	83.3	86.3
	$\Lambda_{B_u}^*$		5.6	26.4	60.0	76.6	83.9	87.9
$T = 250$	Λ_B^*	N	5.0	99.8	100	100	100	100
	$\Lambda_{B_u}^*$		4.0	100	100	100	100	100
	Λ_B^*	D_1	5.0	91.9	99.7	100	100	100
	$\Lambda_{B_u}^*$		5.5	99.2	100	100	100	100
	Λ_B^*	D_2	5.1	99.4	100	100	100	100
	$\Lambda_{B_u}^*$		5.2	96.4	100	100	100	100
	Λ_B^*	D_5	4.4	91.9	99.7	100	100	100
	$\Lambda_{B_u}^*$		4.8	98.3	99.9	100	100	100
	Λ_B^*	D_8	5.0	75.5	98.8	99.9	100	100
	$\Lambda_{B_u}^*$		5.5	78.4	98.0	99.7	99.9	100

Note: DGP with $\eta = 0.5$, $\xi = 0.5$. $N = 1000$, $B = 800$.

Simulations results in Table 6 show that there is some loss in power when the bootstrap algorithm is based on the restricted DGP. However, type I error control is superior when the bootstrap DGP is based on the restricted estimates. This suggests that, for practical implementations, the choice between the restricted and the unrestricted residuals has to be based on the specific application.

5   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

$$PPP_t = P_t - \ddot{P}_t - E_t. \tag{9}$$

Equation (9) 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 Rogoff (1996) defines as the "PPP puzzle".

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. 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. 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 PPP, relationship using the data in Johansen and Juselius (1992) is considered. The data set contains quarterly data from 1971:1 to 1987:1 for the nominal dollar exchange rate for the UK (E_t), the domestic consumer price index (P_t), and the US consumer price index $\left(\ddot{P}_t\right)$. The first two columns in the top of Table 6 summarize the misspecification tests for the unrestricted VAR

$$Y_t = \left[\ \left(P_t - \ddot{P}_t\right) \quad E_t \ \right]'$$

estimated with an unrestricted constant, two lags and $T = 64$. The diagnostic tests involve F_{ar} for the hypothesis that there is no serial correlation against the fourth-order autoregression, χ^2_{no} that residuals are normally distributed, F_{arch} that there is no autoregressive conditional heteroskedasticity (against fourth order) and χ^2_{het} for the hypothesis that there is no heteroskedasticity. Looking at the misspecification tests it emerges that F_{ar} does not reject the null hypothesis of no autocorrelation against fourth order autoregression for both equations determining $\left(P_t - \ddot{P}_t\right)$ and E_t. There is, however, evidence of non normality and heteroskedasticity of the ARCH type for $\left(P_t - \ddot{P}_t\right)$ given that the χ^2_{no} and F_{arch}-tests reject the null hypotheses for this equation. Johansen and Juselius argue that by looking at the residual plots, excess of kurtosis is found to coincide with significant changes in oil prices, and thus they condition their model on Δpo_t and Δpo_{t-1}, where po_t measures the world price of crude oil. Accordingly, in order to investigate the effects of removing "problematic" data features on the empirical sizes of the test statistics an alternative model that includes weakly exogenous variable is estimated. Note that Δpo_t and Δpo_{t-1} enter d_t only in (2). The diagnostic tests for the conditional model are reported in third and fourth columns of the top panel of Table 7. From Table 7 it is clear that there is a significant reduction in the kurtosis associated with $\left(P_t - \ddot{P}_t\right)$, moreover heteroskedasticity and non-normality are no longer a problem.

On the basis of the rank tests reported in the middle panel, it is possible to accept the hypothesis that there is one cointegration vector, since the trace statistics associated with the null hypothesis that $r = 0$ are rejected at the 10% and 5% levels for the unconditional and conditional models, respectively. Note that the trace test is likely to suffer from the same size distortion problem as Λ, but looking at the property of the VAR(2) companion matrix of the unconditional model the moduli of the largest unit root is 0.9157, suggesting that $r = 1$ is correct.

In the bottom panel of Table 7, the empirical sizes and their associated p-values for $\Lambda, \Lambda_B, \Lambda_B^*$ and Λ^* are reported. Empirical levels for Λ_B^* and Λ^* in the second column were obtained using algorithm 2 in Section 2, whereas the levels for Λ_B^* and Λ^* in the forth column were obtained using algorithm 1 in the same section. The null hypothesis under consideration is that $\left(P_t - \ddot{P}_t\right) - E_t$ 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

$$H' = \begin{bmatrix} 1 & -1 \end{bmatrix}$$

versus the alternative $\mathcal{H}_1 : \beta$ unrestricted. Based on the evidence from the misspecification tests the empirical sizes for Λ_B^* and Λ^* are calculated using algorithm 2 in Section 2 for the unconditional VAR(2) model and algorithm 1 for the conditional model with $B = 5,000$. The p-values for Λ are calculated by taking the 95% percentile from the $\chi^2(1)$ and calculating the actual p-value as the frequency of rejection. The test statistic is corrected with the Bartlett correction factor using (4), and the rejection frequency is then recalculated, thus providing corrected p-values for Λ_B.

