

Alternative Bootstrap Procedures for Testing Cointegration in Fractionally Integrated Processes

James Davidson*
Cardiff University

February 2004, forthcoming in *Journal of Econometrics*

Abstract

This paper considers alternative methods of testing cointegration in fractionally integrated processes, using the bootstrap. The special feature of the fractional case is the dependence of the asymptotic null distributions of conventional statistics on the fractional integration parameter. Such tests are said to be asymptotically non-pivotal, and conventional asymptotic tests are therefore not available. Bootstrap tests can be constructed, although these may be less reliable in small samples than in the case of asymptotically pivotal statistics. Bias correction techniques, including the double bootstrap of Beran (1988) and the fast double bootstrap of Davidson and MacKinnon (2000) are considered. The investigation focuses on the issues of (a) choice of statistic, (b) bias correction techniques, and also (c) designing the simulation of the null hypothesis. The latter consideration is crucial for ensuring tests are both correctly sized and powerful.

Three types of test are considered that, being envisaged as routine exploratory tools, are based on the residuals from a putative cointegrating regression. Two are of the null hypothesis of non-cointegration; a conventional residual-based test using the Durbin-Watson statistic, and a test based on the F -statistic, as proposed in Davidson (2002). The third is the Shin (1994) residual-based test of the null hypothesis that cointegration exists. The tests are compared in Monte Carlo experiments whose main object is to throw light on the relative roles of issues (a), (b) and (c) in test performance.

1 Introduction

Except in a few simple situations, tests in econometrics are performed by the method of first-order asymptotic approximation, employing a test statistic whose asymptotic distribution under the null hypothesis is free of nuisance parameters, and so can be tabulated. Such statistics are said to be asymptotically pivotal. They typically depend on nuisance parameters, even though their asymptotic distributions do not, and these must be replaced by consistent estimates. The true rejection probability of the test accordingly differs from the nominal probability, generally by an error of order $T^{-1/2}$ where T is the sample size. Following Davidson and MacKinnon (1999) we refer to this error in rejection probability as the ERP.

In the bootstrap procedure, Monte Carlo simulation is used to estimate the null distribution of a test statistic directly. In this case, estimated parameters are used to construct the data generation process (DGP) from which the replications are drawn.¹ As is well known (among

*Email: davidsonje@cf.ac.uk. Research supported by the ESRC under award L138251025..

¹In this context the expression DGP is used specifically to refer to a computer algorithm that generates a series whose joint distribution is intended to mimic that of an observed sample.

many references, see Hall (1992), and Horowitz (2000) for a survey) bootstrap tests based on asymptotically pivotal statistics may achieve an ERP of smaller order in T than first-order asymptotic tests, yielding the so-called bootstrap refinements. However, the bootstrap method can also be used to derive tests based on non-asymptotically pivotal statistics, allowing consideration of procedures having no counterpart in the usual armoury of asymptotic tests. The potential benefits of this approach are clear. New tests might be constructed, on an *ad hoc* basis and with minimal theoretical labour, with a view to maximizing power against alternatives of interest. The strong assumptions needed to justify particular established procedures might often be dispensed with. Cointegration tests provide an excellent illustration of these points. There also exist bias reduction techniques for bootstrap tests such as prepivoting (Beran 1988), that can confer some of the advantages of asymptotically pivotal tests on the general case.

In this paper, bootstrap methods for tests of cointegration in time series that are fractionally integrated are proposed and compared. In the terminology of Granger (1986), a collection of time series are said to be $CI(d,b)$ if the series are fractionally integrated of order d (or $I(d)$) while there exists a linear combination of the series which is $I(d-b)$ for $b > 0$. A difficulty with this testing problem is that the distributions of the usual test statistics, under the null hypothesis $b = 0$, depend on nuisance parameters asymptotically. While it is possible to devise an asymptotically pivotal test for this case based on the Johansen (1988, 1991) procedure (see Breitung and Hassler 2001) the focus of attention here is on simple residual based statistics, that are not asymptotically pivotal.

Section 2 of the paper describes the fractional cointegration setup, and considers a number of possible tests, both of the hypothesis of non-cointegration and of the hypothesis of cointegration, i.e., of $I(0)$ cointegrating residuals. Section 3 reviews the bootstrap procedure itself, including the ‘double bootstrap’ and ‘fast-double bootstrap’ variants. Section 4 reports the results of Monte Carlo experiments to investigate the properties of the tests. Section 5 gives an application of the procedures to US quarterly consumption, income and wealth data. Section 6 contains concluding remarks. The appendices contain proofs, and some supplementary results on the asymptotics of the cointegration test applied to fractional processes.

2 The Hypotheses and Bootstrap Test Statistics

2.1 Fractional Cointegration Setup

A random process \mathbf{x}_t ($m \times 1$) is said to be fractionally integrated of order d , written $\mathbf{x}_t \sim I(d)$, if

$$\mathbf{v}_t = \Delta^d \mathbf{x}_t \sim I(0) \quad (2.1)$$

where

$$\Delta^d = (1 - L)^d = 1 - \sum_{j=1}^{\infty} \frac{d\Gamma(j-d)}{\Gamma(1-d)\Gamma(j+1)} L^j \quad (2.2)$$

is the the fractional difference operator.² Here, d is in principle any positive real, but we consider cases $d \in (\frac{1}{2}, \frac{3}{2})$. ‘ $I(0)$ ’ is a concept often loosely defined in the literature to mean stationary and/or short-memory, but we shall understand it to mean, strictly, that the process satisfies the

²In this paper, this expansion is truncated at $j = t - 1$ for $t = 1, \dots, n$, for purposes of computation.

weak convergence property³ $\mathbf{V}_T \rightarrow_d \mathbf{B}(\boldsymbol{\Omega})$ where

$$\mathbf{V}_T(\xi) = T^{-1/2} \sum_{t=1}^{[T\xi]} (\mathbf{v}_t - E\mathbf{v}_t) \quad 0 < \xi \leq 1 \quad (2.3)$$

and $\mathbf{B}(\boldsymbol{\Omega})$ is vector Brownian motion with covariance matrix $\boldsymbol{\Omega} = \lim_{T \rightarrow \infty} E(\mathbf{V}_T \mathbf{V}_T')$.

As is well known, \mathbf{x}_t is covariance stationary for $-\frac{1}{2} < d < \frac{1}{2}$, whereas for $\frac{1}{2} < d < \frac{3}{2}$,

$$\mathbf{X}_T(\xi) = T^{1/2-d} \mathbf{x}_{[T\xi]} \xrightarrow{d} \mathbf{X}(\xi) \quad (2.4)$$

where \mathbf{X} is a Gaussian process of the type called fractional Brownian motion, with parameter d . Note that in view of our definition of \mathbf{x}_t , this is Brownian motion of Type 1, in the terminology of Marinucci and Robinson (1999). When $d = 1$, \mathbf{X} corresponds to the usual case whose special property is independent increments. The increments of \mathbf{X} are negatively correlated for $\frac{1}{2} < d < 1$ and positively correlated for $1 < d < \frac{3}{2}$.

We define fractional cointegration as the property that there exists $b > 0$ such that $\boldsymbol{\beta}' \mathbf{x}_t \sim I(d - b)$ for some matrix $\boldsymbol{\beta}$ ($m \times s$) of rank $s \geq 1$. A linear dynamic modelling framework embodying this property is the fractional vector error correction model (FVECM), proposed by Granger (1986). This might be cast in the form⁴

$$[\mathbf{B}(L) + \boldsymbol{\alpha} \boldsymbol{\beta}' (\Delta^{-b} - 1)] \Delta^d \mathbf{x}_t = \mathbf{w}_t \quad (2.5)$$

where $\boldsymbol{\alpha}$ ($m \times s$) has rank s , $\mathbf{B}(L)$ ($m \times m$) is a p -order polynomial matrix having its roots outside the unit circle, and \mathbf{w}_t will be assumed i.i.d. $(\mathbf{0}, \boldsymbol{\Sigma})$; see Davidson (2002), Section 2 for details. Note that (2.5) contains the usual VECM as the special case having $d = b = 1$. It has the conventional ECM feature that $\boldsymbol{\beta}' \mathbf{x}_t$ enters with a lag, noting that the zero-order term of Δ^{-b} is unity.

Let $\mathbf{x}_t = (y_t, \mathbf{z}_t')$, where y_t is a scalar, be a (possibly arbitrary) partition of the data, with a corresponding partition of the limit process, $\mathbf{X}' = (Y, \mathbf{Z}')$. Write

$$y_t = \boldsymbol{\gamma}' \mathbf{z}_t + u_t \quad (2.6)$$

so that the cointegration hypothesis implies the existence of a vector $\boldsymbol{\gamma}$ ($(m - 1) \times 1$) such that $u_t \sim I(d - b)$. In the following sections, we consider tests of the hypothesis $b = 0$ against $b > 0$ (noncointegration) and also of the hypothesis $b = d$ against $b < d$. The latter is stronger than the hypothesis of cointegration as defined, and might be called 'full cointegration' but note that in principle, one might test a hypothesis $b = b^*$ (say) by applying a test to the semi-differenced data $\Delta^{b^*} \mathbf{x}_t$; also see Davidson (2003a) on this question.

The standard tests for the popular case $d = 1$, whether based on regression or reduced-rank VAR estimation, depend critically on the convergence of the relevant trending variables to regular Brownian motion, after normalization. The main problem with adapting these tests is that when d is unknown the statistics are no longer asymptotically pivotal, depending on d as well as covariance parameters. As shown in Davidson (2002) and also in Appendix B, the main components of the asymptotic distributions are functionals of \mathbf{X} . Hence, most such tests are

³The symbols \rightarrow_d and \xrightarrow{d} for "convergence in distribution" are used to denote weak convergence in the relevant space of measures, given by the context.

⁴For simplicity, we assume that the data have zero mean and are free of deterministic trends. Intercepts and trends may be added to the following equations, but apart from requiring some well-known modifications to the asymptotic analysis (replacing limiting stochastic processes with their demeaned and detrended counterparts) add nothing material to the analysis.

not asymptotically pivotal. The bootstrap method is to generate the distributions of suitable test statistics under the null hypothesis by simulation. When asymptotically pivotal statistics are unavailable, one can try to make a virtue of this limitation and consider alternative statistics, not normally adopted but potentially offering good power properties. However, along with the choice of statistic, three other key issues in test design must be considered. First, how to mimic as accurately as possible the distribution of the test variables under the null hypothesis; second, how to keep the computing burden to a reasonable level, appropriate to exploratory and routine testing procedures; and third, how to minimise the distortion caused by substituting estimated for actual parameters. We consider some case-specific solutions to the first two problems, and also investigate some previously proposed techniques for bootstrapping enhancement, to deal with the last one.

2.2 The Conditional Bootstrap

An important feature of the tests proposed is that the variables z_1, \dots, z_T are held fixed at sample values, in the Monte Carlo replications. We shall, accordingly, consider the distributions of the statistics conditional on these values. Some care is needed to distinguish the conditioning concepts in use here, particularly in view of the fact that bootstrap distributions are, in any case, conditional on the sample data, by construction. The following brief résumé of the issues focuses for clarity on the regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, where \mathbf{X} is a random matrix and $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}_T)$.

Letting $\hat{\boldsymbol{\beta}}$ denote the OLS estimator, we can distinguish the conditional distribution of $\hat{\boldsymbol{\beta}}|\mathbf{X}$ (Gaussian with variance $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$) from the unconditional distribution of $\hat{\boldsymbol{\beta}}$ (mixed Gaussian with variance $\sigma^2 E[(\mathbf{X}'\mathbf{X})^{-1}]$). While we are generally more interested in the pivotal t and F statistics in this context, where the distinction matters (in constructing a confidence interval for $\boldsymbol{\beta}$, for example) it is often argued that the conditional distribution is of greatest interest, since \mathbf{X} is what we actually observe. In other words, we should use $s^2(\mathbf{X}'\mathbf{X})^{-1}$ to estimate the variance *even* if we knew $E[(\mathbf{X}'\mathbf{X})^{-1}]$. In the bootstrap context, the same type of argument suggests resampling \mathbf{u} (or in practice, $\hat{\mathbf{u}}$) keeping \mathbf{X} fixed at the sample values in each replication (conditional resampling) rather than resampling from the joint distribution of (\mathbf{X}, \mathbf{u}) .

If \mathbf{X} is stationary, such that $\text{plim } T^{-1}[\mathbf{X}'\mathbf{X} - E(\mathbf{X}'\mathbf{X})] = \mathbf{0}$, the conditional and unconditional distributions actually coincide in the limit. When $\mathbf{X} = I(d)$ for $\frac{1}{2} < d < \frac{3}{2}$, on the other hand, $T^{-2d}\mathbf{X}'\mathbf{X}$ converges weakly, and in this case the limiting distributions differ. The customary use of pivotal statistics relieves us of the need to choose between the conditional and unconditional distributions, but applying the bootstrap to non-pivotal statistics makes this choice explicit. The decision to resample from the conditional distribution is motivated by the fact that this technique uses additional sample information (the realization of \mathbf{X}) that would otherwise be discarded. Other virtues include reduced computational cost, and eliminating the risk of estimation and/or specification errors in modelling the distribution of \mathbf{X} .

The assumption that the distribution of $\mathbf{u}|\mathbf{X}$ does not depend on \mathbf{X} (i.e. the whole sample, not just current values) is necessary for valid random resampling. In a time series model with dynamics, this is in general only true when the columns of \mathbf{X} contains lags and/or leads, as well as contemporaneous values, of the underlying conditioning variables. Such an augmented conditioning set is used, for example, to derive efficient mixed-Gaussian estimators of cointegrating vectors; see Saikkonen (1991). We adopt a similar approach to allow conditional resampling in our bootstrap applications.

2.3 Bootstrap Tests of Noncointegration

Freed from the constraint of requiring our statistics to be asymptotically pivotal, two contrasting tests for the non-cointegration hypothesis are considered. The conventional approach of examining the time profile of the residuals for evidence of stationarity suggests the use of the Dickey-Fuller (1979) type of statistic. However, these statistics have some unexpected properties under fractional integration. It is shown in Davidson (2002) that if

$$\hat{\phi} = \frac{\sum_{t=1}^T \hat{u}_{t-1} \Delta \hat{u}_t}{\sum_{t=1}^T \hat{u}_{t-1}^2} \quad (2.7)$$

where $\hat{u}_t = y_t - \hat{\gamma}' z_t$ is the least squares residual from (2.6), then under H_0 , $|\hat{\phi}| = O_p(T^{\max\{-1, 1-2d\}})$. Since $T\hat{\phi}$ is the simple (unaugmented) Dickey-Fuller statistic of the first type, this implies that with $d < 1$ the usual statistic is diverging even when the null is true. Even more surprising is the fact that, except in the case $d = 1$,

$$t = \frac{\hat{\phi}}{s_e} \sqrt{\sum_{t=1}^T \hat{u}_{t-1}^2} \quad (2.8)$$

(the Dickey-Fuller t statistic) is also diverging under H_0 . Note from Lemma B.1(ii) that $|t| = O_p(T^{\max\{d-1, 1-d\}})$. Moreover, the adjustments for autocorrelation in $\Delta \hat{u}_t$ made in the augmented Dickey-Fuller or Phillips-Perron (1988) statistics are inappropriate when this process is long memory.

However, since it can be renormalized, the rate of divergence of $\hat{\phi}$ under H_0 is irrelevant so long as it is greater under the alternative. The bootstrap tabulation will in any case relate to a fixed sample size. We consider here the use of the Durbin-Watson statistic, asymptotically equivalent to $-2\hat{\phi}$. This statistic converges to 0 at the rates indicated under H_0 and is at most $O_p(1)$ under the alternative. It is shown in Davidson (2002) that the DW statistic yields a consistent test of the hypothesis $b = 0$ against $b > 0$, provided the residuals are mean-reverting ($d - b < 1$).

