Bootstrap Co-integration Rank Testing: The Effect of Bias-Correcting Parameter Estimates[*]

Giuseppe Cavaliere[a], A.M.Robert Taylor[b], and Carsten Trenkler[c]

[a] Department of Statistical Sciences, University of Bologna

[b] School of Economics, University of Nottingham

[c] Department of Economics, University of Mannheim

March 7, 2013

Abstract

In this paper we investigate bootstrap-based methods for bias-correcting the first-stage parameter estimates used in some recently developed bootstrap implementations of the co-integration rank tests of Johansen (1996). In order to do so we adapt the framework of Kilian (1998) which estimates the bias in the original parameter estimates using the average bias in the corresponding parameter estimates taken across a large number of auxiliary bootstrap replications. A number of possible implementations of this procedure are discussed and concrete recommendations made on the basis of finite sample performance evaluated by Monte Carlo simulation methods. Our results show that bootstrap-based bias-correction methods can significantly improve upon the small sample performance of the bootstrap co-integration rank tests. A brief application of the techniques developed in this paper to international dynamic consumption risk sharing within Europe is also considered.

Keywords: Co-integration; trace test; bias-correction; bootstrap.

J.E.L. Classifications: C30, C32.

1 Introduction

The use of tests based on the likelihood ratio [LR] principle for determining the co-integration rank of a VAR system of $I(1)$ variables, see Johansen (1996), is now common-place in empirical research in macroeconomics and finance. However, the finite sample properties of these tests, when based on asymptotic critical values, can be quite poor; see, in particular, Johansen (2002) and the references therein. This has prompted a number of recent studies to propose bootstrap implementations of the LR co-integration rank tests with the aim of delivering tests with empirical rejection frequencies closer to the nominal level; see, in particular, Swensen (2006), Cavaliere, Rahbek & Taylor (2010a) and Cavaliere, Rahbek & Taylor (2012).

Monte Carlo results reported in Swensen (2006) and Cavaliere et al. (2012) suggest that the aforementioned bootstrap LR tests do indeed appear to yield significant improvements on the small sample performance of the asymptotic LR tests, particularly so the smaller the sample size and the larger the dimension of the VAR system under study. However, their results also show that significant size distortions remain in the bootstrap tests when the VAR process contains stationary dynamics, most notably where these display strong positive autocorrelation, such that we have a near-$I(2)$ system. These bootstrap procedures are both based around the use of bootstrap sample data formed using estimates of the stationary dynamics obtained from the original data. As a consequence, their efficacy will clearly be related to the degree of finite sample bias present in these estimates. Swensen (2006) and Cavaliere et al. (2010a) use unrestricted estimates of the stationary dynamics while Cavaliere et al. (2012) estimate the dynamics under the co-integration rank restriction of the null hypothesis being tested. Cavaliere et al. (2012) demonstrate the theoretical validity of the latter approach and that it delivers superior finite sample properties to the approach outlined in Swensen (2006) and Cavaliere et al. (2010a), the theoretical validity of which also remains to be established. We therefore focus attention on the approach of Cavaliere et al. (2012) in this paper.

In the context of providing estimated confidence intervals for impulse response functions from VAR models, be they estimated in levels, first differences or co-integrated VAR form, Kilian (1998) shows that these can be quite inaccurate in small samples owing to the finite sample bias seen in the estimates of the lag coefficient matrices characterising the VAR model; see also the simulation results in Engsted & Pedersen (2011) Engsted and Pedersen (2011). As Kilian (1998) further notes, this bias is systematic and, as a consequence, bootstrap data generated conditional on biased point estimates will tend to result in an even greater bias in the resulting bootstrap estimates of the lag coeffi-

cient matrices, relative to the true parameters. This leads to the potential for a standard bootstrap confidence interval to be less accurate than that based on the original estimates. In order to improve upon the accuracy of the bootstrap confidence levels, Kilian (1998) proposes a so-called bootstrap-after-bootstrap [BaB] method. Here the bootstrap data are generated not using the original point estimates from the VAR model but on bias-corrected estimates which are themselves obtained by bootstrap methods.

The basic idea underlying the bias correction used in the BaB approach of Kilian (1998) is as follows. Suppose one estimates the matrix of parameters on the lagged dependent variable in a stationary VAR model with one lag [VAR(1)] and is interested in obtaining an estimate of the finite sample bias inherent in this estimate. The BaB approach then uses some form of re-sampling, for example i.i.d. re-sampling, from the residuals from this estimated VAR(1) model to obtain a set of bootstrap innovations. A bootstrap analogue of the original sample data is then constructed using a recursion derived from these bootstrap innovations and the estimated lag parameter matrix. A VAR(1) is then estimated on the resulting bootstrap data using the same estimation method as was applied to the original data. This bootstrap procedure is replicated a large number, say B_1 times. Since we have knowledge of the true parameter matrix in these bootstrap samples, one can then obtain an estimate of the bias present in the original parameter estimates by using the average bias taken across the B_1 parameter estimates from the B_1 sets of bootstrap data. This quantity can then be used to bias-correct the original estimate.

In this paper we adapt the use of the BaB approach for use with the bootstrap co-integration rank tests of Cavaliere et al. (2012). As noted above, the finite sample performance of the bootstrap rank tests can vary considerably according to the pattern of the stationary dynamics in the system, and so we anticipate that the resulting bootstrap tests based around the use of bootstrap bias-corrected estimates of the stationary dynamics will further improve upon the finite sample properties of the bootstrap tests. An interesting issue which arises in doing so is that the bootstrap bias-corrected parameter estimates of the stationary dynamics are typically larger than the original estimates. As a consequence, this effects an increase in the number of failures of the root stability checks which are a feature of the bootstrap algorithm of Cavaliere et al. (2012) (and, indeed, of the corresponding algorithm in Swensen, 2006), especially where the stationary dynamics have roots which lie close to the unit circle. We explore the impact of: (i) simply ignoring the root check condition of the bootstrap algorithm, (ii) using a switching bootstrap whereby we revert to the original estimates if the bootstrap bias-corrected estimates fail the root condition; (iii) a root-correction bootstrap, following a suggestion in Kilian (1998). We also explore the possibility, where the co-integration rank being tested is greater than zero,

of applying a bootstrap bias-correction to the estimates of both the stationary dynamics and the long run component of the system.

Monte Carlo simulation methods are used to explore the finite sample properties of the BaB co-integration rank tests and to compare these with the corresponding standard bootstrap rank tests of Cavaliere et al. (2012). Our results show that the BaB method delivers significant improvements upon the finite sample performance of the standard bootstrap tests, particulary in very small samples and where the system lies close to the $I(2)$ boundary in the parameter space. The results also suggest that ignoring the root check conditions discussed above is a safe strategy in practice and yields the best results overall, relative to a switching bootstrap or root-correction bootstrap. Applying the BaB method to the long run component of the system does not appear to deliver any noticeable gains in accuracy once the sort run dynamics of the system have been bias-corrected.

The remainder of the paper is organised as follows. In section 2 we outline our reference co-integrated VAR model and outlines the co-integration rank tests of Johansen (1996), outlining the large sample properties of these, with the standard bootstrap implementations of these tests proposed by Cavaliere et al. (2012) outlined in section 3. In section 4 we introduce our proposed bias-corrected, or bootstrap-after-bootstrap, bootstrap rank tests, together with the variants of this procedure discussed above. The results of our Monte Carlo study comparing the finite sample behaviour of the standard and bias-corrected bootstrap co-integration tests are reported and discussed in section 5. In section 6 we use the bootstrap techniques discussed in this paper to re-assess the empirical evidence provided by Cavaliere, Fanelli & Gardini (2008) on international dynamic consumption risk sharing within a set of European countries. Section 7 concludes.

2   The Model Framework and Rank Tests

We consider the usual VAR(k) model in error correction format:

$$\Delta X_t = \Pi X_{t-1} + \Psi U_t + \mu D_t + \varepsilon_t, \quad t = 1, 2, \ldots, T, \tag{2.1}$$

where $X_t := (X_{1t}, \ldots, X_{pt})'$ and the innovations, $\varepsilon := (\varepsilon_{1t}, \ldots, \varepsilon_{pt})'$, are both $p \times 1$, and $U_t := (\Delta X'_{t-1}, \ldots, \Delta X'_{t-k+1})'$ is $p(k-1) \times 1, \Psi := (\Gamma_1, \ldots, \Gamma_{k-1})$, where $\{\Gamma_i\}_{i=1}^{k-1}$ are $p \times p$ lag coefficient matrices. The impact matrix $\Pi := \alpha\beta'$, where α and β are $p \times r$, $r \leq p$, matrices, with the usual convention that $\alpha\beta'$ is the $p \times p$ matrix of zeroes when $r = 0$ (no co-integration). The initial values, $\mathbb{X}_0 := (X'_0, \ldots, X'_{-k+1})'$, are taken to be fixed. We allow for a restricted linear trend such that $\mu D_t = \mu_1 + \mu_2 t$ with $\mu_2 = \alpha\rho'$ in (2.1). Then, the model, which will often be referred to as model $H(r)$ in what follows, may be written

in the compact form,

$$Z_{0t} = \alpha\beta^{+\prime}Z_{1t} + \Psi^+ Z_{2t} + \varepsilon_t \tag{2.2}$$

with $Z_{0t} := \Delta X_t$, $Z_{1t} := (X'_{t-1}, t)'$, $Z_{2t} := (U'_t, 1)'$, $\beta^+ = (\beta', \rho')'$, and $\Psi^+ = (\Psi, \mu_1)$.