Table 7. Model evaluation diagnostics and test statistics and actual rejection probabilities for the PPP relationship using the UK data.

Misspecification Tests $(k = 2, T = 64)$:

Unconditional Model		*Conditional Model*	
$\left(P_t - \ddot{P}_t\right)$	E_t	$\left(P_t - \ddot{P}_t\right)$	E_t
$F_{ar}(4,50) = 0.761$	$F_{ar}(4,50) = 0.091$	$F_{ar}(4,48) = 1.541$	$F_{ar}(4,48) = 0.807$
$\chi^2_{no}(2) = 10.19^*$	$\chi^2_{no}(2) = 0.886$	$\chi^2_{no}(2) = 3.120$	$\chi^2_{no}(2) = 0.713$
$F_{arch}(4,46) = 6.689^*$	$F_{arch}(4,46) = 0.549$	$F_{arch}(4,44) = 0.745$	$F_{arch}(4,44) = 0.189$
$\chi^2_{het}(14,39) = 1.559$	$\chi^2_{het}(14,39) = 0.429$	$\chi^2_{het}(14,37) = 1.141$	$\chi^2_{het}(14,37) = 0.460$

$I(1)$ *Analysis*:

	Unconditional	*Model*		*Conditional*	*Model*
Rank	$\hat{\lambda}$	*Trace Test*		$\hat{\lambda}$	*Trace Test*
0	0.150	14.420***		0.159	20.168**
1	0.074	4.661		0.149	9.743

Tests for Linear Restrictions (5% *Nominal Level*)

	Unconditional	*Model*		*Conditional*	*Model*
	$\chi^2(1)$	p-val.		$\chi^2(1)$	p-val.
Λ	6.213	0.013		2.161	0.141
Λ_B	4.891	0.031		1.639	0.200
Λ_B^*	3.675	0.055		1.490	0.216
Λ^*	−	0.057		−	0.219

Note: Diagnostic tests involve F_{ar} for the hypothesis that there is no serial correlation against the fourth-order autoregression, that residual are normally distributed $\left(\chi^2_{no}\right)$, that there is no heteroskedasticity $\left(\chi^2_{het}\right)$, and that

there is no autoregressive conditional heteroskedasticity (F_{arch}). *), **), ***) reject the null at 1%, 5% and 10%, respectively. The p-values of a nominal 5% test are obtained by the frequency of rejections. Note that Λ^* does not yield an adjusted LR test.

As far as test results are concerned, from Table 7 it appears that the i.i.d. assumption on the innovations clearly affects the test results. When considering the conditional model PPP relationship cannot be rejected according to test statistics under consideration, thus confirming the results of Johansen and Juselius. However, inference based on the analytical Bartlett correction is affected by heteroskedastic innovations since Λ_B rejects $\mathcal{H}_0$ for the unconditional model, but the opposite is true for the conditional model. On the other side, inferences based on the bootstrap method appear to be robust to deviations from the i.i.d. assumption since for the unconditional model the p-values reported in Table 7 are 0.055 and 0.057 for Λ_B^* and Λ^*, respectively, thus the PPP hypothesis cannot be rejected. It is also noteworthy that Λ_B, Λ_B^* and Λ^* perform equally well when assumptions on the innovations are not violated.

6 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

Bartlett, M.S. (1937). "Properties of Sufficiency and Statistical tests". *Proceedings of the Royal Society*, 160, 263-268.

Bewley, R. and D. Orden (1994). "Alternative Methods of Estimating Long-Run Responses with Application to Australian Import Demand". *Econometric Review*, 13, 179-204.

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.

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.

Cavaliere, G., A. Rahbek, and A.M.R. Taylor, (2010a). "Cointegration Rank testing Under Conditional Heteroskedasticity", *Econometric Theory*, forthcoming.

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.

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.

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.

Hansen, B. E. (1992). "Convergence of Stochastic Integrals for Dependent Heterogenous Processes". *Econometric Theory*, 8, 489-500.

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.

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.

Jacobson, T. and T. Larsson, (1999). "Bartlett Correction in Cointegration Testing", *Computational Statistics and Data Analysis*, 31, 203-225.

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.

Liu, R. Y. (1988). "Bootstrap Procedures Under Some non-i.i.d. Models", *Annals of Statistics*, 16, 1696-1708.

Omtzigt, P., Fachin, S. (2006). "The Size and Power of Bootstrap and Bartlett-Corrected Tests of Hypothesis on the Cointegrating Vectors", *Econometric Review*, 25, 41-60,

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.

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.

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.

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.

Wu, C.F.J. (1986). "Jackknife and Other Resampling Methods in Regression Analysis" *Annals of Statistics*, 14, 1261-1343.