Tests based on $\hat{\phi}$, following the usual approach of modelling the time profile of the residuals, ignore other information in the sample such as ‘goodness of fit’, or the relative magnitudes of regressand and residual. However, in the bootstrap context, this second approach is straightforwardly implemented using the usual F statistic of the regression, following the possible removal of a deterministic trend. Letting

$$F = T \frac{\sum_{t=1}^T (\hat{\gamma}' z_t)^2}{\sum_{t=1}^T \hat{u}_t^2} \quad (2.9)$$

where $\hat{\gamma}$ is the OLS estimator from (2.6), F diverges even under $H_0 : b = 0$, which is the well-known ‘spurious regression’ phenomenon. However, it is shown in Davidson (2002) that when H_0 is true,

$$\frac{F}{T} \xrightarrow{d} \frac{\left(\Psi_{22}^{-1/2} \Psi_{21} + \sqrt{\psi_{11.2}} \mathbf{G}_{22}^{-1} \mathbf{G}_{21} \right)' \mathbf{G}_{22} \left(\Psi_{22}^{-1/2} \Psi_{21} + \sqrt{\psi_{11.2}} \mathbf{G}_{22}^{-1} \mathbf{G}_{21} \right)}{\psi_{11.2} (g_{11} - \mathbf{G}_{12} \mathbf{G}_{22}^{-1} \mathbf{G}_{21})}. \quad (2.10)$$

Here, $\mathbf{G} = \int_0^1 \mathbf{W}_d \mathbf{W}_d' dr$ where \mathbf{W}_d is a m -vector of independent standard fractional Brownian motions, with subscripts denoting the obvious partition with respect to $\mathbf{x}_t = (y_t, z_t)'$, $\Psi = \lim_{T \rightarrow \infty} T^{1-2d} E \mathbf{x}_T \mathbf{x}_T'$ and $\psi_{11.2} = \psi_{11} - \Psi_{12} \Psi_{22}^{-1} \Psi_{21}$. The matrix Ψ has full rank under H_0 , but under the alternative, $\psi_{11.2} = 0$. The test based on F/T is shown in Davidson (2002) to be consistent for all $\frac{1}{2} < d < \frac{3}{2}$ and $0 < b < d$, even in the case $1 \leq d - b < d$ (e.g., with unit root residuals). Of course, F/T is not asymptotically pivotal. However, provided its distribution can

be simulated in a bootstrap exercise, it is a candidate for an alternative test of non-cointegration. This is shown in Davidson (2002), in an empirical example, to offer somewhat more power than the DW-based test.

To bootstrap these statistics, it is necessary to simulate the data generation process under H_0 . When $b = 0$, the second term in (2.5) vanishes and the model reduces to the regular VAR process in the differences,

$$\mathbf{B}(L)\Delta^d \mathbf{x}_t = \mathbf{w}_t. \quad (2.11)$$

In principle, this system might be suitably estimated, and used to construct a bootstrap draw of the full vector of variables. As noted in Section 2.1, we hold the \mathbf{z}_t variables fixed at sample values, and resample y_t alone. For the method to work successfully, however, it is important to specify the conditional model correctly. To fix ideas, consider first how to generate a sample $\{\Delta^d y_t\}_1^T$ direct from the final form of (2.11), conditional on $\{\Delta^d \mathbf{z}_t\}_1^T$ and pre-sample values. The final form equations are

$$\Delta^d \mathbf{x}_t = \frac{1}{|\mathbf{B}(L)|} \text{adj } \mathbf{B}(L) \mathbf{w}_t. \quad (2.12)$$

For simplicity, suppose $m = 2$. Then, the final form equation for the (scalar) $\Delta^d z_t$ is

$$\begin{aligned} \Delta^d z_t = & \lambda_1 \Delta^d z_{t-1} + \dots + \lambda_q \Delta^d z_{t-q} + \kappa_{21,0} w_{1t} + \kappa_{22,0} w_{2t} \\ & + \kappa_{21,1} w_{1,t-1} + \kappa_{22,1} w_{2,t-1} + \dots + \kappa_{21,p} w_{1,t-p} + \kappa_{22,p} w_{2,t-p}. \end{aligned} \quad (2.13)$$

where $1 - \lambda_1 L - \dots - \lambda_q L^q = |\mathbf{B}(L)|$ with $q \leq 2p$, and the coefficients $\kappa_{2i,j}$ are derived in the obvious manner from the second row of $\text{adj } \mathbf{B}(L)$. Assuming the pair (w_{1t}, w_{2t}) are jointly Gaussian, write

$$w_{1t} = \delta w_{2t} + e_t \quad (2.14)$$

where $\delta = \sigma_{12}/\sigma_{22}$ and e_t is independent of $\Delta^d z_t$. Since $\{\Delta^d z_t\}_{-\infty}^T$ and $\{w_{1t}, w_{2t}\}_{-\infty}^0$ are to be treated as fixed, both sides of (2.13) have to be the same in every replication. The generation process therefore proceeds, for $t = 1, 2, \dots, T$, by taking a random drawing e_t , substituting this into (2.14), and then w_{1t} into (2.13), and solving the resulting equation for w_{2t} . The sequence $\{w_{1t}, w_{2t}\}_1^T$ so generated, which preserves the correlation structure of the original data, can then be used to generate $\Delta^d y_t$ using the counterpart of (2.13). In the case $\kappa_{21,0} = 0$, it would be necessary to substitute for $w_{1,t-1}$, and so on. In the case $m > 2$, there will be $m - 1$ recursive equations to solve for $\mathbf{w}_{2t} = (w_{2t}, \dots, w_{mt})$, after setting (with the obvious partition of Σ)

$$w_{1t} = \mathbf{w}'_{2t} \Sigma_{22}^{-1} \sigma_{21} + e_t.$$

However, this approach still has the disadvantage of requiring the full VAR to be explicitly specified, estimated and inverted. Now consider the more direct approach of employing a modified reduced form dynamic equation. First, reduce (2.11) to a single equation by taking a linear combination normalized on $\Delta^d y_t$. Partition $\mathbf{B}(L)$ by its first column as $[\mathbf{b}_1(L) : \mathbf{B}_2(L)]$, and let \mathbf{d} ($m \times 1$) denote the normalizing vector having the property $\mathbf{d}' \mathbf{b}_1(0) = 1$. A candidate regression is then

$$\Delta^d y_t = [1 - \mathbf{d}' \mathbf{b}_1(L)] \Delta^d y_t + [\boldsymbol{\psi}'_0 - \mathbf{d}' \mathbf{B}_2(L)] \Delta^d \mathbf{z}_t + w_t^* \quad (2.15)$$

where $w_t^* = \mathbf{d}' \mathbf{w}_t - \boldsymbol{\psi}'_0 \Delta^d \mathbf{z}_t$, and $\boldsymbol{\psi}_0$ is the least squares bias term reflecting possible endogeneity of $\Delta^d \mathbf{z}_t$ in the system.

In principle, an unrestricted equation of this form captures the reduced form of the dynamics represented by (2.11). However, this is *not* an appropriate equation for constructing the bootstrap draws, because w_t is in general correlated with $\Delta^d \mathbf{z}_{t+j}$ for all $j > 0$, reflecting Granger-causation

of $\Delta^d \mathbf{z}_t$ by $\Delta^d y_t$. To ensure valid conditioning on $\Delta^d \mathbf{z}_{t-j}$ for all j , including negative values, (2.15) must be further augmented. Define $\boldsymbol{\psi}(L^{-1}) = \boldsymbol{\psi}_0 + \boldsymbol{\psi}_1 L^{-1} + \dots + \boldsymbol{\psi}_r L^{-r}$, such that

$$\varepsilon_t = \mathbf{d}' \mathbf{w}_t - \boldsymbol{\psi}'(L^{-1}) \Delta^d \mathbf{z}_t \quad (2.16)$$

is orthogonal to $\Delta^d \mathbf{z}_{t+j}$ for all integers j . In general this holds only for $r = \infty$, but in practice an adequate approximation can be found with finite r depending on sample size. Modifying (2.15) to include these terms, and simplifying notation, the bootstrap distribution is modelled by running a regression of the form

$$\Delta^d y_t = a_1(L) \Delta^d y_{t-1} + \mathbf{a}_2'(L) \Delta^d \mathbf{z}_t + \varepsilon_t \quad (2.17)$$

where $a_1(L)L = [1 - \mathbf{d}' \mathbf{b}_1(L)]$ and $\mathbf{a}_2(L) = [\boldsymbol{\psi}'(L^{-1}) - \mathbf{d}' \mathbf{B}_2(L)]$.

The bootstrap draw is then a matter of either making drawings from the appropriate distribution, or simply randomly resampling the residuals $\hat{\varepsilon}_t$ with replacement, to yield a bootstrap sample $\varepsilon_1^*, \dots, \varepsilon_T^*$. From this series, (2.17) is first solved recursively, to reconstruct the I(0) dynamics. Following re-integration by application of the filter Δ^{-d} , the resulting bootstrap sample y_1^*, \dots, y_T^* has a distribution matching that of the original sample y_1, \dots, y_T , conditional on $\mathbf{z}_1, \dots, \mathbf{z}_T$, when H_0 is true. Pre-sample values in the computation of $\Delta^d(\cdot)$ and $\Delta^{-d}(\cdot)$ must of course be substituted, typically by zeros. These effects are of small order but inevitably entail some distortion of the distribution in finite samples, although one that matches the usual treatment of the observed data in estimation.

Next, consider the properties of these tests. Defining

$$\mathcal{X}_{-\infty}^t = \sigma(\dots, \Delta^d y_{-1}, \Delta^d y_0, \dots, \Delta^d y_t, \dots, \Delta^d \mathbf{z}_{-1}, \Delta^d \mathbf{z}_0, \dots, \Delta^d \mathbf{z}_t)$$

and $\mathcal{Z}_t^\infty = \sigma(\Delta^d \mathbf{z}_t, \Delta^d \mathbf{z}_{t+1}, \dots)$ consider the conditional joint density

$$f(\Delta^d y_1, \dots, \Delta^d y_T | \mathcal{X}_{-\infty}^0, \mathcal{Z}_1^\infty) = \prod_{t=1}^T f(\Delta^d y_t | \mathcal{X}_{-\infty}^{t-1}, \mathcal{Z}_t^\infty). \quad (2.18)$$

The statistics F/T and $\hat{\phi}$ are measurable functionals of the processes $\{y_1, \dots, y_T\}$ and $\{\mathbf{z}_1, \dots, \mathbf{z}_T\}$ and, equivalently, of the infinite stationary processes $\{\dots, \Delta^d y_{-1}, \Delta^d y_0, \Delta^d y_1, \dots, \Delta^d y_T\}$ and $\{\dots, \Delta^d \mathbf{z}_{-1}, \Delta^d \mathbf{z}_0, \Delta^d \mathbf{z}_1, \dots, \Delta^d \mathbf{z}_T\}$. An exact conditional Monte Carlo test yielding true p -values, conducted holding $\{\mathbf{z}_1, \dots, \mathbf{z}_T\}$ fixed at sample values in the replications, would require us to compute F/T and $\hat{\phi}$ from drawings from (2.18) subject to the restrictions of the null hypothesis. Be careful to note that, since the statistics are not pivotal, resampling from (2.18) yields different tests to resampling from $f(\Delta^d y_1, \dots, \Delta^d y_T, \Delta^d \mathbf{z}_1, \dots, \Delta^d \mathbf{z}_T | \mathcal{X}_{-\infty}^0)$, and this is true even asymptotically. The limiting bootstrap distributions depend on fixed \mathbf{Z} (whose role may be likened to a vector of deterministic trends) but, as shown in Davidson (2002), are nevertheless functionals of the fractional Brownian motion Y depending on d and the covariance parameters. The asymptotic validity of the tests can be established formally, as follows (see Appendix A for the proof).

Proposition 2.1 *Assume the generation process of the data is (2.11) such that the stated assumptions on \mathbf{w}_t hold and the null hypothesis of noncointegration is true. Assume that $d \in (\frac{1}{2}, \frac{3}{2})$ and that \mathbf{B}_0 and $\boldsymbol{\Sigma}$ are regular points of the parameter space, such that matrices in an open neighbourhood of these points have constant rank. Let the bootstrap distributions of $\hat{\phi}$ and F/T , conditional on $\mathbf{z}_1, \dots, \mathbf{z}_T$ and pre- and post-sample values, be generated from (2.17) estimated by OLS, where $r = o(T^{1/3})$. Then, the bootstrap tests are correctly sized in the limit as $T \rightarrow \infty$.*

While this result leaves some scope in the choice of r , and assumes p known, in view of the fact that (2.17) is intended to represent the null hypothesis it is very important that data-based model selection criteria *not* be used in these choices. These are inappropriate when the null is false, and over-parameterizing will effectively kill the power of the tests. Some judgement is needed here, to strike a balance between correct test size and sufficient power, in a given sample.

Proposition 2.1 does not establish the rate of ERP convergence. However, given $T^{1/2}$ -consistency of the parameter estimates, the essential requirement for the same convergence rate transferring to the ERPs themselves is that the c.d.f.s are sufficiently smooth functions of the parameters. For the present cases, it is at worst a reasonable conjecture that these mild conditions are satisfied. In particular, there is no counterpart here of the well-known problem of the unstable autoregressive model (see Basawa et al., 1991a,b). Notwithstanding that the unit-root model is embedded as a special case in the continuum of fractional processes, note that there is no discontinuity in the family of distributions at the point $d = 1$, corresponding to the unit autoregressive root – nor, for that matter, at any other points, such as the "stationarity boundary", $d = \frac{1}{2}$.

2.4 Bootstrap Tests of Cointegration

Shin (1994) (which reference we henceforth give as SH, for brevity) has derived a variant of the KPSS test of Kwiatkowski, Phillips, Schmidt and Shin (1992), based on regression residuals, to test the null hypothesis of cointegration in $I(1)$ data. In the case of non-strictly exogenous regressors, SH shows that to obtain an asymptotically pivotal statistic, the test residuals should be derived from the efficient cointegrating regression,

$$y_t = \gamma' z_t + \pi(L)\Delta z_t + e_t \quad (2.19)$$

where $\pi(L) = \pi_{-K}L^{-K} + \dots + \pi_K L^K$ and $K \rightarrow \infty$ with sample size, as derived by Saikkonen (1991), which reference will henceforth be denoted for brevity as SK. We rely in this section on the results of SK and SH which in turn are indebted to the analysis of Said and Dickey (1984).

The assumption underlying the analyses of SH and SK is that

$$y_t = \gamma' z_t + v_{1t} \quad (2.20a)$$

$$\Delta^d z_t = v_{2t} \quad (2.20b)$$

where $d = 1$, and $\mathbf{v}_t = (v_{1t}, \mathbf{v}'_{2t})'$ is a vector of stationary weakly dependent processes.⁵ Assuming the representation

$$v_{1t} = \sum_{j=-\infty}^{\infty} \pi'_j \mathbf{v}_{2,t+j} + e_t. \quad (2.21)$$

such that e_t is totally independent of $\{\mathbf{v}_{2t-j}, -\infty < j < \infty\}$, their analysis depends on the conditions $K = o(T^{1/3})$ and

$$\sum_{|j|>K} \|\pi_j\| = o(T^{-1/2}). \quad (2.22)$$

The limiting null distribution of SH's statistic in the fractional case of (2.20) is derived in Appendix B. Modified versions of the relevant arguments of SH and SK are given for the $I(d)$ case, letting the simple differences (Δ) be replaced by fractional differences (Δ^d), and subject to the modified condition $K = o(T^{\min\{1/3, (2d-1)/3\}})$. This remains compatible with (2.22) when \mathbf{v}_t is

⁵For simplicity we ignore here the possibility allowed in SK of additional regressors appearing only in difference form.

generated by (say) a stable ARMA process, such that $\|\boldsymbol{\pi}_j\| = O(\rho^{|j|})$ for $|\rho| < 1$. The statistic is shown to depend on Brownian functionals derived from the distribution of the regressors, implying that it is not asymptotically pivotal in the $I(d)$ case.