Throughout the paper, the process in (2.1) is assumed to satisfy the following assumptions.

Assumption 1: (a) *All of the characteristic roots associated with* (2.1), *that is the solutions to the characteristic equation* $A(z) := (1-z)I_n - \alpha\beta'z - \Gamma_1 z(1-z) - \cdots - \Gamma_{k-1}z^{k-1}(1-z) = 0$, *lie either outside the unit circle or are equal to unity;* (b) $|\alpha'_\perp \Gamma \beta_\perp| \neq 0$, *with* $\Gamma := I_n - \Gamma_1 - \cdots - \Gamma_{k-1}$.

Assumption 2: *The innovations* $\{\varepsilon_t\}$ *form a martingale difference sequence with respect to the filtration* $\mathcal{F}_t$, *where* $\mathcal{F}_{t-1} \subseteq \mathcal{F}_t$ *for* $t = ..., -1, 0, 1, 2, ...$, *satisfying: (i) the global homoskedasticity condition:*

$$\frac{1}{T}\sum_{t=1}^{T} \mathrm{E}\left(\varepsilon_t \varepsilon'_t | \mathcal{F}_{t-1}\right) \xrightarrow{p} \Sigma > 0, \tag{2.3}$$

and (ii) $\mathrm{E}\left\|\varepsilon_t\right\|^4 \leq K < \infty$.

Assumption 1 is standard in the co-integration testing literature, while Assumption 2, which is used by Cavaliere et al. (2010a), implies that ε_t is a serially uncorrelated, potentially conditionally heteroskedastic process. The latter therefore contrasts with the assumption that ε_t is i.i.d. as made in Johansen (1996) and Swensen (2006).

In this paper we focus attention on the so-called trace test of Johansen (1996) for testing the pair of hypotheses

$$H_0(r) : \mathrm{rk}(\Pi) = r \quad vs. \quad H_1(r) : \mathrm{rk}(\Pi) = p. \tag{2.4}$$

The bias-correction based methods we outline in this paper could equally well be applied to the corresponding maximum eignevalue test of Johansen (1996) in an entirely obvious way.

As is standard, let $M_{ij} := T^{-1}\sum_{t=1}^{T} Z_{it}Z'_{jt}$, $i, j = 0, 1, 2$, with Z_{it} defined as in (2.2), and let $S_{ij} := M_{ij} - M_{i2}M_{22}^{-1}M_{2j}$, $i, j = 0, 1$. Let the p largest solutions to the eigenvalue problem, $\left|\lambda S_{11} - S_{10}S_{00}^{-1}S_{01}\right| = 0$, be denoted by $\hat{\lambda}_1 > \cdots > \hat{\lambda}_p$. The (pseudo) likelihood ratio [PLR] test for the pair of hypotheses in (2.4) then rejects for large values of the trace statistic[1]

---

[1]Notice that the subscript r in LR_r is a generic notation denoting the null rank being tested. If the specific rank $r = 0$ were being tested, for example, then the statistic would be referred to as LR_0. The same convention will be adopted for all other statistics introduced in this paper.

$$LR_r := -T \sum_{i=r+1}^{p} \log(1 - \hat{\lambda}_i).$$

Cavaliere et al. (2010a) derive the limiting null distribution for LR_r for data generated according to (2.1) under Assumptions 1 and 2, and this result is reproduced for convenience in the following theorem.

Theorem 1 *Let X_t be generated as in (2.1) under Assumptions 1 and 2. Then under $H_0(r)$,*

$$LR_r \xrightarrow{w} \mathsf{tr}(\mathcal{LR}_{r,B}) =: LR_{r,\infty} \tag{2.5}$$

where

$$\mathcal{LR}_{r,B} := \int_0^1 (dB_{p-r}(u))F_{p-r}(u)' \left(\int_0^1 F_{p-r}(u)F_{p-r}(u)'du \right)^{-1} \int_0^1 F_{p-r}(u)(dB_{p-r}(u))' \tag{2.6}$$

with $B_{p-r}(\cdot)$ a $(p-r)$-variate standard Brownian motion and $F_{p-r} := (B'_{p-r}, u|1)'$, where the notation $a|b := a(\cdot) - \int a(s)b(s)'ds(\int b(s)b(s)'ds)^{-1}b(\cdot)$ denotes the projection residuals of a onto b.

3 Bootstrap Co-integration Rank Tests

In this section we first outline in Algorithm 1 the standard recursive *i.i.d.* bootstrap trace test. We will then subsequently discuss how to adjust this bootstrap scheme to introduce bias-corrected parameter estimates. The bootstrap outlined in Algorithm 1 is closely related to the algorithm recently proposed in Cavaliere et al. (2012) which uses a bootstrap recursion based on the parameter estimates from (2.1) which obtain under the restriction of $H_0(r)$. As noted in section 1, alternative bootstrap procedures, such as Swensen (2006) and Cavaliere et al. (2010a) which use unrestricted estimates of the parameter matrices on the lagged dependent variables in (2.1), will not be discussed further here because their validity is not guaranteed when the true co-integrating rank exceeds the rank being tested; see Cavaliere et al. (2012) for full details on this point.

Algorithm 1

(1) *Perform a RR regression of (2.1) under the rank hypothesis $H_0(r) : rk(\Pi) = r$ in order to obtain the RR estimators $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\Gamma}_j$, $j = 1, \ldots, k-1$, and the corresponding residuals $\tilde{\varepsilon}_{k+1}, \ldots, \tilde{\varepsilon}_T$.*

(2) Construct the bootstrap sample data, y_t^*, $t = 1, \ldots, T$, recursively from

$$\Delta X_t^* = \tilde{\alpha}\tilde{\beta}' X_{t-1}^* + \sum_{j=1}^{k-1} \tilde{\Gamma}_j \Delta X_{t-j}^* + \varepsilon_t^*, \tag{3.1}$$

with the re-sampled residuals ε_t^* drawn randomly with replacement (i.i.d. re-sampling) from the estimated residuals $\tilde{\varepsilon}_{k+1}, \ldots, \tilde{\varepsilon}_T$. The initial values of the recursion, $X_{-k+1}^*, \ldots, X_0^*$, are set equal to $\mathbf{0}$.

(3) Obtain the bootstrap test statistic, LR_r^{1*} analogous to LR_r.

(4) The bootstrap p-value is then computed as, $p_{r,T}^*(LR_r) := 1 - G_{LR,r,T}^*(LR_r)$, where $G_{LR,r,T}^*(\cdot)$ denotes the conditional (on the original data) cumulative distribution function (cdf) of LR_r^{1*}.

Remark 3.1. Cavaliere et al. (2010a) establish the asymptotic validity of the test based on the bootstrap PLR statistic, LR_r^{1*}, from Algorithm 1. In particular they demonstrate that the bootstrap LR_r^{1*} statistic attains the same first-order limiting null distribution as the LR_r statistic. Formally, they show that under the conditions of Theorem 1, $LR_r^{1*} \overset{w}{\to}_p LR_{r,\infty}$ and that, as a consequence, the associated bootstrap p-value, $p_{r,T}^*(LR_r)$ is (asymptotically) uniformly distributed under the null hypothesis, leading to tests with (asymptotically) correct size.

Remark 3.2. As discussed in Cavaliere et al. (2010a), the unknown cdf, $G_{LR,r,T}^*(\cdot)$, required in Step 4 of Algorithm 1 can be estimated through numerical simulation. This is done by generating B (conditionally) independent bootstrap statistics, $LR_{r,b}^{1*}$, $b = 1, ..., B$, and computing the estimated bootstrap p-value of the test as

$$\tilde{p}_{r,T}^*(LR_r) := \frac{1}{B} \sum_{b=1}^{B} \mathbb{I}(LR_{r,b}^{1*} > LR_r) \tag{3.2}$$

For $B \to \infty$, we have that $\tilde{p}_{r,T}^*(LR_r) \overset{a.s.}{\to} p_{r,T}^*(LR_r)$; see e.g. Hansen (1996). Estimated p-values for the alternative bootstrap procedures discussed below can be obtained in the same way.

Remark 3.3. Algorithm 1 is a simplified version of Algorithm 1 in Cavaliere et al. (2012); the latter incorporates an estimate of the deterministic component in (3.1) and initialises the recursion in (3.1) with $X_{-k+1}^* = X_{-k+1}$, $j = -k + 1, ..., 0$. We adopt the simpler form in (3.1) motivated by the results in Cavaliere, Taylor & Trenkler (2013), but again the principles outlined here could equally well be applied to the bootstrap tests from Algorithm 1 of Cavaliere et al. (2012).