However, consider the case of model (2.5) with $s = 1$, subject to the null hypothesis $H_0 : b = d$, to be tested against alternative $b < d$. Assuming a single cointegrating vector normalised (possibly arbitrarily) on y_t , (2.5) can be arranged in the form considered by SK and SH by writing the final form equations as

$$\begin{aligned}\Delta^d \mathbf{x}_t &= [\mathbf{B}(L) + \boldsymbol{\alpha}\boldsymbol{\beta}'(\Delta^{-d} - 1)]^{-1} \mathbf{w}_t \\ &= [\Delta^d(\mathbf{B}(L) - \boldsymbol{\alpha}\boldsymbol{\beta}') + \boldsymbol{\alpha}\boldsymbol{\beta}']^{-1} \Delta^d \mathbf{w}_t \\ &= \mathbf{G}(L)' \Delta^d \mathbf{w}_t\end{aligned}$$

(defining $\mathbf{G}(L)$). Letting $\boldsymbol{\beta}' = (1, -\boldsymbol{\gamma}')$ ($1 \times m$), and defining the partition $\mathbf{G}(L) = [\mathbf{g}_1(L) : \mathbf{G}_2(L)]$, the system can be given the form (2.20) where

$$v_{1t} = [\boldsymbol{\beta}' - (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\mathbf{B}(L)]\mathbf{G}(L)'\Delta^d \mathbf{w}_t + (\boldsymbol{\alpha}'\boldsymbol{\alpha})^{-1}\boldsymbol{\alpha}'\mathbf{w}_t \quad (2.23a)$$

$$\mathbf{v}_{2t} = \mathbf{G}_2(L)'\Delta^d \mathbf{w}_t. \quad (2.23b)$$

According to these formulae, the stationary processes v_{1t} and \mathbf{v}_{2t} feature hyperbolic lag distributions in general, due to the presence of the anti-persistent components, although to derive these forms analytically for the general case appears intractable. Under these conditions, whether condition (2.22) holds for eligible rates of divergence of K is no longer self-evident, and becomes an issue for investigation. We return to this question below.

To derive a bootstrap version of the Shin test, the system may be reduced as in the non-cointegration case to a single dynamic equation, in which all variables but y_t are held conditionally fixed, but in this case embodying the hypothesized cointegration. Following the same argument that leads to equation (2.17) in the non-cointegration case, the counterpart for this case is obtained as follows. First, premultiply the system (2.5) by $c^{-1}\boldsymbol{\alpha}'$, where $c = \boldsymbol{\alpha}'\mathbf{b}_1(0)$. Defining $a_1(L)$ by $a_1(L)L = 1 - c^{-1}\boldsymbol{\alpha}'\mathbf{b}_1(L)$, and letting $\mathbf{a}_2(L) = \boldsymbol{\psi}'(L^{-1}) - c^{-1}\boldsymbol{\alpha}'\mathbf{B}_2(L)$ where $\boldsymbol{\psi}$ is defined as in (2.16), the proposed generating equation is

$$\Delta^d y_t = a_1(L)\Delta^d y_{t-1} + \mathbf{a}_2(L)'\Delta^d \mathbf{z}_t + \mu[(1 - \Delta^d)y_t - \boldsymbol{\gamma}'(1 - \Delta^d)\mathbf{z}_t] + \varepsilon_t \quad (2.24)$$

where

$$\varepsilon_t = c^{-1}\boldsymbol{\alpha}'\mathbf{w}_t - \boldsymbol{\psi}'(L^{-1})\Delta^d \mathbf{z}_t \quad (2.25)$$

and since $\boldsymbol{\beta} = (1, -\boldsymbol{\gamma}')$, $\mu = c^{-1}\boldsymbol{\alpha}'\boldsymbol{\alpha}$, and $\mu < 0$ is a stability condition that the system must observe under H_0 , so that all the terms of (2.24) are $I(0)$. After substituting for y_t from $y_t = \boldsymbol{\gamma}'\mathbf{z}_t + u_t$, (2.24) can be rearranged in the form

$$u_t = a_1(L)\Delta^d u_{t-1} + (1 + \mu)(1 - \Delta^d)u_t + [\mathbf{a}_2(L) + \boldsymbol{\gamma}(a_1(L)L - 1)]'\Delta^d \mathbf{z}_t + \varepsilon_t. \quad (2.26)$$

After fitting (2.24) by least squares, solving out the implied estimate of the cointegrating vector $\hat{\boldsymbol{\gamma}}$, and rearranging as (2.26), the bootstrap draws can now be taken as follows. First, given randomly resampled residuals $\varepsilon_1^*, \dots, \varepsilon_T^*$, (2.26) is solved recursively to generate u_1^*, \dots, u_T^* , given fixed $\Delta^d \mathbf{z}_1, \dots, \Delta^d \mathbf{z}_T$. Then, y_1^*, \dots, y_T^* are obtained from $y_t^* = \hat{\boldsymbol{\gamma}}'\mathbf{z}_t + u_t^*$, and the KPSS statistic computed from these data.

Let $\|\mathbf{x}\|$ represent the Euclidean norm of a vector \mathbf{x} .

Proposition 2.2 *Suppose that in model (2.5) the null hypothesis $b = d$, $s = 1$ is true, and that \mathbf{B}_0 and $\boldsymbol{\Sigma}$ are regular points of the parameter space. Let the bootstrap distribution of Shin's statistic,*

conditional on $\mathbf{z}_1, \dots, \mathbf{z}_T$ and pre- and post-sample values, be generated from (2.24) estimated by OLS. Subject to the conditions $r = o(T^{1/3})$ and

$$\sum_{j=r}^{\infty} \|\boldsymbol{\psi}_j\| = o(T^{-1/2}) \quad (2.27)$$

the bootstrap tests are correctly sized in the limit as $T \rightarrow \infty$.

(For the proof, see Appendix C.) Note that (2.27) has to be asserted as an additional assumption in this case, because it does not follow from the form of the assumed model, as is the case in Proposition 2.1. It is likely that if (2.22) holds then (2.27) will also hold, but note that (2.22) is not necessary for the validity of Proposition 2.2. Its failure would typically mean that the asymptotic distribution of the statistic is different, but it will still be reproduced by the bootstrap draws subject to the stated conditions.

However, the resampling procedure described thus far does not yield a consistent test, for reasons analogous to that arising over the selection of lags in the tests of non-cointegration. When H_0 is false, the unconstrained estimate of μ in (2.24) is converging to 0. To impose the restrictions of the null hypothesis in the bootstrap draws, μ must be constrained to be strictly negative in (2.26), regardless of the estimated value. When the null hypothesis is true, an appropriately chosen restriction should bind only with probability zero when the sample is large enough, and in that case the true null distribution is generated asymptotically by the bootstrap procedure. On the other hand, it will bind with probability 1 in the limit under the alternative hypothesis, ensuring that the test is consistent. In finite samples it may bind with positive probability under H_0 , and the value chosen once again represents a trade-off between size distortion and power. This is an issue that can best be investigated by simulation experiments.

A more direct approach to this problem is to estimate the cointegrating residual \hat{u}_t by a direct method such as Saikkonen (1991), and then model this under H_0 as a function of differences only (lagged, in the case of the regressand) to obtain the shocks for resampling. That is, run the regression

$$\hat{u}_t = a_1(L)\Delta^d y_{t-1} + \mathbf{a}_2'(L)\Delta^d \mathbf{z}_t + \varepsilon_t \quad (2.28)$$

where, similarly to above, a_1 is a lag polynomial and \mathbf{a}_2 is a lag-lead polynomial. This approach makes no attempt to represent the underlying FVECM, but under the null hypothesis, $u_t \sim I(0)$ and so should have a representation of this form, provided the lag/lead length is sufficient to make ε_t serially uncorrelated and orthogonal to $\Delta^d \mathbf{z}_1, \dots, \Delta^d \mathbf{z}_T$. The bootstrap draw is made as before by substituting out $\Delta^d y_{t-1}$, so the equation to be solved recursively from a resampled set $\varepsilon_1^*, \dots, \varepsilon_T^*$, conditional on $\Delta^d \mathbf{z}_1, \dots, \Delta^d \mathbf{z}_T$, is

$$\hat{u}_t^* = \hat{a}_1(L)\Delta^d \hat{u}_{t-1}^* + [\hat{\mathbf{a}}_2(L) + \hat{\boldsymbol{\gamma}}\hat{a}_1(L)L]'\Delta^d \mathbf{z}_t + \varepsilon_t^* \quad t = 1, \dots, T. \quad (2.29)$$

With finite order lags, this equation cannot yield a nonstationary solution at any values of the parameters, and also note that the form of the solution is the same as (2.26) with the constraint $\mu = -1$ imposed. However, this procedure is not assured to precisely reproduce the null distribution of the statistic. Like the ‘constrained μ ’ test, it needs to be evaluated by simulation experiments.

3 Implementation of the Bootstrap Tests

3.1 The Parametric Bootstrap

This section briefly reviews the methodology of parametric bootstrap tests, and considers some possible refinements. If F_{T0} , the c.d.f. under H_0 of a test statistic t_T in a sample of size T is

known (the whole function, not just particular quantiles), then the p -value

$$g_T = P_{T0}(t > t_T) = 1 - F_{T0}(t_T) \quad (3.1)$$

itself serves as a convenient test statistic. In a test of level (size) α the hypothesis is rejected if $g_T < \alpha$, equivalent to the criterion ‘reject if $t_T > c$ ’ where $P_{T0}(t > c) = \alpha$. A useful property of c.d.f.s is the following.⁶

Theorem 3.1 *Let $F(\cdot)$ be a c.d.f., let the r.v. X be distributed according to F , and let $\phi(y) = \inf\{x : y \leq F(x)\}$. Then*

$$P(F(X) \leq y) = \begin{cases} y, & y = F(\phi(y)) \\ F(\phi(y)-), & F(\phi(y)-) \leq y < F(\phi(y)) \end{cases} \quad (3.2)$$

for $0 \leq y \leq 1$.

Thus,

$$P_{T0}(g_T \leq \alpha) = \alpha \quad (3.3)$$

for every α that is the image of a continuity point of F_{T0} . If t_T is continuously distributed then g_T is uniformly distributed on the unit interval ($U(0, 1)$).

If the DGP under H_0 were known, g_T could be computed directly by Monte Carlo. Let

$$\hat{F}_{T0}(x) = \frac{1}{N} \sum_{i=1}^N I(t_{ni}^* \leq x) \quad (3.4)$$

represent the empirical distribution function (EDF), where $I(\cdot)$ is the indicator function of its argument, and t_{ni}^* is the i th drawing from the resampled null distribution of the test statistic. As a consequence of the Glivenko-Cantelli theorem, \hat{F}_T can be set arbitrarily close to F_T (uniformly, with probability 1) by taking N large enough, making the test effectively exact, although in practice a modest number of drawings may give sufficient accuracy. In what follows, formulae will generally be given in terms of the true c.d.f. F_T , with the understanding that this can be estimated in practice by the EDF \hat{F}_T .

In the more usual case of a composite null, the distribution of t_T is $F_{T\theta}$, depending on parameters $\theta \in \Theta$, where Θ denotes the parameter space under the maintained hypothesis. The null hypothesis takes the form $H_0 : \bar{\theta} \in \Omega_0$ where $\bar{\theta}$ is the vector that generated the observed data set, and $\Omega_0 \subset \Theta$ represents the set of admissible values under H_0 . F_{T0} now represents $F_{T\theta}$ at the point $\theta = \bar{\theta}$. In the regular parametric bootstrap procedure, the unknown θ are replaced by sample-based estimates $\hat{\theta}_T$. A nominal α -level bootstrap test is performed by rejecting if $\hat{g}_T < \alpha$ where

$$\hat{g}_T = 1 - F_{T\hat{\theta}}(t_T). \quad (3.5)$$

This is equivalent to using the critical value c defined implicitly by

$$\alpha = P_{T\hat{\theta}}(t > c) = 1 - F_{T\hat{\theta}}(c). \quad (3.6)$$

By contrast, the true level of this test is α' defined by

$$\alpha' = P_{T0}(t > c) = 1 - F_{T0}(c) \quad (3.7)$$

⁶Note that F is arbitrary here, and there is no requirement that it be either continuous or everywhere strictly increasing. Distributions on the unit interval arising in the application of the double and fast-double bootstraps may have mass points at 0 and/or 1.

where c is equated in (3.7) and (3.6). Under consistency of $\hat{\theta}_T$, such that in particular

$$\sup_x |F_{T\hat{\theta}}(x) - F_{T0}(x)| \xrightarrow{pr} 0 \quad (3.8)$$

as $T \rightarrow \infty$, this test is asymptotically of true level α , but in finite samples α and α' will be different. Subject to regularity conditions, the parametric bootstrap has a ERP ($\alpha - \alpha'$ in this notation) of $O(T^{-1})$ or smaller when the test is asymptotically pivotal, whereas a non-asymptotically pivotal test has an ERP of only $O(T^{-1/2})$ or smaller under the same conditions (see Beran 1988). In this case, the bootstrap test is comparable with the usual asymptotic test. The regularity conditions in question require the existence of an asymptotic expansion of the distribution with differentiable first-order term, to permit a linearization argument.⁷ Test statistics appropriate to the hypothesis of (non-)cointegration do not have distributions expressible in closed form, so the rate of convergence of the ERP remains conjectural in these cases.

3.2 The Double Bootstrap

Additional refinement may be possible by the method of *prepivoting*, proposed by Beran (1988). If the ERP of the bootstrap test is $o(1)$, then \hat{g}_T in (3.5) is an asymptotically pivotal statistic, since its limiting null distribution is that of g_T in (3.1), free of nuisance parameters in view of (3.3). If

$$H_{T0}(x) = P_{T0}(\hat{g}_T \leq x) \quad (3.9)$$

were known then $H_{T0}(\hat{g}_T)$ would be an exact p -value, being $U(0, 1)$ -distributed if the distribution of \hat{g}_T is continuous. In practice, Monte Carlo can be used to generate samples from the ‘‘pseudo-true’’ DGP $\hat{\theta}$ and construct $H_{T\hat{\theta}}$, the corresponding distribution of \hat{g}_T . This yields the *double-bootstrap* p -value

$$\tilde{g}_T = H_{T\hat{\theta}}(\hat{g}_T). \quad (3.10)$$

An ERP due to estimating $H_{T0}(\hat{g}_T)$ by \tilde{g}_T remains, but is of smaller order than that of \hat{g}_T , under regularity conditions. Since \hat{g}_T itself is computed by Monte Carlo the number of replications needed to generate $H_{T\hat{\theta}}$ to comparable accuracy to $F_{T\hat{\theta}}$ is in principle of order N^2 . However, it can be much reduced in practice by the use of stopping rules, and by simulating only the relevant part of the distribution, say, the region $0 \leq \tilde{g}_T \leq 0.15$. See Nankervis (2001) for details of these methods.

In econometric applications the bootstrap procedure, like any conventional tests, is subject to the hazard of specification uncertainty. Here it is manifested in the need to specify the DGP under H_0 , which inevitably only approximates the ‘real’ distribution of the observed data. It is important not to overlook the fact that the double bootstrap will not compensate for specification errors. It can only correct the ERP due to the estimation error that *would* exist, had the observed data been drawn from a case of the experimental model. In practice, this correction could be inappropriate and could actually increase the ERP.

3.3 The Fast Double Bootstrap

Since the double bootstrap is still expensive to compute, Davidson and MacKinnon (2000) have proposed two procedures that may confer some of the benefits of the double bootstrap, while requiring only a single run of replications. The c.d.f. $F_{T\hat{\theta}}$ is a drawing from a distribution of probability measures, derived from the distribution of $\hat{\theta}_T$. Since it is the difference between $F_{T\hat{\theta}}$

⁷See Davidson and MacKinnon (1999) for an exposition.

and F_{T_0} that induces the bootstrap ERP, the size of the difference can be estimated by making repeated drawings from the same distribution, as follows. On each bootstrap draw, first compute the test statistic t_T^* as usual. Then, let the model be estimated from this bootstrap sample, and call this new parameter set $\check{\theta}_T$. A second bootstrap sample is drawn, using these parameters as the DGP, and this sample is used to compute the test statistic again, say t_T^{**} .

The distribution of the t_T^{**} so generated can be represented as

$$\begin{aligned}\check{F}_{T\hat{\theta}}(\cdot) &= E(F_{T\theta}(\cdot)|\hat{\theta}_T) \\ &= \int_{\Theta} F_{T\theta}(\cdot) dP(\theta|\hat{\theta}_T)\end{aligned}\tag{3.11}$$

where $P(\theta|\hat{\theta}_T)$ represents the distribution induced on Θ by the resampling procedure. Davidson and MacKinnon's (2000) first fast double bootstrap method (FDB1) is to estimate the p -value by

$$\begin{aligned}\check{g}_{T1} &= 1 - F_{T\hat{\theta}}(\check{F}_{T\hat{\theta}}^{-1}(1 - \hat{g}_T)) \\ &= \hat{g}_T - [F_{T\hat{\theta}}(\check{F}_{T\hat{\theta}}^{-1}(1 - \hat{g}_T)) - (1 - \hat{g}_T)] \\ &= \hat{g}_T + [F_{T\hat{\theta}}(t_T) - F_{T\hat{\theta}}(\check{F}_{T\hat{\theta}}^{-1}(F_{T\hat{\theta}}(t_T)))].\end{aligned}\tag{3.12}$$

Note that $\check{F}_{T\hat{\theta}}^{-1}(1 - \hat{g}_T)$ represents the quantile corresponding to $1 - \hat{g}_T$ under \check{F} . Davidson and MacKinnon's second fast double bootstrap method (FDB2) is to estimate the p -value by

$$\begin{aligned}\check{g}_{T2} &= 2\hat{g}_T - (1 - \check{F}_{T\hat{\theta}}(t_T)) \\ &= \hat{g}_T + [\check{F}_{T\hat{\theta}}(t_T) - F_{T\hat{\theta}}(t_T)].\end{aligned}\tag{3.13}$$

Negative estimates of the p -value, and also estimates exceeding 1, are possible using this latter formula, and the truncated distribution on $[0, 1]$ is therefore likely to have positive probability mass on the boundaries. The FDB1 method avoids these problems, and may therefore perform better in practice.

These methods can be viewed as alternative bias corrections. To understand how they are intended to work, first define

$$u_T = 1 - F_{T_0}(t_T) \sim U(0, 1)\tag{3.14}$$

where this random variable derives its properties from the fact that t_T is a drawing from F_{T_0} itself, by hypothesis. Write

$$\check{F}_{T_0}(\cdot) = \int_{\Theta} F_{T\theta}(\cdot) dP(\theta|\bar{\theta})\tag{3.15}$$

to represent the counterpart of (3.11) when the distribution $P(\theta|\bar{\theta})$ is generated by 'nature' (the processes generating the observed sample from the distribution with parameters $\bar{\theta}$) of which $P(\theta|\hat{\theta}_T)$ is the experimental counterpart. Let t_T be held conditionally fixed, while taking expectations with respect to $P(\theta|\bar{\theta})$. Then, combining (3.5) with (3.14),

$$E(\hat{g}_T|t_T) = u_T - [\check{F}_{T_0}(t_T) - F_{T_0}(t_T)].\tag{3.16}$$

When the sample is not too small, so that $\bar{\theta}$ can be approximated by consistent $\hat{\theta}_T$,

$$E(\hat{g}_T|t_T) \approx u_T - [\check{F}_{T\hat{\theta}}(t_T) - F_{T\hat{\theta}}(t_T)]$$

and hence, from (3.13),

$$\check{g}_{T2} \approx u_T + [\hat{g}_T - E(\hat{g}_T|t_T)].\tag{3.17}$$

In other words, for a conditionally fixed t_T , the error in measuring the true p -value u_T by \check{g}_{T2} has, up to an asymptotic approximation, zero conditional mean. Given the approximations involved, the implication of these facts is that the correction in (3.13) should make a difference in cases where the distribution of random c.d.f.s drawn from $P(\theta|\hat{\theta})$ is not centred close (in an appropriate sense) to F_{T0} itself. Similarly, note that by (3.16),

$$\begin{aligned} E[F_{T\hat{\theta}}(t_T) - F_{T\hat{\theta}}(\check{F}_{T0}^{-1}(F_{T0}(t_T)))|t_T] &= \check{F}_{T0}(t_T) - F_{T0}(t_T) \\ &= u_T - E(\hat{g}_T|t_T). \end{aligned} \quad (3.18)$$

Up to an asymptotic approximation, substituting $\check{F}_{T\hat{\theta}}^{-1}F_{T\hat{\theta}}$ for $\check{F}_{T0}^{-1}F_{T0}$, the correction embodied in \check{g}_{T1} is therefore a random variable whose expectation conditional on t_T is the correction in \check{g}_{T2} .

The validity of the conditioning on t_T in the above argument needs to be validated; specifically, the issue is whether $\check{F}_{T\hat{\theta}}(t_T)$ does converge to the same limit as $\check{F}_{T0}(t_T)$, since the sampling distributions potentially have some salient differences. In the former case, the resampled DGPs $\check{\theta}_T$ are independent of t_T by construction, which is not the case with $\hat{\theta}_T$ and t_T . The differences should vanish if t_T and $\hat{\theta}_T$ are distributed independently in the limit, as argued by Davidson and MacKinnon (2000). In other cases, it is possible that an asymptotic bias could be introduced by \check{g}_{T1} and \check{g}_{T2} that is not present with \hat{g}_T . The question to be resolved here is whether these methods offer a degree of bias reduction in finite samples sufficient to prefer them to the regular bootstrap, in cases where the computation of the double bootstrap is excessively burdensome.

4 Monte Carlo Experiments

4.1 Experimental Setup

This section of the paper is intended to throw some experimental light on the alternative merits of the methods surveyed above.⁸ Data are generated from normally distributed random numbers and the bivariate case of model (2.5). Even in this simple set-up, several dimensions of variation of the model exist, and experimentation shows that the characteristics of tests can vary significantly from one model to another. It is impossible to report a manageable set of results capturing all the relevant features. Rather, the aim is to work in a restricted set-up, and vary a few key features to point up potential problems and suggest further specific lines of investigation.

The sample size is held to 100 observations in all experiments. This is about as large as is computationally feasible in simulation experiments,⁹ and is also as small as is reasonable for asymptotic approximations to be useful, delineating a ‘worst case’ that larger samples may generally improve on. Comparisons across sample size are also tricky because of the complicating factor of choices of bandwidth; specifically, lead/lags in the simulation regressions and bandwidth in the nonparametric consistent variance estimator in the Shin statistic. These should depend on sample size but cannot be selected on the basis of the data, for the reasons previously noted. The main model parameters to be varied are the value of d , the contemporaneous correlation between the series increments, and the serial correlation in the increments.

All the experiments were performed with 100 bootstrap draws for each test. In the case of the double bootstrap, 100 draws were also specified to generate the distribution of \hat{g}_T , although using the stopping rules devised by Nankervis (2001), and making only the replications required

⁸All computations are coded in Ox 3.20. See Doornik (1999).

⁹The distributions charted in Figures 1-4 each take around 8-10 hours of CPU time to compute on 550 MHz Pentium III Xeon processors.

to tabulate the EDF over the interval $[0, 0.15]$, substantially reduced the total number of draws required. Without these, the Monte Carlo study would not have been feasible. 3000 Monte Carlo replications were used for the size evaluations, and 1000 replications for each case of the alternative, to estimate powers.

Four FVECM models are used to generate the test data. In Case II (independence across time and contemporaneously across series) w_{1t} and w_{2t} in the bivariate case of (2.5) are generated as a pair of independent $N(0, 1)$. In Case CI (contemporaneous correlation) w_{1t} and w_{2t} are generated as $w_{2t} = N(0, 1)$ and $w_{1t} = (w_{2t} + N(0, 1))/\sqrt{2}$, so that each process has the same variance. In both of these cases, $\mathbf{B}(L) = \mathbf{I}_2$. In cases IA and CA, on the other hand,

$$\mathbf{B}(L) = \begin{bmatrix} 1 - 0.3L & 0 \\ 0 & 1 - 0.3L \end{bmatrix}. \quad (4.1)$$

with the shocks generated as in II and CI, respectively. In the tests of noncointegration, the null hypothesis is represented by the case $b = 0$. In the Shin tests of cointegration the null hypothesis is $b = d$, and the ECM is specified by $\boldsymbol{\alpha} = (-.5, .5)'$ and $\boldsymbol{\beta} = (1, -1)'$. These settings are also used for power estimation in the noncointegration tests. In all cases, note that the structure of the model is symmetric in $y_t = x_{1t}$ and $z_t = x_{2t}$, and the endogeneity status of the variables is the same in each case.

4.2 Non-Cointegration Tests

Tests of non-cointegration based on the F and DW statistics are investigated, using equation (2.17) to generate the data conditionally on (z_1, \dots, z_T) with $p = r = 1$ in the II and CI cases (representing a degree of over-fitting, since the true lead/lag lengths are zero in these models) and $p = r = 2$ or 4 in the other cases – see Figures 1 and 2 for details. Figures 1 and 2 show the EDFs for the alternative p -values under H_0 , plotted over the interval $[0, 0.15]$. The three cases shown in each chart are the regular bootstrap (\square) the double bootstrap (\diamond) and the FDB1 method (\circ).¹⁰ If the test is exact, such that a test of level α rejects in $100\alpha\%$ of replications, these EDFs should lie on the diagonal of the squares. The ERPs for the tests are therefore easily assessed from the charts. Each figure gives the EDFs for three values of d and the four model alternatives, II, CI, IA, and CA.

Estimates of d for both variables are computed (in each replication) to generate the fractional differences. Note that the estimate for z_t is needed to compute the dynamic model coefficients, even though these data are not resampled. The estimation is done using univariate ARFIMA models and maximizing the Whittle likelihood. The ARFIMA(1, d ,0) was used in all cases except the II and CI models of non-cointegration, where the ARFIMA(0, d ,0) was used. Experiments with less efficient estimation procedures, such as Robinson's (1994) nonparametric estimator, did not yield such satisfactory results, although investigation of these issues is ongoing. Note that equal d s are not an implication of the null hypothesis and therefore, these parameters are estimated separately for each process.

Estimated powers of the regular bootstrap against selected alternatives are shown in the first panel of Table 2. The size-corrected powers have been computed using the reported EDFs from Figures 1 and 2. In other words, if \hat{g}_T is the regular bootstrap p -value then the corrected p -value is $H_{0T}(\hat{g}_T)$, where H_{0T} is the EDF of \hat{g}_T under H_0 , estimated by simulation. According to Theorem 1, these size-corrected tests are exact within the chosen numerical accuracy, by construction. Be careful to note the distinction between these corrected p -values and double bootstrap p -values; the

¹⁰The FDB2 p -values were also computed, but these results are omitted to save space. This method suffers from over-rejection at small significance levels for the reason cited, and is not otherwise notably superior to FDB1.

latter are computed by resampling from the artificial null hypothesis, and hence are susceptible to specification error, while the former are got by replicating the ‘true’ DGP, in the context of the Monte Carlo exercise.

4.3 Cointegration Tests

The two variants of the Shin test of cointegration are investigated, using the same statistic but different methods of drawing the bootstrap samples. The results are reported, as for the non-cointegration tests, in Figures 3 and 4 and the second panel of Table 2. The ‘ECM’ method uses equation (2.26) after fitting equation (2.24) with $p = 2$ and $r = 5 \approx T^{1/3}$. To impose the stationarity constraint under H_0 , the parameter μ is set to the smaller of its estimated value and -0.5 . The ‘I(0) Regression’ method uses equation (2.29) after fitting equation (2.28), again with $p = 2$ and $r = 5$. In this case, γ has to be estimated extraneously, and the Saikkonen estimator, as used to compute the test statistic itself, is used. Note that the bandwidth for the latter estimator is set to $K = 5$ for all cases. Because equal integration orders is a restriction of H_0 in this case, the value of d was estimated (as described previously) for z_t only. The same value is also used to difference y_t .¹¹

The value of r is large relative to the non-cointegration tests, because of the relatively slow decay of the ψ_j coefficients following from (2.23b). As noted in Section 2.4, the asymptotic validity of the bootstrap distributions depends on assumption (2.27) which it turns out may fail, especially for the smaller values of d . However, the condition can be investigated in specific cases, by numerical approximation.

We proceed as follows. Letting n denote the truncation value, the n leading coefficients of the MA(∞) form of (2.23b) are evaluated. We denote by $\theta_{21}(L)$ and $\theta_{22}(L)$ the polynomials applied to w_{1t} and w_{2t} , respectively, in the equation for v_{2t} . Define the sequence of pairs $\{(w_{1h}^j, w_{2h}^j), h = 0, 1, 2, \dots, n\}$ where

$$w_{kh}^j = \begin{cases} 1_{\{h=0\}}(h), & k = j \\ 0, & k \neq j \end{cases}$$

and $1_A(h)$ is the indicator function of the set A . The sequences $\{\theta_{2jh}, h = 0, 1, 2, \dots, n\}$, for $j = 1, 2$, can be found by solving (2.23b) recursively for (w_{1h}^j, w_{2h}^j) , for $j = 1, 2$ respectively, where n is chosen large enough that the tail can be ignored; in practice; $n = 2000$ and $n = 6000$ gave similar results. (If desired, $\theta_{11}(L)$ and $\theta_{12}(L)$ could be computed similarly.) The sequence $\omega_{22h} = E(v_{2t}v_{2t+h})$ is now computed for $h = 1, \dots, T \ll n$ from the approximation formula

$$\omega_{22h} \approx \sum_{k=1}^2 \sum_{l=1}^2 \sum_{s=0}^{n-h} \sigma_{kl} \theta_{2ks} \theta_{2l, s+h}$$

where $\sigma_{kl} = E(w_{kt}w_{lt})$. Similarly note that

$$\gamma_{2h} = E(v_{2, t+h} \boldsymbol{\alpha}' w_t) \approx \sum_{k=1}^2 \sum_{l=1}^2 \sigma_{kl} \alpha_k \theta_{2l, s+h}.$$

Apart from the factor c^{-1} , the coefficients ψ_j in (2.25) can be thought of as the limiting case as $T \rightarrow \infty$ of the vector

$$\boldsymbol{\psi}_T = \boldsymbol{\Omega}_{22}^{-1} \boldsymbol{\gamma}_2 \tag{4.2}$$

¹¹Note that in the general case of a multiple regression, the value to use for this purpose should be the largest d of the regressors.