Remark 3.4. Algorithm 1 of Cavaliere et al. (2012) additionally includes the requirement that the following root condition

$$|\tilde{A}^*(z)| = 0 \;\Rightarrow\; |z| > 1 \text{ or } z = 1, \tag{3.3}$$

holds on the estimated lag polynomial matrix

$$\tilde{A}^*(z) := (1-z)I_k - \tilde{\alpha}\tilde{\beta}'z - \tilde{\Gamma}_1(1-z)z - ... - \tilde{\Gamma}_{k-1}(1-z)z^{k-1}.$$

Cavaliere et al. (2012) show that (3.3) is guaranteed to hold in large samples; however, if condition (3.3) is not met in a finite sample then the resulting bootstrap samples may become explosive. Consequently, it has been suggested not to implement a bootstrap algorithm in such a case; cf. Swensen (2006, Remark 1). However, as we will show later in the paper, considering bootstrap iterations in which (3.3) is violated does not appear to be harmful for the finite sample performance of the bootstrap co-integration rank test. Indeed, our results suggest that discarding these iterations can actually worsen the performance of the resulting bootstrap. We therefore argue that condition (3.3) can be safely ignored in applied work. Technically, Algorithm 1, in common with Algorithm 1 of Cavaliere et al. (2012), requires that $|\tilde{\alpha}'_\perp \tilde{\Gamma} \tilde{\beta}_\perp| \neq 0$, where $\tilde{\Gamma} := I_k - \tilde{\Gamma}_1 - ... - \tilde{\Gamma}_{k-1}$, but this condition is always satisfied in practice and so may be safely ignored.

Remark 3.5. Algorithm 1 outlines (a simplified version of) the bootstrap co-integration rank tests of Cavaliere et al. (2012) which are based on the use of an i.i.d. re-sampling device in step (2). Recently, Cavaliere, Rahbek & Taylor (2013a) have proposed an analogous bootstrap procedure whereby re-sampling is done via the wild bootstrap rather than i.i.d. re-sampling. They show that the resulting wild bootstrap rank tests are asymptotically valid in the case where the innovations, ε_t, in (2.1) display conditional and/or unconditional heteroskedasticity. For economy of notation and space we will outline our proposed methods in section 4 in the context of the tests from Algorithm 1 only. However, the same principles can equally be applied to the corresponding tests based on the wild bootstrap, thereby allowing their use in cases where the innovations are heteroskedastic. Indeed, we will provide some Monte Carlo results relating to the wild bootstrap and conditionally heteroskedastic shocks in section 5.

4   Implementing Bias-Correction

In this section we discuss how to perform bias-corrected based implementations of the bootstrap procedure outlined in Algorithm 1. We adopt the framework of Kilian (1998), in order to obtain a simple bootstrap-after-bootstrap [BaB] approach to bias-correcting

the VECM (restricted) parameter estimates used in the bootstrap recursion, (3.1). This approach is outlined below and essentially uses averages (taken across the bootstrap replications) of the deviations of the parameter estimates from estimating the equation in (3.1) under $H(r)$ relative to the corresponding original estimates from (2.1) to proxy the underlying bias in the estimates from (2.1). We will focus attention on bias-correction of the estimated stationary dynamics of the system, that is of the parameters on the lagged dependent variables in (2.1), since the finite sample size distortions observed in the standard bootstrap trace tests appear to vary according to the stationary dynamics of the system, for a given co-integrating rank. It is also possible, in cases where we are testing the null of rank $r > 0$, to bias-correct the estimated long-run parameter estimates from (2.1) and we will briefly discuss this possibility too.

Algorithm 2 (Bootstrap-after-Bootstrap (BaB))

(1) *Proceed as in steps 1 and 2 of Algorithm 1 in order to obtain the restricted estimate $\tilde{\Psi} := (\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{k-1})$ from (2.1) and to generate B_1 bootstrap data sets.*

(2) *Fit the model (2.1) to each bootstrap sample in order to obtain the average $\bar{\tilde{\Psi}}^* = \frac{1}{B_1} \sum_{b=1}^{B_1} \tilde{\Psi}_b^*$, where $\tilde{\Psi}_b^* := (\tilde{\Gamma}_1^*, \ldots, \tilde{\Gamma}_{k-1}^*)$ is the estimate of Ψ based on the b-th bootstrap sample. Define the resulting bootstrap bias estimate as $C_\psi^* := \bar{\tilde{\Psi}}^* - \tilde{\Psi}$. Then, construct the bias-corrected estimate $\hat{\Psi} := (\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}) = \tilde{\Psi} - C_\psi^*$.*

(3) *Apply Algorithm 1 whereby the estimates $\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{k-1}$ used in step (2) are replaced with the corresponding bias-corrected estimate $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}$.*

Remark 4.1. It is important to note that the bootstrap procedure used in step (1) of the BaB in Algorithm 2 is a completely separate bootstrap procedure from that used in step (3), and they are based on independent sets of bootstrap draws from the underlying residuals obtained in step (1) of Algorithm 1.

As discussed in Kilian (1998), the bias-correction device outlined above typically leads to larger estimates for the parameter values on the stationary dynamics from (2.1), and so it is likely that the root condition in (3.3) will be violated more frequently than when the original uncorrected estimates are used. As a response to this we will also consider two alternative bootstrap schemes. The first of these we label the switching bootstrap; here we use the uncorrected estimates if the bias-corrected estimates lead to a violation of (3.3). The second approach, labelled the root-correction bootstrap, follows the suggestion

of Kilian (1998) (see also Takimoto & Hosoya 2006) whereby we progressively reduce the magnitude of the bias-correction until we avoid a violation of (3.3) in the bias-corrected estimates. These two methods are detailed below.

Switching Bootstrap:

(1) *Proceed as in the bootstrap-after-bootstrap in order to obtain the bias-corrected estimates in $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}$. Check wether or not the root condition in (3.3) is violated with the estimates $\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{k-1}$ replaced by $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}$.*

(2) *If the root condition is not violated, then apply Algorithm 1, where the estimates $\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{k-1}$ used in step (2) are replaced by the corresponding bias-corrected estimates $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}$. Otherwise, proceed as described in Algorithm 1; i.e., use the uncorrected parameter estimates obtained in step (1).*

Root-Correction Bootstrap:

(1) *If the root condition in (3.3) is violated, then apply Algorithm 1.*

(2) *If the root condition in (3.3) is not violated, perform the bias correction as in the bootstrap-after-bootstrap. If the root condition in (3.3) is not violated with the estimates $\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{k-1}$ replaced by the biased-corrected ones $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}$, then apply Algorithm 1, where $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}$ are used in step (2) instead of the estimates $\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{k-1}$.*

(3) *If the use of $\hat{\Gamma}_1, \ldots, \hat{\Gamma}_{k-1}$ leads to a violation of the root condition in (3.3), then iteratively reduce the bias correction according to $C^*_{\psi,i+1} := \delta_i C^*_{\psi,i}$ with $\delta_{i+1} = \delta_i - 0.01$, $\delta_1 = 1$, and $C^*_{\psi,1} := C^*_\psi$, until the root condition in (3.3) is satisfied. Use the resulting shrunk bias-correction term to bias-correct the parameter estimates. Then, apply Algorithm 1 using the resulting bias-corrected estimates in step (2) of Algorithm 1 instead of the estimates $\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{k-1}$.*

Remark 4.2. Notice that the estimated parameters obtained in step (1) of Algorithm 1 are used in the root-correction bootstrap in cases where they violate the root condition in (3.3); i.e. no bias-correction takes place in such cases since it is very likely that bias-correcting here would further inflate the absolute values of the resulting roots. For the same reason, the switching bootstrap also typically employs the estimated parameters obtained in step (1) of Algorithm 1 if they violate the root condition (3.3).

Remark 4.3. The asymptotic validity of the three methods of bias-correction outlined above follows straightforwardly from the consistency results established in Cavaliere et al. (2012) for the parameter estimates obtained in step (1) of Algorithm 1. As a result, the bias-correction methods outlined above are asymptotically vanishing, such that for each of the three bias-correction schemes the resulting bootstrap trace statistic will be asymptotically equivalent to the standard bootstrap trace statistic from Algorithm 1. See also Kilian (1998).