	II	CI	IA	CA
$d = 0.7$	-1.9	-1.93	-1.9	-1.96
$d = 1$	-26	-43	-30	-16
$d = 1.3$	-2.45	-2.55	-2.6	-2.78

Table 1: Convergence Rates for ψ_j Coefficients

where $\gamma_2 = (\gamma_{20}, \dots, \gamma_{2T})'$ and $\{\Omega_{22}\}_{ij} = \omega_{22,|i-j|}$. Approximating the ψ_j by (4.2) with $T = 100$, the slope coefficients in the regressions of $\log|\psi_j|$ on $\log j$ are shown in Table 1. (The R^2 s are around 0.99 in each case.)

Defining δ by $\sum_{j=r}^{\infty} |\psi_j| = o(r^{-\delta})$, since we need $r = o(T^{1/3})$ the assumptions of Proposition 2.2 require $\delta > \frac{3}{2}$, which is equivalent to $|\psi_j| = o(j^{-\frac{5}{2}})$. This condition appears to be satisfied, or nearly so, for the case $d = 1.3$, and also for the case $d = 1$ where convergence is geometric, but not in the case $d = 0.7$. This indicates that with $d = 0.7$ the bootstrapped Shin test may not be correctly sized asymptotically, even without the constraint on μ . Of course, the actual magnitude of the distortion depends on the absolute magnitude of the tail coefficients as well as their rate of convergence. It is gratifying to note from Figures 3 and 4 that the ERP is not noticeably larger for the case $d = 0.7$ than for the other cases, particularly noting that this is a worst-case example, in which the elements of α have equal magnitude in the two equations of the DGP. Under the weak exogeneity condition $\alpha_2 = 0$, the coefficients can converge at worst geometrically. However, note that the size distortion is greatest, paradoxically, in the ‘II’ cases where there is no correlation or autocorrelation in the increments. This might be accounted for by the fact that in these cases the hyperbolic components of (2.23b) dominate, and their influence is not ‘diluted’ by the short-run dynamics.

4.4 Discussion

While simulation evidence is always specific to the cases investigated, these results are encouraging as far as they go, and point the direction for further investigations. In many cases, the regular bootstrap appears to provide a reasonably well-sized test. When the difference is noteworthy, the double bootstrap often appears to yield a better approximation, and also a result close to the fast double bootstrap, which is encouraging evidence in favour of the latter. We have noted that there are two sources of error inherent in any practical bootstrap procedure, as indeed in any econometric exercise; estimation error and specification error. Generating the data by a model different from the conditional regressions used in the simulations, and in particular, having the regressor potentially endogenous, allows some assessment of the error from both sources. The performance of the tests in the face of estimation error alone can be judged most clearly by considering case II (top rows of charts) in Figures 1 and 2. Since under H_0 the data are a pair of independent pure fractional series, the assumed null model is correct, even though over-parameterised in this case. The regular bootstrap tends to under-reject in these cases when $d = 0.7$, and the double and fast double procedures both provide notable improvements in the ERP. The same is true for case IA of the F procedure and case CI of the DW procedure, where the deviation of the true model from the fitted model should be irrelevant. Of course, the degree of improvement observed here is related to the fact that these statistics are not asymptotically pivotal, especially in respect of the covariance parameters, so that the double bootstrap is predicted to yield an order-of-magnitude reduction in ERP.

In other cases, much as might be expected to occur in an application, the ‘truth’ of the null model is dependent (at least) on the choice of lag lengths and ARFIMA specifications. Where there is a tendency to over-reject, the double bootstrap methods often exhibit worse

size distortion than the regular bootstrap, which is apparently benefitting from an off-setting of the two sources of bias. It would be generally dangerous to assume that the double or fast double bootstrap procedures will always yield size corrections, although this fact has nothing to do with the inherent virtues of the method. Since the F test is evidently more sensitive to mis-specification of the short-run cross-correlations, and the DW test more likely to over-reject in the presence of short-run autocorrelation (see the second and third rows of Figures 1 and 2, respectively) there are grounds to advocate using these two statistics in conjunction. If one were to reject but not the other, one should be correspondingly cautious in evaluating this evidence. The lead/lag lengths of 4 used in the CA cases reduced over-rejection that was apparent with length 2, but there may be a corresponding reduction in corrected power.

Interpreting the Shin tests of cointegration is subject to two differences from the noncointegration tests. First, in all of these cases the null hypothesis is being approximated, so some specification error is to be expected. On the other hand, the statistics are asymptotically pivotal in respect of the covariance parameters, although not with respect to d . This is likely to explain why the differences between the regular and double bootstraps is much smaller overall. Such biases as we observe can be attributed mainly to specification error. The $I(0)$ regression method (Figure 4) has, unsurprisingly, a general tendency to over-reject, and this method may not be worth pursuing. On the other hand, for the ECM method (Figure 3) the regular bootstrap shows minimal size distortion in most cases, although the power of these procedures is not too high. Further experimentation with the choice of μ constraint might be desirable.

5 Empirical Application

Lettau and Ludvigson (2001) report estimates of a cointegrating relationship between the logarithms of US consumption, labour income and household asset wealth. They investigate the role of the cointegrating residual from this relation in forecasting asset returns. The interesting feature of their data set (quarterly, 1952Q4-2001Q1, 194 observations)¹² is that the estimated values of d for the consumption and income series significantly exceed unity, implying that the growth rates of these variables exhibit long memory. Estimates of ARFIMA(p,d,q) models are shown in Table 3. These are computed by the Whittle frequency domain maximum likelihood method, and in each case, the model reported is ARFIMA(0, d ,0), which is the case that optimizes the Schwarz model selection criterion (SBC) out of the range of models satisfying $0 \leq p + q \leq 4$. There is a little evidence of neglected autocorrelation in the model of the consumption series according to the Box-Pierce statistic,¹³ and therefore the most parsimonious model satisfying this additional criterion is also shown. These two cases are referred to as Consumption Model 1 and Consumption Model 2.

Lettau and Ludvigson report tests of cointegration using the Phillips-Ouliaris (1990) and Johansen (1998, 1991) procedures, but of course, these tests impose the assumption that the data are $I(1)$, and use the customary first-order asymptotic approximations to critical values. The present methods provide an opportunity to repeat their tests under the milder assumption that the true d values are unknown, and may differ from unity. Note that the null hypothesis does not include the restriction that the d -values of the variables are equal even to each other, so

¹²The data are made available by the authors on the internet at <http://www.ny.frb.org/rmaghome/economist/ludvigson/ludvigson.html> and are used here with grateful acknowledgement.

¹³' $Q(12)$ ' denotes the number of terms in the sum of squared autocorrelations. The distribution of this statistic in the residuals from an ARFIMA process is unknown. This is being interpreted as approximately $\chi^2(12)$ in the ARFIMA(0, d ,0) case, but with due caution.

that rejecting it is a correspondingly strong finding. The wealth variable is permitted to have a smaller d than the other variables even under the alternative.

The results of the tests of noncointegration are given for both consumption models in Table 4. The cases for both 2 leads/lags and 3 leads/lags in the bootstrap model are quoted, noting that the F test results are equivocal in the second case, with non-rejection at the 5% level. However, the costs of over-parameterizing the model, in terms of lost power, can be considerable, and the time series evidence from Table 3 gives adequate grounds to consider the 2 lead/lag test reliable. In any case, the overall evidence from these tests appears to favour the alternative hypothesis.

Two variants of the Shin test of cointegration are reported in Table 5. Note that the null hypothesis entails equality of the largest d s in the model, and accordingly, that of consumption is measured by that of labour income. Implicitly, note that this comes close to adopting Consumption Model 2. Interestingly, the hypothesis that the cointegrating residuals are $I(0)$ is also rejected at the 5% level. If we estimate d for the cointegrating residual (for this purpose the efficiently estimated residual reported by Lettau and Ludvigson (2001) was used) we obtain, for an ARFIMA(0, d ,0), the results $\hat{d} = 0.86$ with standard error of 0.08, and a residual $Q(12) = 8.63$. In other words, the evidence indicates that the series cointegrate with $b > 0$ but also $b < d$, so that the cointegrating residual is long memory and nonstationary. Apparently, neither $I(1)$ nor $I(0)$ are relevant cases in this analysis.

6 Concluding Remarks

This paper has suggested some possible ways to test cointegration hypotheses using the bootstrap, in situations where the usual tests are either unavailable, or potentially misleading. There will be many occasions in applied research where the assumptions underlying the application of standard tests are questionable, and in these cases, our methods may often provide a more robust approach to the testing problem. Considering that they are not asymptotically pivotal, the tests perform reasonably well in a small sample. While there are size distortions in some situations, these are almost certainly no worse than are encountered in the conventional asymptotic tests. While the findings point to the benefits of implementing either the double bootstrap or fast double bootstrap in some situations, the simulation evidence suggests that test biases due to estimation error and specification error are offsetting, so that the double bootstrap cannot be treated as a panacea. Use of a variety of tests with contrasting characteristics, and investigating the effects of changing lag lengths, etc., are desirable precautions in any implementation.

Two possible extensions of the analysis, not considered explicitly in this paper, should be noted. The first is that, with a modified form of (2.5) as detailed in Davidson (2002), it is possible to have some elements of \mathbf{x}_t integrated to order less than d . The necessary condition for cointegration is that at least two elements share the same, maximal, order, one of which must of course correspond to y_t in (2.6). Note that the omission of such a variable, of integration order $I(d - b)$, with $b > 0$, would nonetheless represent a case of the alternative. However, if that variable reduces the integration order of the residuals to $I(d - b - b^*)$ for $b^* > 0$, tests of noncointegration would be expected to have greater power against the second alternative than the first.

The second extension is to test for cointegrating rank of order greater than 1. This case may be handled in the same framework by noting that if $s \geq 2$ then a variable can always be dropped from \mathbf{x}_t , such that the remaining $m - 1$ variables are cointegrated (see Davidson 1998), and the latter hypothesis can be tested by the same methods. Of course, the variable to be discarded could be y_t , so the choice of normalization should always be considered carefully.

Appendices

A Proof of Proposition 2.1

To fix ideas, assume initially that $\mathbf{w}_t \sim IN(\mathbf{0}, \Sigma)$. Under the assumptions we have

$$E(\Delta^d y_t | \mathcal{X}_{-\infty}^{t-1}, \mathcal{Z}_t^\infty) = [1 - \mathbf{d}' \mathbf{b}_1(L)] \Delta^d y_t - \mathbf{d}' \mathbf{B}_2(L) \Delta^d \mathbf{z}_t + E(\mathbf{d}' \mathbf{w}_t | \mathcal{Z}_t^\infty). \quad (\text{A-1})$$

Expanding equations (2.12) in MA(∞) form expresses $\Delta^d \mathbf{z}_t$ as a linear function of $(\mathbf{w}_{t-j}, j \geq 0)$, with coefficients declining geometrically as the order of lag increases. In view of the Gaussianity it follows that the linear projection $E(\mathbf{d}' \mathbf{w}_t | \mathcal{Z}_t^\infty) = \sum_{j=0}^{\infty} \boldsymbol{\psi}'_j \Delta^d \mathbf{z}_{t+j}$ holds, for a sequence of coefficients $\{\boldsymbol{\psi}_j\}$, also declining geometrically as j increases. Substituting this expression into (A-1) yields the regression equation represented by (2.17), where ε_t defined in (2.16) is a Gaussian martingale difference, and so serially independent, and also totally independent of \mathcal{Z}_1^∞ .

If the variables measurable with respect to $\mathcal{X}_{-\infty}^0$ and \mathcal{Z}_1^∞ were observed and the true parameters $d, \mathbf{B}_0, \dots, \mathbf{B}_p$, and Σ known, randomly drawing a sample $\varepsilon_1, \dots, \varepsilon_T$ and substituting into (2.17) would yield a drawing from (2.18), and so provide the basis for exact bootstrap tests. These ideal tests are not feasible, but we show that our feasible resampling procedure yields tests that are correctly sized in the limit as $T \rightarrow \infty$. The steps that need to be established are the following: (i) the same limiting distributions are obtained with non-Gaussian data, subject to the stated assumptions; (ii) replacing \mathbf{a} in (2.17), and also d , by $T^{1/2}$ -consistent estimates also yields correctly sized tests in the limit; (iii) $T^{1/2}$ -consistency of the OLS coefficients $\hat{\mathbf{a}}$ in (2.17) holds, under truncation of the infinite lead distributions (estimating $\boldsymbol{\psi}_j$ by 0 for $j > r$).

(i) If the data are non-Gaussian, projection (A-1) does not induce independence of the disturbances, so that the independent bootstrap draws do not match the distribution in (2.18) exactly. However, the covariances induced between the variables by the resampling procedure (that is, under the joint distribution of $(\Delta^d y_t, \Delta^d \mathbf{z}'_t)$) match those of the sample data, as in the Gaussian case. Under the assumptions the limiting distributions of F/T and $\hat{\phi}$ are functionals of fractional Brownian motion Y depending only on the realization \mathbf{Z} , d and the covariance parameters. By taking T large enough the bootstrap draws therefore match the distribution in (2.18) arbitrarily closely.

(ii) Next consider the case where $T^{-1/2}$ -consistent estimates $\hat{\mathbf{a}}$, and an extraneous $T^{-1/2}$ -consistent estimate \hat{d} , are substituted for the true \mathbf{a} and d to generate the bootstrap distribution. We establish the conditions for correct asymptotic size given in Proposition 4.3 of Beran and Ducharme (1991). These are: consistency of the estimates; the ‘triangular array convergence condition’; and continuity of the family of limit distributions on the parameter space. Consistency will be addressed under (iii) below. By Davidson (2002), as noted, the limit distributions of F/T and $\hat{\phi}$ belong to parametric families of probability measures depending continuously on $\mathbf{B}_0, \dots, \mathbf{B}_p$, Σ and d , whose elements belong to the vector $\boldsymbol{\theta} \in \Theta$.¹⁴ In the following discussion, which may be applied to either case, consider a family of limiting distributions $F_A(\cdot, \boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. Although these are conditional probability measures depending on \mathbf{Z} , they are identical with the unconditional measures when viewed as members of the parametric family F_A at a given point in the sample space. By the assumption of regularity, at the ‘true point’ $\boldsymbol{\theta} = \bar{\boldsymbol{\theta}}$ the matrices \mathbf{B}_0 and Σ have constant rank under small enough perturbations of their elements, and hence $F_A(\cdot, \boldsymbol{\theta})$ depends continuously on elements of $\boldsymbol{\theta}$ in the neighbourhood of $\bar{\boldsymbol{\theta}}$. With no loss of generality Θ can be defined to exclude non-regular points, since the model would have no solution at such points. Next let $F_T(\cdot, \boldsymbol{\theta})$ denote the generic family of finite sample distributions, derived from (2.18).

¹⁴Under the assumptions, the finite sample distributions may depend on additional parameters such as skewness and kurtosis coefficients, but the limiting distributions are invariant with respect to these.