Remark 4.4. As noted at the start of this section, the three schemes outlined above bias-correct only the estimated short run dynamics. In addition, where we are testing $H(r)$ with $r > 0$, we might also consider bias-correcting the estimates of α and β, and hence Π, from (2.1) under $H(r)$. This can be done in various ways: one could either bias correct $\tilde{\Pi} := \tilde{\alpha}'\tilde{\beta}$, or separately bias-correct $\tilde{\alpha}$ and $\tilde{\beta}$. On the face of it there would appear little to be gained in making these corrections because the estimates of α and β are restricted to satisfy the null co-integrating rank. However, simulation evidence in, among others Johansen (2002) and Cavaliere, Rahbek & Taylor (2010b), suggests that the finite sample behaviour of both the asymptotic and standard bootstrap trace tests can depend on α (but not β) and so it may be worth bias-correcting only $\tilde{\alpha}$. In order to bias-correct any of these estimates, the same approach as outlined in the three schemes above is used with an obvious change in notation. Again the impact of these bias-correction schemes would be asymptotically vanishing given the consistency of $\tilde{\alpha}$ and $\tilde{\beta}$ when $r > 0$. We will investigate these possibilities further in section 5.3.

5 Numerical Results

In this section we use Monte Carlo (MC) methods to compare the finite sample size and power properties of the various bootstrap tests outlined in sections 3 and 4.

We use a VAR(2) process for a variety of parameter configurations as our simulation DGPs; cf. Johansen (2002) and Swensen (2006). We will consider processes of dimensions $p = 2, 4$ and set the true co-integrating rank, r_0, equal to either zero, one, and two. The general model we use is therefore given by

$$\Delta X_t = \alpha\beta'x_{t-1} + \Gamma_1\Delta X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, I_p), \quad t = 1, ..., T, \qquad (5.1)$$

where α and β are $p \times r_0$ vectors and $\Gamma_1 = \xi I_n$ with $-1 < \xi < 1$. We consider $\xi = 0.5, 0.8, 0.9$. Notice that values of ξ close to one represent near-I(2) systems which is known to be a problem case for co-integration rank tests; see Johansen (2002) and Cavaliere et al. (2012), among others. For the non co-integrated case ($r_0 = 0$), we have $\alpha = \beta = 0$ (we

label this Case 0). For the $r_0 = 1$ case we follow, among others, Johansen (2002) and Swensen (2006), and consider the parameter combinations $\beta = (1, 0, ..., 0)'$ and $\alpha = (a_1, a_2, 0, ..., 0)'$, leading to the model (omitting equation (5.2c) when $p = 2$)

$$\Delta x_{1,t} = a_1 x_{1,t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_{1,t} \tag{5.2a}$$

$$\Delta x_{2,t} = a_2 x_{1,t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_{2,t} \tag{5.2b}$$

$$\Delta x_{i,t} = \Gamma_1 \Delta x_{t-1} + \varepsilon_{i,t}, \qquad i = 3, 4. \tag{5.2c}$$

We focus on the case $a_1 = a_2 = -0.4$ (which we label Case 1). Finally, for $r_0 = 2$ (which we label Case 2) we consider a four-dimensional[2] VAR and use $\alpha = \left(\begin{smallmatrix} -0.4 & -0.4 & 0 & 0 \\ 0 & -0.6 & 0 & 0 \end{smallmatrix} \right)'$, $\beta = \left(\begin{smallmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{smallmatrix} \right)'$. Hence, we have,

$$\Delta x_{1,t} = -0.4 x_{1,t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_{1,t} \tag{5.3a}$$

$$\Delta x_{2,t} = -0.4 x_{1,t-1} - 0.6 x_{2,t-1} + \Gamma_1 \Delta x_{t-1} + \varepsilon_{2,t} \tag{5.3b}$$

$$\Delta x_{i,t} = \Gamma_1 \Delta x_{t-1} + \varepsilon_{i,t}, \qquad i = 3, 4. \tag{5.3c}$$

All tests are run at the nominal 5% significance level. The computations are performed using the RNDNS function (with fixed seed) of GAUSS 10.0 for Windows. Note that the same sets of randomly generated error term vectors are used for all DGPs of the same dimension in order to eliminate this source of Monte Carlo variation from a comparison of the performance of the tests. The number of Monte Carlo replications is set at $R = 5000$. For determining the quantiles of the empirical bootstrap distributions we use $B = 499$ bootstrap replications. The number of first-level bootstrap replications used in the step (1) of the BaB scheme is set to $B_1 = 299$.

5.1 Empirical Size of Tests: $H_0 : r = 0$ and $H_0 : r = 1$

The results in Table 1 for tests of $H_0 : r = 0$ under Case 0 show that in those cases where excessive size distortions are observed in the standard bootstrap (Algorithm 1), these can be significantly reduced, or even completely avoided in many cases, by using the BaB (Algorithm 2). Indeed the BaB displays excellent finite sample size control throughout the results in Table 1. However, it is noteworthy that the bias-corrected estimate, $\hat{\Psi}$, used within the BaB quite often violates the root condition in (3.3) in very small samples ($T = 50$) and for large values of ξ. The large number of root violations for $\xi = 0.9$ is perhaps to be expected given that this parameter value represents a near-I(2) system. The use of the switching algorithm is also seen to improve on the standard bootstrap.

---

[2]Notice that $p = 2$ is not considered when $r_0 = 2$ because this would correspond to a purely $I(0)$ system.

Nevertheless, the BaB appears preferable because it avoids the over-rejections seen in the switching bootstrap in the four-dimensional system and/or in the case where $\xi = 0.9$.

It would appear from these results therefore that ignoring violations of the root condition (3.3) does not seem to be harmful for the finite sample performance of the bootstrap. One might expect, given these findings, that the rejection frequencies of the BaB would be lower for those replications where root violations occur than for those replications where root violations do not occur. However, this seems to generally not be the case as the auxiliary results in Table 2 demonstrate. Indeed, for $T = 50$ the rejection frequencies are rather higher on average where the root condition is violated than where it is not. To try and explain this result let us focus on the case where $p = 2$, $T = 50$ and $\xi = 0.9$. In this case, there are 1733 Monte Carlo replications in which the root condition in (3.3) is violated when using the bias-corrected estimates $\hat{\Psi}$ obtained within Algorithm 2. Part C of Panel B in Table 2 tells us that the rejection frequency is 0.0646 if the corrected estimates $\hat{\Psi}$ are used. We also simulated the standard bootstrap using the original estimates $\tilde{\Psi}$ over these 1733 replications and this resulted in a rejection frequency of 0.0952. Hence, conditional on the Monte Carlo replications in which (3.3) is violated when applying $\hat{\Psi}$, using $\hat{\Psi}$ in place of $\tilde{\Psi}$ results in lower rejection frequencies. Thus, it is correct to say that not using the Monte Carlo replications with root violations results in higher rejection frequencies, even though the rejection frequency of the BaB can be higher in case of root violations than in cases of no root violations. The same patterns are also found for the DGPs with $r_0 = 1$ (Case 1) and $r_0 = 2$ (Case 2) discussed below. Hence, the comments made above would appear to describe a general feature of the bootstrap rank tests.

The results for the root-correction bootstrap are rather similar to those for the BaB. Hence, and in contrast to switching to the uncorrected estimates $\tilde{\Psi}$, reducing the bias correction does not appear to alter the behaviour of the resulting bootstrap tests to any significant degree. The only exception is the setup with $p = 4$, $T = 50$, and $\xi = 0.9$ for which the root-correction bootstrap over-rejects somewhat more than the BaB. This again supports the view that one can safely ignore violations of the root condition in practice.

Consider now the results for the tests of $H_0 : r = 1$ under Case 1 which are reported in Table 3. In the case of $p = 2$, the standard bootstrap already performs well. Accordingly, not much is gained when using the BaB. In contrast, for $p = 4$ the standard bootstrap displays significant over-size for $\xi = 0.9$ and also displays under-sizing for $T = 50$ and $\xi = 0.5$. The BaB avoids this over-sizing although it can also be rather conservative for $T = 50$ when $\xi = 0.5$ or $\xi = 0.8$. The root-correction bootstrap shows similar performance to the BaB. The switching bootstrap also shows similar performance to the BaB for $p = 2$, but is not as well behaved as the BaB when $p = 4$. In the latter case it displays some over-

13

size for $\xi = 0.9$, although this is not as bad as for the standard bootstrap, and replicates the under-size seen in the BaB when $T = 50$ and $\xi = 0.5$. The number of root violations seen with the BaB is somewhat smaller than was the case for the corresponding results when testing $H_0 : r = 0$ under Case 0 in Table 1.

Overall, the use of the BaB appears to be beneficial in terms of reducing finite sample size distortions relative to those seen in the standard bootstrap, particularly so where strong positive serial correlation is seen in the stationary dynamics of the system. The benefits of using the BaB also appear clearer the larger the dimension of the system and the smaller the sample size. The performance of the root-correction bootstrap is generally quite similar to that of the BaB although it has a greater tendency to be over-sized than does the BaB. The switching bootstrap appears in general to be inferior to the BaB. On the basis of the numerical results presented in this subsection we recommend that practitioners may safely ignore any violations of the root condition in (3.3) which they encounter.

5.2   Finite Sample Power of Tests: $H_0 : r = 0$ and $H_0 : r = 1$

As the results in Tables 4 and 5 for Case 1 with $r_0 = 1$ and Case 2 with $r_0 = 2$, respectively, show, the finite sample power of the BaB can in some cases be a little lower than that of the standard bootstrap. This is of course purely an artefact of the over-sizing of the standard bootstrap tests discussed in the previous subsection; see Tables 1 and 3. For example, for $T = 50$ and $\xi = 0.8$ (Case 1, 2) or $\xi = 0.9$ (Case 2) the BaB has rather lower power than the standard bootstrap but of course it is precisely these examples where the most pronounced differences were seen between the empirical sizes of the BaB and standard bootstrap tests in Tables 1 and 3. Where the empirical sizes of the standard bootstrap and the BaB are similar to one another then, not surprisingly, we see very similar power properties across the two tests. This is reassuring because it suggests that there is no genuine loss in finite sample power from using the BaB, yet it can offer significant improvements in finite sample size over the standard bootstrap.

Again in line with the empirical size results from Tables 1 and 3, we observe that the switching bootstrap has higher power than the BaB in some cases because of its corresponding over-size under the null, while the power of the root-correction bootstrap is very similar to that of the BaB.

14

5.3 Bias-Correcting the Adjustment Matrix α

As discussed in section 4, we might also consider bias-correcting the estimate of α, in addition to the estimates of the parameters of the short-run dynamics contained in Ψ. Accordingly, some results for the null hypothesis $H_0 : r = 1$ for both $r_0 = 1$ (Case 1, empirical size) and $r_0 = 2$ (Case 2, empirical power) are reported in columns 2 - 8 and columns 9 - 12, respectively, of Table 6.

The results suggest that additionally bias-correcting $\tilde{\alpha}$ does not really bring any significant further improvements in finite sample size over bias correcting only the short-run dynamics. The only differences of note are seen in very small sample sizes for the larger dimensional system, i.e. $T = 50$ and $p = 4$, in relation to $\xi = 0.5$ and $\xi = 0.8$, where this effects a marginal reduction in the degree of under-size seen in the BaB tests; see Table 3 and Table 6 (columns 2 - 8). However, for $\xi = 0.9$ when $T = 50$ and $p = 4$ we see that the BaB and root-correction bootstrap tests which also bias-correcting $\tilde{\alpha}$ now over-reject the null. Moreover, the problem of excessive size distortion for the switching algorithm test discussed in subsection 5.1 is seen to worsen when $\tilde{\alpha}$ is also bias-corrected. Based on these results, we therefore recommend bias-correcting only the estimator of Ψ for all three bias-corrected bootstrap methods discussed in this paper.

5.4 Conditionally Heteroskedastic Error Terms

Finally, we consider two DGPs with conditionally heteroskedastic error terms covered by our Assumption 2. To be precise, we use two- and four-dimensional DGPs of the form given in (5.1) except that we replace the error term specification by the following two cases:

$$\text{ARCH}(1) : \varepsilon_{i,t} = h_{i,t}^{1/2} v_{i,t}, \ h_{i,t} = 0.5 + 0.5\varepsilon_{i,t-1}^2, v_{i,t} \sim iid\ N(0,1), \ i = 1, 2, 3 \qquad (5.4)$$

$$\text{Bilinear} : \varepsilon_{i,t} = e_{i,t}e_{i,t-1}, e_{i,t} \sim iid\ N(0,1), \ i = 1, 2, 3. \qquad (5.5)$$

Recall that we require the fourth moments of the error terms to be finite. This requirement holds for the bilinear specification in (5.5). Following Ling & McAleer (2000) finite fourth moments for ARCH(1) models require that $3a^2 < 1$, where a is the coefficient on $\varepsilon_{i,t-1}$. Hence, our ARCH(1) specification in (5.4) can therefore be seen to satisfy this condition.

Given the results of the previous subsections 5.1-5.3 we focus on the standard bootstrap test (Algorithm 1) and the BaB test (Algorithm 2) and we only consider bias-correcting $\tilde{\Psi}$. In addition to the i.i.d. re-sampling scheme applied in Algorithms 1 and 2 we also applied a re-sampling scheme based on the wild bootstrap, as discussed in Remark 3.5.

In this latter scheme, the bootstrap residuals are obtained by $\varepsilon_t^* = \tilde{\varepsilon}_t w_t$, $t = k + 1, \ldots, T$, where $\{w_t\}_{t=k+1}^T$ denotes an independent $N(0, 1)$ scalar sequence. Accordingly, we consider the corresponding wild bootstrap versions of the standard and BaB tests that can be obtained by appropriately adjusting Algorithms 1 and 2 to wild bootstrap re-sampling. We therefore analyse four bootstrap tests: the standard i.i.d. bootstrap test, the i.i.d. BaB test, the standard wild bootstrap test, and the wild BaB test.

Selected results relating to the empirical size properties of the four bootstrap tests are summarised in Table 7. We focus on Case 0 ($r_0 = 0$) with $\xi = 0.8$ and $\xi = 0.9$. Qualitatively similar results are obtained for the other cases and parameter specifications considered previously and, hence, are not reported here.

A few remarks on the results are in order. First, the rejection frequencies of the standard i.i.d. and BaB bootstrap tests tend to be higher where the error term is conditionally heteroskedastic than for i.i.d. errors; compare the results in Tables 1 and 7. Indeed under conditional heteroskedasticity the standard bootstrap test over-rejects quite strongly. Even the BaB test can have rather high rejection frequencies, in particular for $T = 50$. The number of violations of the root condition (3.3) is slightly higher on average for the conditionally heteroskedastic error term specifications than was the case under i.i.d. errors but the pattern was qualitatively similar and, as a consequence, we do not report these in Table 7. Our results on the DGPs with conditionally heteroskedastic error terms again suggest that one can safely ignore violations of the root condition (3.3).

The standard wild bootstrap test is less affected by excessive size distortions than the standard i.i.d. bootstrap test. Yet, the wild bootstrap version still rejects the true null hypothesis $H_0 : r = 0$ too often in most of the considered situations. The rejection frequencies are sometimes even higher than those of the i.i.d. BaB test; see, for example, the ARCH case where $p = 4$ and $\xi = 0.9$. Indeed, overall the i.i.d. BaB test displays better finite sample size control than the standard wild bootstrap test in case of the four-dimensional DGP.

Finally, the over-rejection which is observed for the i.i.d. BaB test in case of $T = 50$ is seen to be avoided by applying the wild BaB bootstrap test. However, the results also show that the wild BaB bootstrap test has the tendency to be conservative in some cases. Nevertheless, the BaB seems clearly preferable to the standard bootstrap framework for both the i.i.d. and wild bootstrap versions of the tests. These findings, coupled with those relating to i.i.d. innovations discussed previously in this section, lead us to strongly recommend the use of the BaB in practice.

6 Empirical Application

In this section we consider an empirical application based on Cavaliere et al. (2008) who study international dynamic consumption risk sharing within a set of European countries. As pointed out by Cavaliere et al. (2008) there is strong empirical evidence that domestic aggregate consumption is not well insured against idiosyncratic shocks; that is, risks are poorly shared internationally. Hence, the so-called 'full risk sharing hypothesis' (FRS) is not supported. This empirical result is explained by restrictions on international factor mobility and goods' trade as well as by other reasons such as habit persistence. Cavaliere et al. (2008) argue that if these frictions are not sufficiently large to keep consumers permanently away from the FRS-equilibrium, then departures from this equilibrium can be analysed through a dynamic model capturing the adjustment process.

Cavaliere et al. (2008) consider a set of countries forming a risk-sharing pool with one so-called 'leader' country. Based on a standard international business cycle model, they show that the FRS implies an equilibrium relationship between logged consumption of a country in the pool, say c_t^i, logged consumption of the leader country, say c_t^0, the logged real exchange rate between the countries, say r_t^i, and a linear trend. If c_t^i, c_t^0, and r_t^i are integrated of order one, this equilibrium relationship translates into a co-integration relation. Hence, one may interpret the FRS-equilibrium relation as a long-run anchor in the process of risk sharing adjustment. While Cavaliere et al. (2008) formally and empirically analyse the adjustment dynamics in detail we focus on co-integration testing as a pre-requisite for this type of analysis.

We use the same data set as analysed by Cavaliere et al. (2008). Accordingly, we consider Germany, France, Italy, Spain, the Netherlands, Belgium, Portugal, and Austria representing the core European Monetary Union (EMU) member countries and the United Kingdom as the most important non-EMU country of the European Union (EU). Germany will serve as the leader country such that we analyse eight country pairs with respect to Germany. Specifically, c_t^0 and c_t^i represent the log of real per-capita private final consumption expenditure for Germany and the other countries, respectively, and r_t^i is the log of the real exchange rates of the countries with respect to Germany. The annual data cover the relatively long period from 1961 to 2003, with the effective sample size $T = 41$. For data sources and further details we refer the reader to Cavaliere et al. (2008).