The ‘triangular array convergence condition’ is $F_T(\cdot, \boldsymbol{\theta}_T) \rightarrow F_A(\cdot, \bar{\boldsymbol{\theta}})$, where the arrow denotes weak convergence and $\{\boldsymbol{\theta}_T\}$ denotes an arbitrary sequence in Θ converging to $\bar{\boldsymbol{\theta}}$. However, under the assumptions, $F_T(\cdot, \boldsymbol{\theta}) \rightarrow F_A(\cdot, \boldsymbol{\theta})$ for arbitrary fixed points $\boldsymbol{\theta} \in \Theta$. Given any sequence $\{\boldsymbol{\theta}_M\}$ converging to $\bar{\boldsymbol{\theta}}$ in Θ , the array $F_T(\cdot, \boldsymbol{\theta}_M)$ converges weakly to the sequence $F_A(\cdot, \boldsymbol{\theta}_M)$ (use the diagonal method, see Davidson 1994, p. 35), and continuity ensures $F_A(\cdot, \boldsymbol{\theta}_M) \rightarrow F_A(\cdot, \bar{\boldsymbol{\theta}})$. The required condition is established by constructing the convergent sequence $F_T(\cdot, \boldsymbol{\theta}_T)$, setting $M = T$ for $T = 1, 2, 3, \dots$

(iii) We adapt the proof of Saikkonen’s (1991) Theorem 4.1. Write

$$\mathbf{q}_t = (\Delta^d y_{t-1}, \dots, \Delta^d y_{t-p}, \Delta^d \mathbf{z}'_{t-p}, \dots, \Delta^d \mathbf{z}'_{t+r})' \quad ((2q + r + 1) \times 1).$$

The regression errors-of-estimate have the form

$$T^{1/2}(\hat{\mathbf{a}} - \mathbf{a}) = \hat{\mathbf{R}}^{-1} T^{-1/2} \sum_{t=1+p}^{T-r} \mathbf{q}_t (\varepsilon_t + s_t)$$

where $s_t = \sum_{j=r+1}^{\infty} \boldsymbol{\psi}'_j \Delta^d \mathbf{z}_{t+j}$ and

$$\hat{\mathbf{R}} = T^{-1} \sum_{t=1+p}^{T-r} \mathbf{q}_t \mathbf{q}'_t.$$

Let $\mathbf{R} = E(\hat{\mathbf{R}})$. Saikkonen’s Lemmas A1-A4 can then be used to show that

$$\|\mathbf{R}^{-1}\|_1 = O(1), \quad \|\hat{\mathbf{R}}^{-1} - \mathbf{R}^{-1}\|_1 = O_p(r/T^{1/2}) \quad (\text{A-2})$$

where $\|\cdot\|_1$ is the matrix norm defined by Said and Dickey (1984), p. 601. Subject to the condition

$$\sum_{j=r}^{\infty} \|\boldsymbol{\psi}_j\| = o(T^{-1/2}) \quad (\text{A-3})$$

where $\|\cdot\|$ denotes the Euclidean norm, we can then follow Saikkonen (1991) Lemmas A5 and A6 to show that

$$T^{-1/2} \sum_{t=1+p}^{T-r} \mathbf{q}_t \varepsilon_t = O_p(r^{1/2}), \quad T^{-1/2} \sum_{t=1+p}^{T-r} \mathbf{q}_t s_t = o_p(r^{1/2}). \quad (\text{A-4})$$

Combining (A-2) with (A-4) leads to the conclusion that $\hat{\mathbf{a}} - \mathbf{a} = O_p(T^{-1/2})$ if $r = o(T^{1/3})$, and also that the asymptotic distribution of $T^{1/2}(\hat{\mathbf{a}} - \mathbf{a})$ is invariant to the omission of the leads $\Delta^d \mathbf{z}'_{t+j}$ for $j > r$. Condition (A-3) certainly holds if $\|\boldsymbol{\psi}_j\| = \rho^j$ for some $|\rho| < 1$, which is true by assumption. ■

B The Shin Statistic under Fractional Integration

As in the text, let SH refer to Shin (1994) and SK to Saikkonen (1991). These authors use similar but not identical notation. In the discussion following, we adopt the appropriate notation, as far as possible, when considering the generalization of one or other source.

The KPSS statistic for the hypothesis that $\varepsilon_t \sim \text{I}(0)$ is

$$\hat{\eta}_\mu = \frac{1}{T^2 s^2(l)} \sum_{t=1}^T S_t^2 \quad (\text{B-1})$$

where $S_t = \sum_{s=1}^t (\varepsilon_s - \bar{\varepsilon})$ and

$$s^2(l) = T^{-1} \left(\sum_{t=2}^T \varepsilon_t^2 + 2 \sum_{j=1}^l w_{jl} \sum_{t=j+2}^T \varepsilon_t \varepsilon_{t-j} \right) \quad (\text{B-2})$$

is the Newey-West (1987)-type of autocorrelation-consistent variance estimator where typically the Bartlett weights $w_{jl} = 1 - j/(l+1)$ are adopted, and l is the bandwidth, diverging at a fractional power of T . Consistency of the test against the alternative of $I(1)$ requires $l = o(T^{1/2})$. In SH's application, ε_t is the residual from a hypothesized cointegrating regression, estimated by SK's efficient procedure described in the text. In this appendix we outline the generalizations of SH's Theorems 2 and 3 to the fractional case. These derive the asymptotic distribution under null and alternative respectively. We show that, under suitable restrictions to be specified, the same null distribution applies with fractional BMs replacing regular BMs, and that the test is consistent against alternatives $b < d$ where $l = o(T^{2(d-b)})$.

For brevity we consider only the pure fractional process without corrections for mean or trend, but these extensions are easily provided by analogy with SH's arguments. The key results on the weak convergence of nonstationary fractional processes are given in Davidson and de Jong (2000) and Davidson (2003b). We state the essentials as follows:

Lemma B.1 *Let $\mathbf{x}_t \sim I(d)$ for $\frac{1}{2} < d < \frac{3}{2}$, and let $\mathbf{v}_t \sim I(0)$. Then, subject to the additional assumptions of the cited results,*

- (i) $T^{1/2-d} \mathbf{x}_{[Tr]} \xrightarrow{d} \mathbf{X}_d(r)$, $0 \leq r \leq 1$, $\frac{1}{2} < d < \frac{3}{2}$
- (ii) $T^{-2d} \sum_{t=1}^T \mathbf{x}_t \mathbf{x}_t' \xrightarrow{d} \int_0^1 \mathbf{X}_d \mathbf{X}_d' dr$, $\frac{1}{2} < d < \frac{3}{2}$
- (iii) $T^{-d} \sum_{t=1}^T \mathbf{x}_t \mathbf{v}_t' \xrightarrow{d} \int_0^1 \mathbf{X}_d d\mathbf{V}' + \mathbf{\Lambda}_{XV}$, $1 \leq d < \frac{3}{2}$
- (iv) $T^{-1} \sum_{t=1}^T \mathbf{x}_t \mathbf{v}_t' \xrightarrow{pr} \mathbf{\Lambda}_{XV}$, $\frac{1}{2} < d < 1$, $\mathbf{\Lambda}_{XV} \neq \mathbf{0}$
- (v) $T^{-d} \sum_{t=1}^T \mathbf{x}_t \mathbf{v}_t' \xrightarrow{d} \int_0^1 \mathbf{X}_d d\mathbf{V}'$, $\frac{1}{2} < d < 1$, $\mathbf{\Lambda}_{XV} = \mathbf{0}$

where \mathbf{X}_d is a fractional Brownian motion with parameter d , \mathbf{V} is a regular Brownian motion, and $\mathbf{\Lambda}_{XV} = \lim_{T \rightarrow \infty} T^{-\max(1,d)} \sum_{t=1}^T E(\mathbf{x}_t \mathbf{v}_t') < \infty$.

Parts (i) and (ii) are by Davidson and de Jong 2000, Theorem 3.1, and the continuous mapping theorem. Parts (iii)-(v) are by Davidson and de Jong 2000, Theorems 4.1-4.3, and Theorem 4.1 of Davidson (2003b). Note that the additional assumptions cited characterize the short memory and finite variance properties of the increments required for weak convergence. They are complex and hence not given here explicitly, but are generally fairly mild.

These results have to be used in place of the weak convergences to Brownian motion assumed explicitly or implicitly in the proofs of Theorems 1 and 2 and Lemma 1 of SH. In these arguments (adopting SK's notation) let the symbol Δ be replaced by Δ^d , and let N^{-1} be replaced by N^{-d} , in the definition of the matrix \mathbf{D}_T , where $N = T - 2K$, and use the results of Lemma B.1 where appropriate.

Let \mathbf{z}_t be as defined in (2.20b), and take the case of Lemma B.1 where $\mathbf{x}_t = \mathbf{z}_t$ and $\mathbf{v}_t = \Delta^d \mathbf{z}'_{t-j}$ for $-K \leq j \leq K$. Observe that

$$T^{-d-1/2} \sum_{t=1}^T \mathbf{z}_t \Delta^d \mathbf{z}'_t = O_p(T^{\max(1/2-d, -1/2)}) \quad (\text{B-3})$$

by Lemma B.1(iii) or (iv). It follows that, with the normalization matrix \mathbf{D}_T defined as in SK (page 18) but with $N^{-d}\mathbf{I}_{n_2}$ as its top-left block, Lemmas A2 and A4 of SK hold for $d \geq 1$. Otherwise, note that the convergence rates specified in these lemmas become $O_p(K/T^{d-1/2})$. Therefore, the condition for the terms $\|E_2\|$ and $\|E_3\|$ to vanish in SK's equation (A.13) is $K^{3/2}/T^{\min(1/2, d-1/2)} \rightarrow 0$, or

$$K = o(T^{\min\{1/3, (2d-1)/3\}})$$

which accordingly must replace SK's condition (21). To determine whether this condition will be compatible with SK's condition (22), define $\delta > 0$ by $\sum_{|j|>K} \|\boldsymbol{\pi}_j\| = o(K^{-\delta})$, and $\mu < 0$ by $K = T^\mu$. It can be verified that there exists μ to satisfy SK's conditions (21) and (22) if $\delta \max\{1, 2d-1\} > \frac{3}{2}$.

Now consider the formulae in (2.23). The weights on \mathbf{w}_{t-j} cannot decline less slowly than $O(j^{-d-1})$, so that the cross-autocovariances $E(v_{1t}\mathbf{v}_{2,t+h})$ converge as $|h| \rightarrow \infty$ not less slowly than $O(h^{-2d-1})$. Note that the sequence $\{\boldsymbol{\pi}_j\}_{j=-\infty}^{\infty}$ has the same order of magnitude. It can be thought of as the limit of the finite sequences $\{\boldsymbol{\pi}_{-K}, \dots, \boldsymbol{\pi}_K\}$ where

$$\begin{bmatrix} \boldsymbol{\pi}_{-K} \\ \vdots \\ \boldsymbol{\pi}_K \end{bmatrix} = \boldsymbol{\Omega}_{2K}^{-1} \boldsymbol{\omega}_{12K}. \quad (\text{B-4})$$

Here, $\boldsymbol{\Omega}_{2K}$ is a matrix of block-Toeplitz form with blocks defined by $\{\boldsymbol{\Omega}_{2K}\}_{jk} = E(\mathbf{v}_{2,t+j}\mathbf{v}'_{2,t+k})$ for $j, k = -K, \dots, K$, and $\boldsymbol{\omega}_{12K} = (E(\mathbf{v}'_{2,t-K}v_{1t}, \dots, \mathbf{v}'_{2,t+K}v_{1t}))'$. Thus, $\delta \geq 2d$, at worst. It can be verified that the inequality $2d \max\{1, 2d-1\} > \frac{3}{2}$ is satisfied for $d > 0.911$. For smaller values of d there is the possibility that the results of this appendix do not hold. To proceed, we assume that δ satisfies whatever bound is required.

Now, consider the case $\mathbf{v}_t = \varepsilon_t$, in SH's notation (corresponding to our e_t in (2.19)). It is easy to see that (in the notation of Lemma B.1)

$$T^{-d} \sum z_t \varepsilon_t \xrightarrow{d} \int_0^1 \mathbf{X}_d d\mathbf{V}' \quad (\text{B-5})$$

for $\frac{1}{2} < d < \frac{3}{2}$ when the processes are asymptotically independent, such that $\boldsymbol{\Lambda}_{XV} = \mathbf{0}$ holds. Hence, given his Lemma 1, and noting that the calculus of fractional Brownian motion obeys the familiar rules of the regular case, subject to Lemma B.1, SH's Theorem 2 follows directly. The convergence indicated at the bottom of page 112 in SH here becomes (adapting SH's notation to our own)

$$\begin{aligned} T^{-1/2} S_{[Tr]} &= T^{-1/2} \sum_{j=1}^{[Tr]} \varepsilon_j - T^{-1/2-d} \sum_{j=1}^{[Tr]} \mathbf{Z}'_j T^d (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) + o_p(1) \\ &\rightarrow Q_{Bd} \end{aligned} \quad (\text{B-6})$$

Lemma B.1 has been used to assign orders of magnitude in the second equality, and of course the indicated weak limit is different from SH's case, in that it involves functionals of fractional Brownian motion.

To prove consistency of the test, we adapt SH's Theorem 3, but only establish orders of magnitude in this instance. Under the alternative, $y_t - \boldsymbol{\gamma}'\mathbf{z}_t = u_t \sim \text{I}(d-b)$ for $b < d$. For $b > 0$ this alternative still represents cointegration in the sense that the noncointegration hypothesis is also false, and for $d-b < \frac{1}{2}$ the residual is stationary. However, we cannot assume that SH's Lemma 1 will hold in respect of this long memory residual. Therefore, the $\text{I}(0)$ increments $\Delta^{d-b}u_t$ and $\Delta^d z_t$ may be correlated, in general. In this situation, Davidson (2002), Section 4.1 shows

that $\hat{\gamma} - \gamma = O_p(T^{\max\{-b, 1-2d\}})$. The ‘worst case’ from the point of view of the consistency of the test is however $\hat{\gamma} - \gamma = O_p(T^{-b})$, since in the other case the second term in the middle member of (B-6) is actually diverging more rapidly than the first one. Assuming this case, we therefore have, in place of the expression on page 114 of SH,

$$\begin{aligned} T^{b-d-1/2} S_{[Tr]} &= T^{b-d-1/2} \sum_{j=1}^{[Tr]} \varepsilon_j - T^{-d-1/2} \sum_{j=1}^{[Tr]} \mathbf{Z}'_j T^b (\hat{\gamma} - \gamma) \\ &= O_p(1) \end{aligned} \tag{B-7}$$

and hence (at worst, note)

$$T^{-2} \sum_{t=1}^T S_t^2 = O_p(T^{2(d-b)}). \tag{B-8}$$

Now consider the denominator of the KPSS statistic, $s^2(l)$ in (B-2). In the case $d - b < \frac{1}{2}$, ε_t is covariance stationary and the autocovariance estimators contained in this expression are $O_p(1)$. Hence, $s_l^2 = O_p(l)$, and the test is consistent whenever $l = o(T^{2(d-b)})$, with $\hat{\eta}_\mu = O_p(T^{2(d-b)}/l)$. If $d - b > \frac{1}{2}$ then $T^{-1} \sum_{t=2}^T \varepsilon_t^2 = O_p(T^{2(d-b)-1})$. Thus, in this case $\hat{\eta}_\mu = O_p(T/l)$, which includes the case of the I(1) alternative in particular.

C Proof of Proposition 2.2

The argument follows the lines and notation of Proposition 2.1, with appropriate variations. Under H_0 the final form equations for $\Delta^d \mathbf{x}_t$, from (2.5), are given in (2.23b). Except under strong exogeneity, such that the polynomial matrix is block-triangular, the fractional difference operator cannot be cancelled on the right-hand side. It follows, noting the discussion in Appendix B, that the MA(∞) representation of $\Delta^d \mathbf{z}_t$ contains \mathbf{w}_{t-j} with lag weights that generally decline hyperbolically, at a rate not slower than $O(j^{-1-d})$.¹⁵ However, the same reasoning as before leads to the conclusion that, under Gaussianity, the disturbances of (2.24) are serially independent, and totally independent of \mathcal{Z}_1^∞ . We conclude that resampling from this distribution, as described in the text, yields a drawing from the conditional distribution of $\{y_1, \dots, y_T | \mathcal{X}_0, \mathcal{Z}_1^\infty\}$, and hence, of the conditional distribution of the Shin test statistic.