In addition to the asymptotic PLR rank test results reported in Cavaliere et al. (2008) we also report the corresponding results for the standard bootstrap and the BaB tests, in each case for both the i.i.d. and the wild bootstrap versions. We follow Cavaliere et al. (2008) and base the co-integration tests on three-dimensional VAR(2) models which

17

contain c_t^0, c_t^i, and r_t^i regarding the eight country pairs, respectively.[3] The resulting VEC model contains a linear trend restricted to the error-correction term, as specified in (2.2).

The results are summarized in Tables 8 and 9. We display the p-values of the rank tests and the estimates of the diagonal elements of Γ_1, the parameter matrix associated with the first lag of Δy_t in the VECM on which the corresponding null hypothesis is imposed. These estimates will facilitate the interpretation of the results.

Applying the asymptotic PLR rank test, one finds evidence for one co-integration relation for the France-Germany, Italy-Germany, and Belgium-Germany pairs.[4] In case of the Portugal-Germany pair two co-integration relations are indicated by the asymptotic tests. Basing inference on the bootstrap approach, however, overturns these test decisions with the exception of the Portugal-Germany pair. In fact, the bootstrap tests provide no evidence for co-integration except for the Portugal-Germany pair. Indeed, for the Portugal-Germany pair, while the wild bootstrap tests yield the same conclusion as the asymptotic test results, at least at the 10% significance level, the i.i.d. bootstrap tests suggest a co-integration rank of one at this significance level as the theory of international risk sharing implies. Hence, in this example we see that the choice of a specific bootstrap test can be crucial.

In general we see that the p-values of the bootstrap tests are higher than those of the corresponding asymptotic rank test. This would not be a surprising result if the true rank were equal to zero. The simulation results of Cavaliere et al. (2013) show that the asymptotic test is significantly oversized when testing the true null hypothesis $H_0 : r = 0$. For the bivariate and four-dimensional VAR(2) processes with $\xi = 0.5$ considered in our simulation study, the empirical rejection frequencies of the asymptotic test are 0.1140, 0.0778, and 0.0646 ($p = 2$) and 0.4708, 0.2086, and 0.1076 ($p = 4$) for $T = 50$, $T = 100$, and $T = 200$, respectively. These size distortions are even further amplified if we set ξ to 0.8 or 0.9. Since only small samples are available we may expect rather large size distortions for the asymptotic test resulting in relatively low p-values if the true rank is $r_0 = 0$. In contrast, even the standard i.i.d. bootstrap test can greatly reduce the size distortions in such a situation as our simulation results in Table 1 demonstrate.

In relative terms, the differences between the p-values of the i.i.d. BaB tests and the corresponding standard i.i.d. bootstrap tests are much smaller than the differences be-

---

[3]Note that Cavaliere et al. (2008) use $k = 1$ lag for the Austria-Germany system. However, we decided to set $k = 2$ for illustrative purposes since there would be otherwise no parameter estimates with respect to the lagged first differences.

[4]If a VAR(1) model is used for the Austria-Germany pair as in Cavaliere et al. (2008), then the asymptotic test also suggests $r = 1$ for this pair. Using the critical values from Johansen (1996), as used in Cavaliere et al. (2008), one co-integration relation is found at the 10% level for the UK-Germany pair.

tween the rejection frequencies of the latter and the asymptotic tests. A similar conclusion holds for the wild bootstrap test outcomes. The simulation results from section 5 suggest that it is where diagonal elements of Γ_1 are large that we might expect to see significant differences between the p-values associated with the BaB and standard bootstrap tests; see, for example, Table 1. The estimates of the diagonal elements of Γ_1, shown in the last three columns of Tables 8 and 9, are in the main smaller than 0.5 suggesting that one might not expect very large changes from bias-correcting these parameter estimates. In general the empirical results appear to follow this prediction. However, in almost all cases bias-correcting a given bootstrap test leads to an increased p-value, with this being more pronounced the larger the estimates of the diagonal elements of Γ_1.

7 Conclusions

Bootstrap implementations of the likelihood ratio co-integrating rank tests of Johansen (1996) are known to deliver significant improvements in finite sample behaviour, relative to the use of asymptotic critical values. However, notable finite sample size distortions are still observed with these bootstrap tests when the stationary dynamics of the underlying VAR process shows strong serial correlation and/or conditional heteroskedasticity. In this paper we have shown that these size distortions can be all but eliminated by the use of a bootstrap-after-bootstrap method, whereby the parameter estimates from the estimated co-integrated VAR model are bias-corrected using the framework of Kilian (1998). Here estimates of the bias in the original parameter estimates are obtained from the average bias in the corresponding parameter estimates taken across a large number of auxiliary bootstrap replications. A number of possible implementations of this procedure were discussed.

Our results lead us to strongly recommend the use of the bootstrap-after-bootstrap method, either in conjunction with i.i.d. bootstrap or wild bootstrap co-integration rank tests, and that violations of the so-called root conditions in these two bootstrap procedures can be safely ignored in practice, as may extensions to the basic bootstrap-after-bootstrap method involving either a switching bootstrap or root-correction bootstrap. Moreover, applying the bootstrap-after-bootstrap method to the long run component of the system did not appear to deliver any noticeable gains in accuracy once the sort run dynamics of the system had been bias-corrected.

An application of the bootstrap techniques discussed in this paper to re-assess the empirical evidence provided by Cavaliere et al. (2008) on international dynamic consumption risk sharing within a set of European countries was also considered. We found consider-

ably less evidence in favour of co-integration than was found by Cavaliere et al. (2008) on the basis of asymptotic co-integration rank tests, casting doubt upon the existence of a risk-sharing equilibrium serving as a long-run anchor for consumption streams within the EU.

Finally, in this paper we have followed Cavaliere et al. (2012) and focused our attention on the case where either no deterministic component is present or where either a restricted constant or restricted trend model is considered. The same principles as are outlined in this paper can also be straightforwardly applied in the context of the unrestricted constant or unrestricted trend models, using the corresponding bootstrap co-integration rank tests recently proposed in Cavaliere, Rahbek & Taylor (2013b).

References

Cavaliere, G., Fanelli, L. & Gardini, A. (2008). International dynamic risk sharing, *Applied Econometrics,* **23**: 1–16.

Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2010a). Co-integration rank testing under conditional heteroskedasticity, *Econometric Theory,* **26**: 1719–1760.

Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2010b). Testing for co-integration in vector autoregressions with non-stationary volatility, *Journal of Econometrics,* **158**: 7–24.

Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2012). Bootstrap sequential determination of the Co-integration rank in VAR models, *Econometrica,* **80**: 1721–1740.

Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2013a). Bootstrap Determination of the Co-integration Rank in Heteroskedastic VAR Models, *Econometric Reviews,* **forthcoming**.

Cavaliere, G., Rahbek, A. & Taylor, A.M.R. (2013b). Bootstrap Determination of the co-integration rank in VAR models with unrestricted deterministic time trends, Manuscript.

Cavaliere, G., Taylor, A.M.R. & Trenkler, C. (2013). Bootstrap Co-integration Rank Testing: The role of deterministic variables and initial values in the bootstrap recursion, forthcoming in *Econometric Reviews*.

Engsted, T. & Pedersen, T.Q. (2011). Bias-correction in vector autoregressive models: a simulation study, *CREATES Research Paper 2011-18*.

Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica* **64**: 413–30.

Johansen, S. (1996). *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford (second edition).

Johansen, S. (2002). A small sample correction for the test of cointegrating rank in the vector autoregressive model, *Econometrica* **70**: 1929–1961.

Kilian, L. (1998). Small-sample confidence intervals for impulse response functions *The Review of Economics and Statistics* **80**: 218–230.

Ling, S. & McAleer, M. (2000). Stationarity and the existence of moments of family of GARCH processes, *Journal of Econometrics* **106**: 109–117.

MacKinnon, J. G., Haug, A. A & Michelis, L. (1999). mackinnon:99, *Journal of Applied Econometrics,* **14**: 563–577.

Swensen, A. R. (2006). Bootstrap algorithms for testing and determining the cointegration rank in VAR models, *Econometrica* **74**: 1699–1714.

Takimoto, T. & Hosoya, Y. (2006). Inference on the cointegration rank and a procedure for VARMA root-modification, *Journal of the Japan Statistical Society* **36**: 149-171.

Table 1. Number of Violations of Root Condition (3.3) and Rejection Frequencies of Tests for VAR(2) (5.1) with $r_0 = 0$ (Case 0: $a_1 = a_2 = 0$), $H_0 : r = 0$ (Empirical Size), Significance Level 0.05

	Panel A: $p = 2$						Panel B: $p = 4$					
	Standard Bootstrap		BaB Bootstrap		Switching Bootstrap	Root-corr. Bootstrap	Standard Bootstrap		BaB Bootstrap		Switching Bootstrap	Root-corr. Bootstrap
	rej. freq.	root viol.	rej. freq.	root viol.	rej. freq.	rej. freq.	rej. freq.	root viol.	rej. freq.	root viol.	rej. freq.	rej. freq.
	Part A: $\xi = 0.5$						Part A: $\xi = 0.5$					
$T = 50$	0.0532	0	0.0484	0	—	—	0.0644	0	0.0456	0	—	—
$T = 100$	0.0520	0	0.0500	0	—	—	0.0602	0	0.0522	0	—	—
$T = 200$	0.0518	0	0.0504	0	—	—	0.0520	0	0.0504	0	—	—
	Part B: $\xi = 0.8$						Part B: $\xi = 0.8$					
$T = 50$	0.0636	0	0.0446	232	0.0464	0.0446	0.1058	0	0.0458	989	0.0594	0.0468
$T = 100$	0.0512	0	0.0436	0	—	—	0.0792	0	0.0448	3	0.0448	0.0448
$T = 200$	0.0548	0	0.0530	0	—	—	0.0584	0	0.0448	0	—	—
	Part C: $\xi = 0.9$						Part C: $\xi = 0.9$					
$T = 50$	0.0766	13	0.0524	1733	0.0630	0.0530	0.1392	33	0.0678	3413	0.1162	0.0802
$T = 100$	0.0638	0	0.0416	261	0.0428	0.0414	0.1114	0	0.0498	1213	0.0670	0.0514
$T = 200$	0.0592	0	0.0502	0	—	—	0.0758	0	0.0416	3	0.0416	0.0416

Note: The entries in the columns labeled "root viol." represent the number of MC replications in which the root condition (3.3) is violated by the relevant parameter estimates used within the corresponding bootstrap algorithm.

Table 2. Results on the Bootstrap-after-Bootstrap: Number of Violations of Root Condition (3.3) and Rejection Frequencies for VAR(2) (5.1) with $r_0 = 0$ (Case 0: $a_1 = a_2 = 0$), $H_0 : r = 0$ (Empirical Size), Significance Level 0.05.

	Panel A: $p = 2$				Panel B: $p = 4$			
	overall rej. freq.	root okay rej. freq.	root viol. rej. freq.	# root violations	overall rej. freq.	root okay rej. freq.	root viol. rej. freq.	# root violations
	Part A: $\xi = 0.5$				Part A: $\xi = 0.5$			
$T = 50$	0.0484	0.0484	—	0	0.0456	0.0456	—	0
$T = 100$	0.0500	0.0500	—	0	0.0522	0.0522	—	0
$T = 200$	0.0504	0.0504	—	0	0.0504	0.0504	—	0
	Part B: $\xi = 0.8$				Part B: $\xi = 0.8$			
$T = 50$	0.0446	0.0440	0.0560	232	0.0458	0.0421	0.0607	989
$T = 100$	0.0436	0.0414	—	0	0.0448	0.0448	0.0000	3
$T = 200$	0.0530	0.0508	—	0	0.0448	0.0448	—	0
	Part C: $\xi = 0.9$				Part C: $\xi = 0.9$			
$T = 50$	0.0524	0.0459	0.0646	1733	0.0678	0.0447	0.0785	3413
$T = 100$	0.0416	0.0414	0.0460	261	0.0498	0.0481	0.0552	1213
$T = 200$	0.0502	0.0502	—	0	0.0416	0.0416	0.0000	3

Note: The entries in the columns labeled "overall rej. freq." represent the rejection frequencies of the BaB over all MC replications. By contrast, in the columns labeled "root okay rej. freq." ("root viol. rej. freq."), the rejection frequencies of the BaB for the MC replications that satisfy (do not satisfy) the root condition are shown. The number of MC replications in which the root condition (3.3) is violated is shown in the columns labeled with "# root violations".

Table 3. Number of Violations of Root Condition (3.3) and Rejection Frequencies of Tests for VAR(2) (5.1) with $r_0 = 1$ (Case 1: $a_1 = a_2 = -0.4$), $H_0 : r = 1$ (Empirical Size), Significance Level 0.05. Only the $\tilde{\Gamma}_i$'s are adjusted.

	Panel A: $p = 2$						Panel B: $p = 4$					
	Standard Bootstrap		BaB Bootstrap		Switching Bootstrap	Root-corr. Bootstrap	Standard Bootstrap		BaB Bootstrap		Switching Bootstrap	Root-corr. Bootstrap
	rej. freq.	root viol.	rej. freq.	root viol.	rej. freq.	rej. freq.	rej. freq.	root viol.	rej. freq.	root viol.	rej. freq.	rej. freq.
Part A: $\xi = 0.5$												
$T=50$	0.0470	0	0.0468	0	—	—	0.0326	0	0.0264	2	0.0264	0.0254
$T=100$	0.0498	0	0.0492	0	—	—	0.0512	0	0.0482	0	—	—
$T=200$	0.0536	0	0.0536	0	—	—	0.0526	0	0.0512	0	—	—
Part B: $\xi = 0.8$												
$T=50$	0.0536	2	0.0488	129	0.0488	0.0488	0.0792	4	0.0396	813	0.0464	0.0410
$T=100$	0.0520	0	0.0496	0	—	—	0.0640	0	0.0444	2	0.0444	0.0444
$T=200$	0.0548	0	0.0544	0	—	—	0.0552	0	0.0490	0	—	—
Part C: $\xi = 0.9$												
$T=50$	0.0590	49	0.0460	1076	0.0494	0.0468	0.1044	76	0.0530	3040	0.0864	0.0588
$T=100$	0.0546	0	0.0480	82	0.0482	0.0480	0.0818	0	0.0414	743	0.0470	0.0418
$T=200$	0.0508	0	0.0490	0	—	—	0.0646	0	0.0430	1	0.0430	0.0430

Note: see Table 1.

Table 4. Number of Violations of Root Condition (3.3) and Rejection Frequencies of Tests for VAR(2) (5.1) with $r_0 = 1$ (Case 1: $a_1 = a_2 = -0.4$), $H_0 : r = 0$ (Empirical Power), Significance Level 0.05

| | Panel A: $p = 2$ | | | | | | Panel B: $p = 4$ | | | | | |
| | Standard Bootstrap | | BaB Bootstrap | | Switching Bootstrap | Root-corr. Bootstrap | Standard Bootstrap | | BaB Bootstrap | | Switching Bootstrap | Root-corr. Bootstrap |
	rej. freq.	root viol.	rej. freq.	root viol.	rej. freq.	rej. freq.	rej. freq.	root viol.	rej. freq.	root viol.	rej. freq.	rej. freq.
							Part A: $\xi = 0.5$					
$T = 50$	0.7498	0	0.7334	0	—	—	0.2814	0	0.2288	0	—	—
$T = 100$	1.0000	0	1.0000	0	—	—	0.9150	0	0.9052	0	—	—
$T = 200$	1.0000	0	1.0000	0	—	—	1.0000	0	1.0000	0	—	—
							Part B: $\xi = 0.8$					
$T = 50$	0.9924	0	0.9890	199	0.9892	0.9890	0.8262	1	0.7264	794	0.7420	0.7288
$T = 100$	1.0000	0	1.0000	0	—	—	0.9998	0	0.9996	6	0.9996	0.9996
$T = 200$	1.0000	0	1.0000	0	—	—	1.0000	0	1.0000	0	—	—
							Part C: $\xi = 0.9$					
$T = 50$	0.9994	9	0.9994	1175	0.9994	0.9994	0.9694	26	0.9452	3026	0.9564	0.9486
$T = 100$	1.0000	0	1.0000	141	1.0000	1.0000	1.0000	0	1.0000	882	1.0000	1.0000
$T = 200$	1.0000	0	1.0000	0	—	—	1.0000	0	1.0000	2	1.0000	1.0000

Note: see Table 1.

Table 5. Number of Violations of Root Condition (3.3) and Rejection Frequencies of Tests four-dimensional VAR(2) (5.1) with $r_0 = 2$ (Case 2: $a_{11} = a_{21} = -0.4$, $a_{22} = -0.6$), $H_0 : r = 1$ (Empirical Power), Significance Level 0.05.

	Standard Bootstrap		Bootstrap-after Bootstrap		Switching Bootstrap	Root-corr. Bootstrap
	rej. freq.	root viol.	rej. freq.	root viol.	rej. freq.	rej. freq.
	Part A: $\xi = 0.5$					
$T = 50$	0.1344	0	0.1142	0	—	—
$T = 100$	0.7676	0	0.7604	0	—	—
$T = 200$	1.0000	0	1.0000	0	—	—
	Part B: $\xi = 0.8$					
$T = 50$	0.4666	57	0.3616	691	0.3750	0.3652
$T = 100$	0.9872	1	0.9824	14	0.9824	0.9824
$T = 200$	1.0000	0	1.0000	0	—	—
	Part C: $\xi = 0.9$					
$T = 50$	0.7532	533	0.6540	3110	0.7128	0.6842
$T = 100$	0.9996	90	0.9988	850	0.9988	0.9988
$T = 200$	1.0000	3	1.0000	19	1.0000	1.0000

Note: see Table 1.