Subject to the conditions of the theorem, the OLS estimates of the equation are again consistent, notwithstanding the inclusion of the $I(d)$ regressors, which form a cointegrating combination under H_0 (compare e.g., Stock 1987). Similarly to the non-cointegration test case, under the assumptions the asymptotic bootstrap distribution is invariant to non-Gaussianity, and the substitution of estimated parameters. ■

D Proof of Theorem 3.1

Since F is monotone nondecreasing and right continuous, the points $y \in [0, 1]$ belong to one of three equivalence classes,

$$\begin{aligned} A &= \{y : y = F(\phi(y)) < F(x) \text{ for } x > \phi(y)\} \\ B &= \{y : y = F(\phi(y)) = F(x) \text{ for } \phi(y) < x \leq \phi(y) + \delta, \delta > 0\} \end{aligned}$$

¹⁵The exceptional case is $d = 1$, where the coefficients in the expansion of Δ^d are truncated at $j = 1$, so the geometric rate of convergence applies overall. Note, the coefficients are negative for $j > 1$ when $\frac{1}{2} < d < 1$, and positive for $j > 1$ when $1 < d < \frac{3}{2}$.

$$C = \{y : F(\phi(y)-) \leq y < F(\phi(y))\}.$$

A corresponds to (the images of) points where F is strictly increasing, B to points where F is constant, and C to jump points of F . Consider each case in turn.

In the case $y \in A$, $F(x) \leq F(\phi(y))$ iff $x \leq \phi(y)$. It follows that the events $\{X \leq \phi(y)\}$ and $\{F(X) \leq y\}$ have the same probability. Since $P(X \leq \phi(y)) = y$ by definition of F , the first case of (3.2) follows.

In the case $y \in B$, $\phi(y)$ is by definition the infimum of an interval on which F is constant, open above by right continuity. Letting $\phi(y) + \delta$ for $\delta > 0$ denote the supremum of this interval, suppose that $F(\phi(y) + \delta) \in A$. Then,

$$\begin{aligned} P(F(X) \leq y) &= P(F(X) \leq F(\phi(y) + \delta)). \\ &= P(X \leq \phi(y) + \delta) \\ &= F(\phi(y) + \delta) \\ &= y \end{aligned} \tag{D-1}$$

as required for the first case of (3.2), where the second equality holds by the previous argument. and the last equality is by the continuity of F at $\phi(y) + \delta$. Otherwise, necessarily, $F(\phi(y) + \delta) \in C$ and $F((\phi(y) + \delta)-) = F(\phi(y)) = y$ by definition. Note that in this case, $F(X) \leq F((\phi(y) + \delta)-)$ iff $X < \phi(y) + \delta$, and therefore

$$\begin{aligned} P(F(X) \leq y) &= P(F(X) \leq F((\phi(y) + \delta)-)) \\ &= P(X < \phi(y) + \delta) \\ &= F((\phi(y) + \delta)-) \\ &= y. \end{aligned} \tag{D-2}$$

In the case $y \in C$,

$$\begin{aligned} P(F(X) \leq y) &= P(F(X) \leq F(\phi(y)-)) + P(F(\phi(y)-) < F(X) \leq y) \\ &= P(F(X) \leq F(\phi(y)-)) \end{aligned} \tag{D-3}$$

where the second equality holds because if $F(X) > F(\phi(y)-)$ then $F(X) > y$ with probability 1. To evaluate the remaining term, consider $P(F(X) \leq F(\phi(y) - \delta))$ for $\delta > 0$. By right-continuity, taking δ small enough ensures that the point $F(\phi(y) - \delta)$ belongs either to class A or class B , and hence by previous arguments,

$$P(F(X) \leq F(\phi(y) - \delta)) = F(\phi(y) - \delta). \tag{D-4}$$

The second case of equation (3.2) now follows by letting $\delta \rightarrow 0$. ■

References

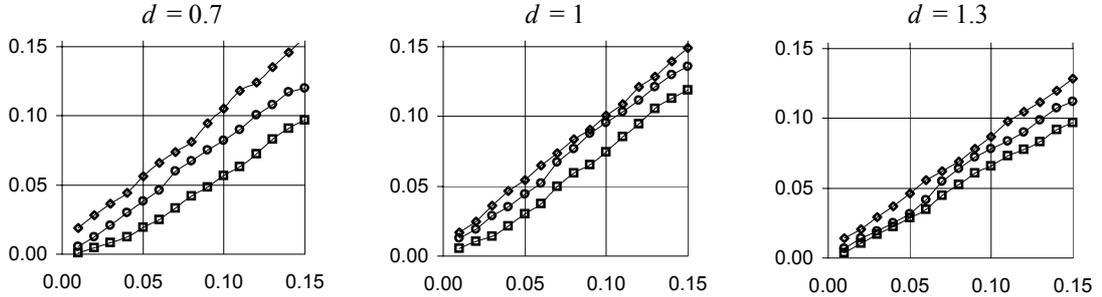
- Basawa, I. V., Mallik, A. K., McCormick W. P., Reeves, J. H., Taylor, R. L., 1991a, Bootstrapping unstable first order autoregressive processes, *Annals of Statistics* 19, 2, 1098–1101
- Basawa, I. V., Mallik, A. K., McCormick W. P., Reeves, J. H. and Taylor, R. L., 1991b, Bootstrap test of significance and sequential bootstrap estimation for unstable first order autoregressive process, *Communications in Statistics A*, 20 1015–1026.
- Beran, R. 1988, Prepivoting test statistics: a bootstrap view of asymptotic refinements, *Journal of the American Statistical Association* 83, 687–697.

- Beran, R. and G. R. Ducharme 1991, *Asymptotic Theory for Bootstrap Methods in Statistics*. Centre de Recherches Mathematiques, Universite de Montreal, Canada.
- Breitung, J., and U. Hassler 2002, Inference on the cointegration rank in fractionally integrated processes, *Journal of Econometrics* 110(2), 167–185.
- Davidson, J. 1994, *Stochastic Limit Theory*. Oxford, Oxford University Press
- Davidson, J. 1998, Structural relations, cointegration and identification: some simple results and their application, *Journal of Econometrics* 87, 87–113.
- Davidson, J. 2002 A model of fractional cointegration, and tests for cointegration using the bootstrap. *Journal of Econometrics* 110 (2) 187–212.
- Davidson, J. 2003a Testing for fractional cointegration: a reappraisal of the relationship between government popularity and economic performance in the UK. Working paper at <http://www.cf.ac.uk/carbs/davidsonje/bootpop2.pdf>.
- Davidson, J. 2003b, Convergence to stochastic integrals with fractionally integrated integrator processes: theory, and applications to fractional cointegration, Working paper at <http://www.cf.ac.uk/carbs/davidsonje/fracstochint7.pdf>
- Davidson, J. and R. M. de Jong 2000, The functional central limit theorem and convergence to stochastic integrals II: fractionally integrated processes. *Econometric Theory* 16, 5, 643–666.
- Davidson, R. and J. MacKinnon 1999, The size distortion of bootstrap tests, *Econometric Theory* 15, 361-376
- Davidson, R. and J. MacKinnon 2000 Improving the reliability of bootstrap tests. Working paper, at <http://qed.econ.queensu.ca/pub/faculty/mackinnon/>
- Dickey, D. A. and W. A. Fuller 1979: Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association* 74, 427–31.
- Doornik, J. A. 1999, *Object-Oriented Matrix Programming Using Ox*, 3rd Edn. London: Timberlake Consultants Press (<http://www.nuff.ox.ac.uk/Users/Doornik/>).
- Granger, C. W. J. 1986, Developments in the study of cointegrated economic variables, *Oxford Bulletin of Economics and Statistics* 48, 213–228.
- Hall, P. 1992, *The Bootstrap and Edgeworth Expansion*, New York: Springer-Verlag.
- Horowitz, J. L. 1997, Bootstrap methods in econometrics: theory and numerical performance, Ch. 7 of *Advances in Economics and Econometrics: Theory and Applications* eds. Kreps and Wallis, (Econometric Soc. 7th World Congress). Cambridge University Press.
- Horowitz, J. L. 2000, The Bootstrap *Handbook of Econometrics* Vol. 5 North-Holland Elsevier (forthcoming).
- Johansen, Søren 1988, Statistical Analysis of Cointegration Vectors, *Journal of Economic Dynamics and Control* 12, 231–54.
- Johansen, Søren, 1991, Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models, *Econometrica* 59(6), 1551–80.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159–78.
- Lettau, M. and S. Ludvigson 2001, Consumption, aggregate wealth and expected stock returns. *Journal of Finance* LVI(3), 815-849.
- Marinucci, D. and P. M. Robinson, 1999, Alternative forms of fractional Brownian motion. *Journal of Statistical Inference and Planning* 80, 111–122.
- Nankervis, J. C. 2001, Stopping rules for double bootstrap tests. Working paper, University of Surrey.

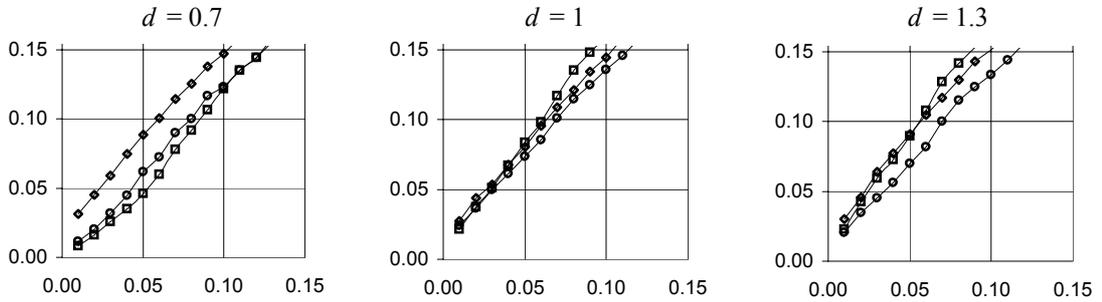
- Newey, W. K. and K. D. West 1987, A simple, positive semi-definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–8.
- Phillips, P. C. B. and S. Ouliaris 1990, Asymptotic properties of residual based tests for cointegration *Econometrica* 58, 165–93.
- Phillips, P. C. B. and P. Perron 1988, Testing for unit roots in time series regression. *Biometrika* 75, 335–46.
- Robinson, P. M. 1994, Semiparametric analysis of long memory time series, *Annals of Statistics* 22 (1), 515–539.
- Said, S. E. and D. A. Dickey 1984, Testing for unit roots in autoregressive moving average models of unknown order. *Biometrika* 71, 599–607.
- Saikkonen, P. 1991, Asymptotically efficient estimation of cointegration regressions, *Econometric Theory* 7, 1–21.
- Shin, Y. 1994, A residual-based test of the null of cointegration against the alternative of no cointegration, *Econometric Theory* 10, 91–115.

Figure 1
 Empirical CDFs of bootstrap p -values under H_0 :
 Test for noncointegration based on the F statistic

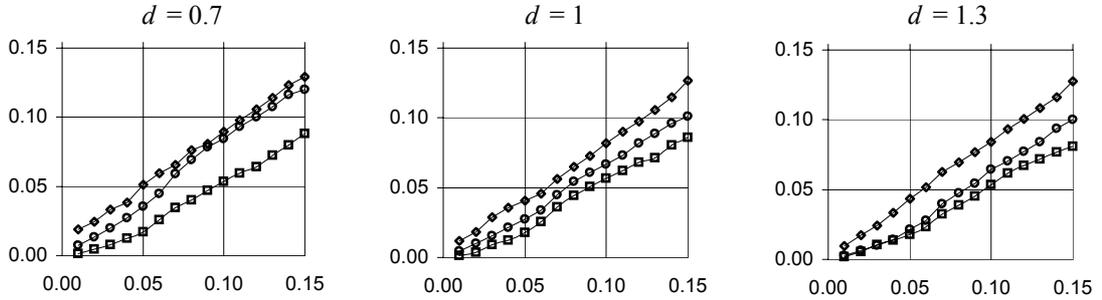
Case II (independent increments). ARFIMA(0, d , 0), 1 lead, 1 lag.



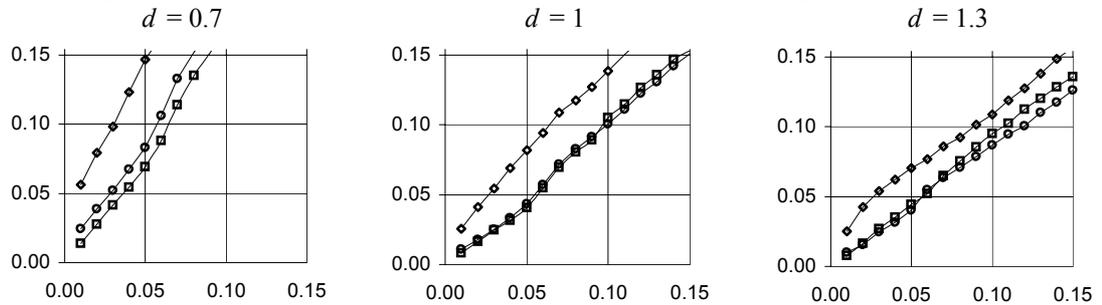
Case CI (correlated increments). ARFIMA(0, d , 0), 1 lead, 1 lag.



Case IA (autoregressive increments). ARFIMA(1, d , 0), 2 leads, 2 lags.



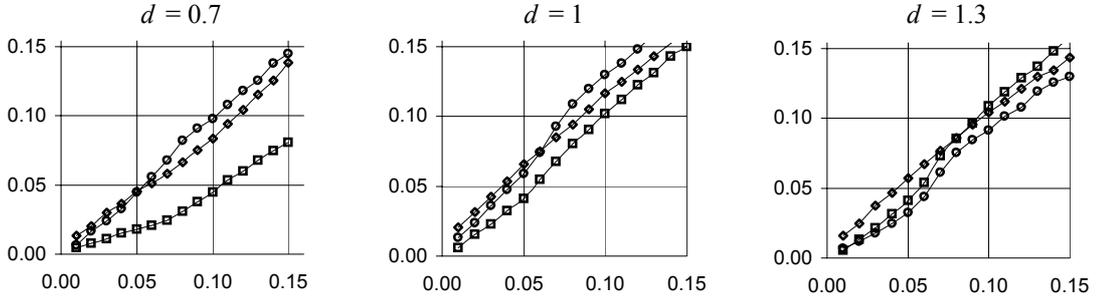
Case CA (autoregressive, correlated increments). ARFIMA(1, d , 0), 4 leads, 4 lags.