Table 6. Number of Violations of Root Condition (3.3) and Rejection Frequencies of Tests for VAR(2) (5.1) with $r_0 = 1$, $H_0 : r = 1$, Significance Level 0.05, $\tilde{\Psi} = (\tilde{\Gamma}_1, \ldots, \tilde{\Gamma}_{k-1})$ and $\tilde{\alpha}$ are adjusted.

	Panel A: $p = 2$, Case 1 ($a_1 = a_2 = -0.4$, Empirical Size)				Panel B: $p = 4$, Case 1 ($a_1 = a_2 = -0.4$, Empirical Size)				Panel C: $p = 4$, Case 2 ($a_{11} = a_{21} = -0.4$, $a_{22} = -0.6$, Empirical Power)			
	BaB Bootstrap		Switching Bootstrap	Root-corr. Bootstrap	BaB Bootstrap		Switching Bootstrap	Root-corr. Bootstrap	BaB Bootstrap		Switching Bootstrap	Root-corr. Bootstrap
	rej. freq.	root viol.	rej. freq.	rej. freq.	rej. freq.	root viol.	rej. freq.	rej. freq.	rej. freq.	root viol.	rej. freq.	rej. freq.
Part A: $\xi = 0.5$												
$T = 50$	0.0446	1	0.0446	0.0446	0.0286	2	0.0286	0.0286	0.1160	22	0.1160	0.1160
$T = 100$	0.0492	0	—	—	0.0488	0	—	—	0.7526	0	—	—
$T = 200$	0.0538	0	—	—	0.0520	0	—	—	1.0000	0	—	—
Part B: $\xi = 0.8$												
$T = 50$	0.0482	144	0.0484	0.0484	0.0478	842	0.0532	0.0492	0.3926	1338	0.4020	0.4022
$T = 100$	0.0488	0	—	—	0.0460	2	0.0460	0.0460	0.9830	393	0.9830	0.9832
$T = 200$	0.0544	0	—	—	0.0492	0	—	—	1.0000	43	1.0000	1.0000
Part C: $\xi = 0.9$												
$T = 50$	0.0470	1090	0.0502	0.0472	0.0698	3062	0.0906	0.0716	0.6908	3367	0.7258	0.7130
$T = 100$	0.0478	83	0.0480	0.0478	0.0436	739	0.0490	0.0442	0.9988	1699	0.9988	0.9990
$T = 200$	0.0482	0	—	—	0.0434	1	0.0434	0.0434	1.0000	328	1.0000	1.0000

Note: see Table 1.

Table 7. Rejection Frequencies of Tests for VAR(2) (5.1) with $r_0 = 0$ (Case 0: $a_1 = a_2 = 0$), $H_0 : r = 0$ (Empirical Size), Significance Level 0.05

	Panel A: $p = 2$				Panel B: $p = 4$			
	i.i.d.-Bootstraps		Wild Bootstraps		i.i.d.-Bootstraps		Wild Bootstraps	
	Standard Bootstrap	BaB Bootstrap	Standard Bootstrap	BaB Bootstrap	Standard Bootstrap	BaB Bootstrap	Standard Bootstrap	BaB Bootstrap
Part A: ARCH-error term specification (5.4), $\xi = 0.8$								
$T = 50$	0.0800	0.0578	0.0522	0.0348	0.1480	0.0814	0.0952	0.0450
$T = 100$	0.0608	0.0528	0.0480	0.0374	0.0994	0.0638	0.0674	0.0366
$T = 200$	0.0600	0.0570	0.0506	0.0472	0.0716	0.0570	0.0582	0.0456
Part B: ARCH-error term specification (5.4), $\xi = 0.9$								
$T = 50$	0.1006	0.0714	0.0660	0.0428	0.1904	0.1140	0.1242	0.0700
$T = 100$	0.0750	0.0572	0.0548	0.0386	0.1426	0.0706	0.1020	0.0396
$T = 200$	0.0704	0.0596	0.0532	0.0458	0.0928	0.0544	0.0728	0.0414
Part C: Bilinear-error term specification (5.5), $\xi = 0.8$								
$T = 50$	0.0914	0.0714	0.0596	0.0438	0.1384	0.0786	0.0838	0.0388
$T = 100$	0.0622	0.0538	0.0478	0.0400	0.0904	0.0536	0.0618	0.0316
$T = 200$	0.0596	0.0564	0.0536	0.0494	0.0714	0.0592	0.0554	0.0426
Part D: Bilinear-error term specification (5.5), $\xi = 0.9$								
$T = 50$	0.1070	0.0852	0.0760	0.0490	0.1914	0.1122	0.1168	0.0554
$T = 100$	0.0766	0.0592	0.0596	0.0400	0.1382	0.0712	0.0996	0.0392
$T = 200$	0.0644	0.0552	0.0542	0.0454	0.0898	0.0522	0.0686	0.0344

Note: p-values for the asymptotic rank test have been obtained from the approach of MacKinnon, Haug & Michelis (1999).

Table 8. Cointegration Test Results (p-values) and estimates of Γ_{ii}, $i = 1, 2, 3$, for the country-specific VAR(2) systems, Part 1

H_0	Panel A: p-values					Panel B: Parameter Estimates		
	asymptotic rank test	i.i.d.-Bootstraps Standard Bootstrap	i.i.d.-Bootstraps BaB Bootstrap	Wild Bootstraps Standard Bootstrap	Wild Bootstraps BaB Bootstrap	$\hat{\Gamma}_{11}$	$\hat{\Gamma}_{22}$	$\hat{\Gamma}_{33}$
Part A: France								
$r = 0$	0.0181	0.1030	0.1146	0.1622	0.1746	0.164	0.451	0.089
$r = 1$	0.3074	0.0467	0.4931	0.3557	0.3739	0.172	0.369	0.146
$r = 2$	0.2503	0.2291	0.2365	0.1920	0.2032	0.334	0.227	0.148
Part B: Italy								
$r = 0$	0.0520	0.2058	0.2283	0.3705	0.3873	0.277	0.262	0.185
$r = 1$	0.3408	0.5053	0.5395	0.4527	0.4689	0.503	0.388	-0.014
$r = 2$	0.3629	0.2561	0.2573	0.2437	0.2347	0.396	0.214	-0.025
Part C: Spain								
$r = 0$	0.1972	0.5055	0.5493	0.6751	0.7039	0.287	0.588	0.175
$r = 1$	0.4208	0.5795	0.6089	0.6967	0.7229	0.165	0.570	0.339
$r = 2$	0.2451	0.2351	0.2841	0.2611	0.2957	0.052	0.600	0.333
Part D: Netherlands								
$r = 0$	0.3456	0.7001	0.7398	0.6631	0.7181	0.088	0.672	0.117
$r = 1$	0.4755	0.6671	0.7019	0.6451	0.6741	0.006	0.565	0.170
$r = 2$	0.4158	0.4927	0.5187	0.4927	0.5273	0.005	0.593	0.090

Note: p-values for the asymptotic rank test have been obtained from the approach of MacKinnon et al. (1999).

Table 9. Cointegration Test Results (p-values) and estimates of Γ_{ii}, $i = 1, 2, 3$, for the country-specific VAR(2) systems, Part 2

| | | Panel A: p-values | | | | | Panel B: Parameter Estimates | | |
| | | | i.i.d.-Bootstraps | | Wild Bootstraps | | | | |
H_0	asymptotic rank test	Standard Bootstrap	BaB Bootstrap	Standard Bootstrap	BaB Bootstrap	$\hat{\Gamma}_{11}$	$\hat{\Gamma}_{22}$	$\hat{\Gamma}_{33}$
				Part A: Belgium				
$r = 0$	0.0756	0.2629	0.2851	0.2088	0.2293	0.308	0.272	0.276
$r = 1$	0.5750	0.7402	0.7582	0.6711	0.6927	0.285	0.146	0.476
$r = 2$	0.3273	0.2088	0.2080	0.2026	0.2038	0.197	0.165	0.483
				Part B: Portugal				
$r = 0$	0.0069	0.0654	0.0792	0.0488	0.0598	0.314	0.643	0.219
$r = 1$	0.0281	0.1320	0.1312	0.0802	0.0798	0.345	-0.026	0.250
$r = 2$	0.2203	0.3801	0.3813	0.2939	0.2981	0.356	-0.026	0.517
				Part C: Austria				
$r = 0$	0.2158	0.4359	0.4521	0.3681	0.3873	0.353	-0.001	0.186
$r = 1$	0.5207	0.6863	0.6929	0.5681	0.5781	0.266	-0.064	0.198
$r = 2$	0.5160	0.5565	0.5509	0.5945	0.5969	0.328	-0.086	0.091
				Part D: United Kingdom				
$r = 0$	0.1025	0.3251	0.3519	0.2497	0.2827	0.345	0.362	0.295
$r = 1$	0.1815	0.2973	0.3189	0.2337	0.2617	0.321	0.362	0.410
$r = 2$	0.3280	0.3909	0.3969	0.3209	0.3405	0.190	0.477	0.456

Note: see Table 8.