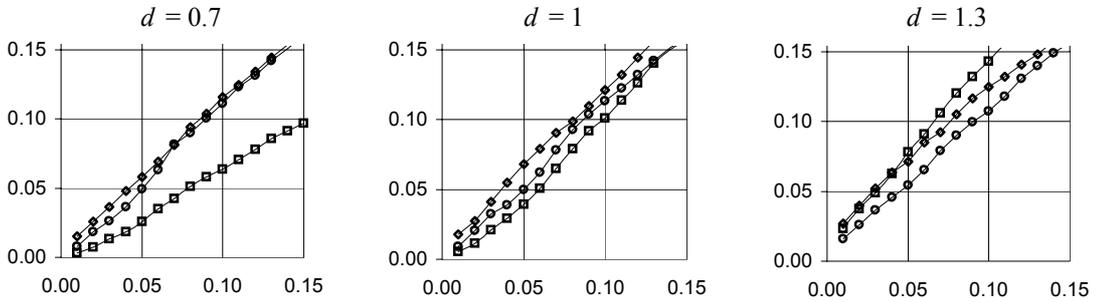
Key: \square = regular bootstrap, \diamond = double bootstrap, \circ = fast double (FD1)

Figure 2
 Empirical CDFs of bootstrap p -values under H_0 :
 Test for noncointegration based on the DW statistic

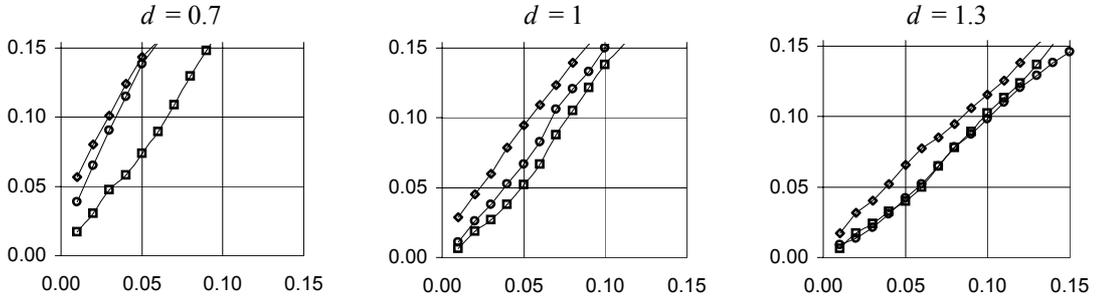
Case II (independent increments). ARFIMA(0, d , 0), 1 lead, 1 lag.



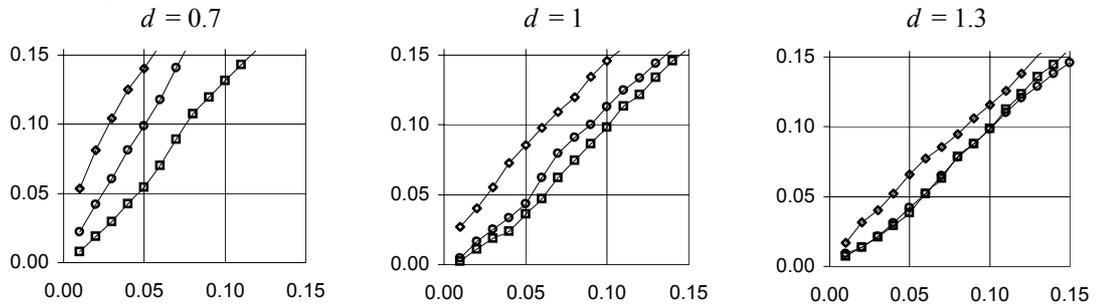
Case CI (correlated increments). ARFIMA(0, d , 0), 1 lead, 1 lag.



Case IA (autoregressive increments). ARFIMA(1, d , 0), 2 leads, 2 lags.



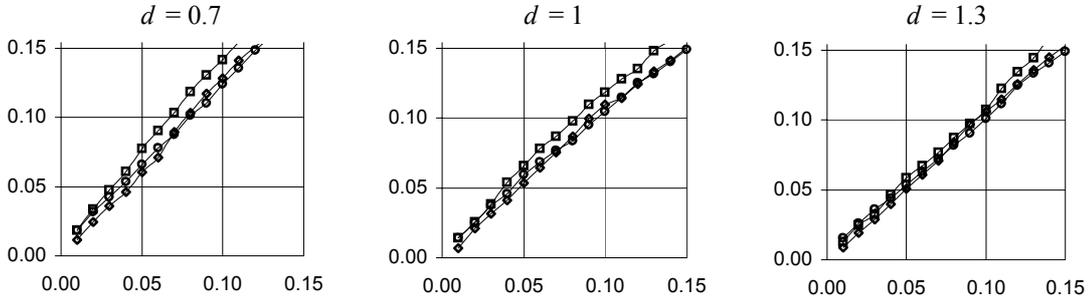
Case CA (autoregressive, correlated increments). ARFIMA(1, d , 0), 4 leads, 4 lags.



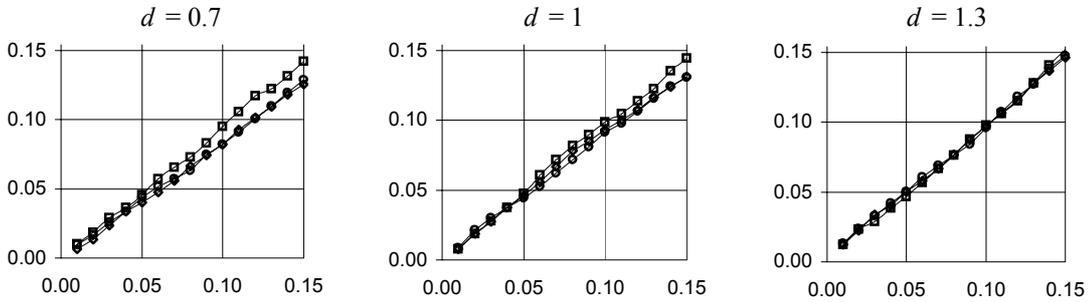
Key: \square = regular bootstrap, \diamond = double bootstrap, \circ = fast double (FD1)

Figure 3
 Empirical CDFs of bootstrap p -values under H_0 :
 Shin's test for cointegration, restricted ECM method ($\mu \leq -0.5$)

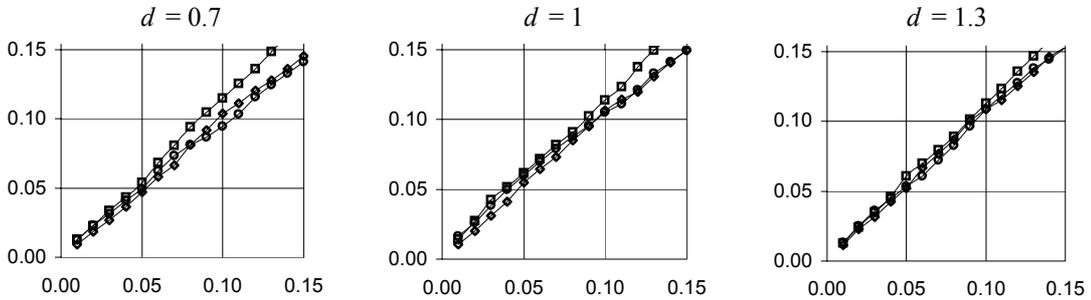
Case II (independent increments). ARFIMA(1, d , 0), 5 leads, 2 lags.



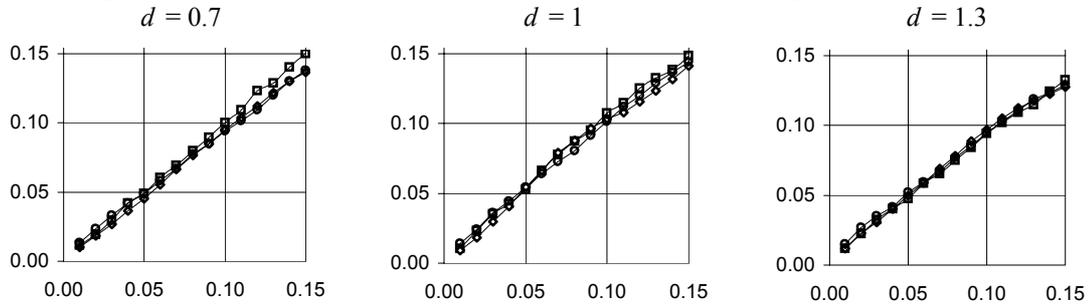
Case CI (correlated increments). ARFIMA(1, d , 0), 5 leads, 2 lags..



Case IA (autoregressive increments). ARFIMA(1, d , 0), 5 leads, 2 lags.



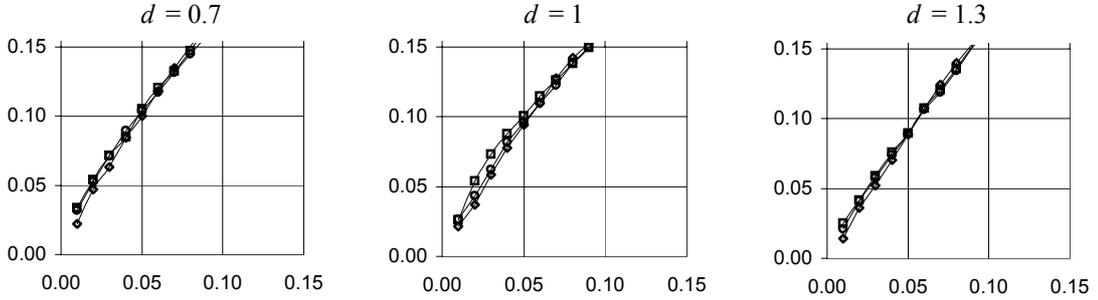
Case CA (autoregressive correlated increments). ARFIMA(1, d , 0), 5 leads, 2 lags.



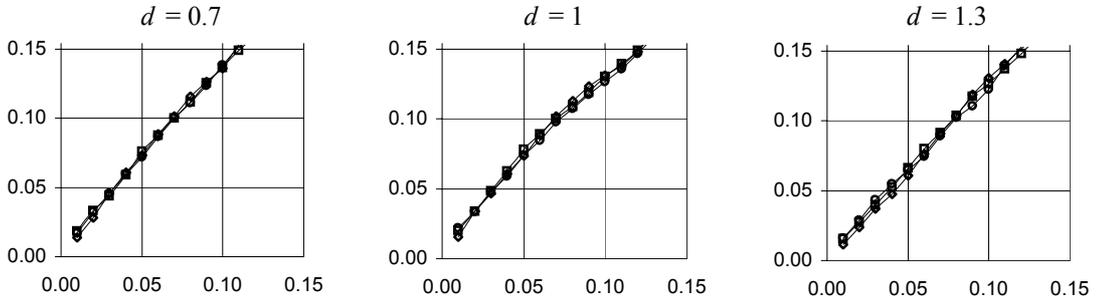
Key: \square = regular bootstrap, \diamond = double bootstrap, \circ = fast double (FD1)

Figure 4
 Empirical CDFs of bootstrap p -values under H_0 :
 Shin's test for cointegration, $I(0)$ regression method.

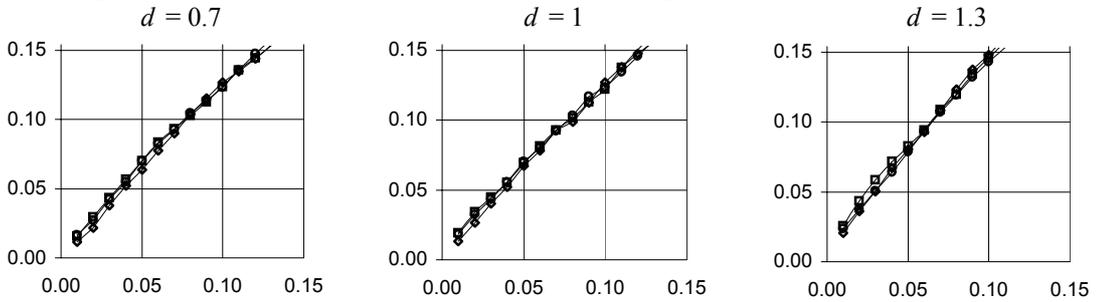
5 leads, 2 lags.



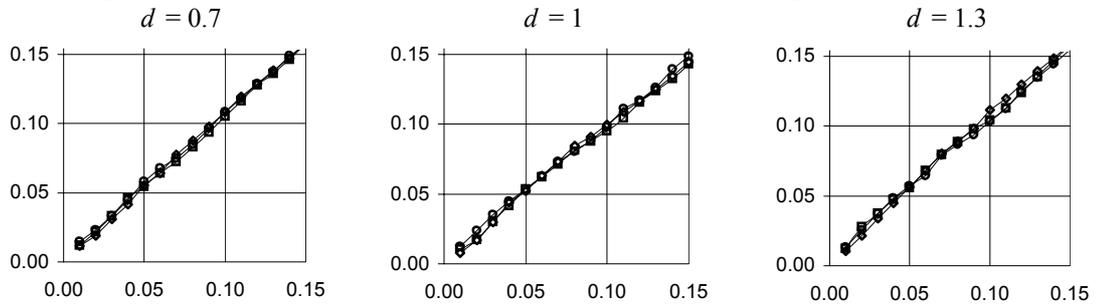
Case CI (correlated increments). ARFIMA(0, d ,0), 5 leads, 2 lags.



Case IA (autoregressive increments). ARFIMA(1, d ,0), 5 leads, 2 lags.



Case CA (autoregressive correlated increments). ARFIMA(1, d ,0), 5 leads, 2 lags.



Key: \square = regular bootstrap, \diamond = double bootstrap, \circ = fast double (FD1)

Table 2.

Rejection frequencies for 0.05-level tests (regular bootstrap) under selected alternatives

H ₀ : $b = 0$						
Model	Alternative		F Statistic		DW Statistic	
	d	b	Nominal	Corrected	Nominal	Corrected
II	0.7	0.4	0.176	0.294	0.088	0.233
	0.7	0.7	0.305	0.485	0.214	0.456
	1	0.5	0.445	0.515	0.386	0.432
	1	1	0.857	0.917	0.902	0.926
	1.3	0.5	0.348	0.467	0.421	0.380
	1.3	1	0.966	0.986	0.975	0.968
CI	0.7	0.4	0.496	0.525	0.483	0.604
	0.7	0.7	0.798	0.811	0.884	0.943
	1	0.5	0.746	0.762	0.742	0.767
	1	1	0.998	0.998	1	1
	1.3	0.5	0.781	0.473	0.797	0.572
	1.3	1	0.998	0.961	0.999	0.997
CA	0.7	0.4	0.811	0.65	0.827	0.688
	0.7	0.7	0.964	0.892	0.986	0.962
	1	0.5	0.941	0.815	0.951	0.848
	1	1	1	0.995	1	1
	1.3	0.5	0.977	0.814	0.970	0.815
	1.3	1	1	0.999	1	1

H ₀ : $b = d$			Shin Statistic			
Model	Alternative		ECM, $\mu \leq -0.5$		I(0) Regression	
	d	b	Nominal	Corrected	Nominal	Corrected
II	0.7	0.3	0.271	0.204	0.395	0.222
	0.7	0	0.463	0.391	0.614	0.403
	1	0.5	0.359	0.282	0.491	0.293
	1	0	0.717	0.667	0.774	0.630
	1.3	0.8	0.352	0.327	0.443	0.331
	1.3	0.3	0.733	0.711	0.814	0.732
CI	0.7	0.3	0.227	0.227	0.349	0.294
	0.7	0	0.490	0.490	0.601	0.540
	1	0.5	0.289	0.289	0.431	0.359
	1	0	0.733	0.733	0.788	0.741
	1.3	0.8	0.286	0.286	0.429	0.364
	1.3	0.3	0.710	0.710	0.770	0.716
CA	0.7	0.3	0.200	0.200	0.290	0.266
	0.7	0	0.555	0.555	0.618	0.582
	1	0.5	0.277	0.243	0.385	0.342
	1	0	0.777	0.752	0.761	0.745
	1.3	0.8	0.304	0.304	0.381	0.352
	1.3	0.3	0.706	0.706	0.765	0.752

Table 3.
Estimated d values from Whittle estimation of ARFIMA model

	Consumption 1	Consumption 2	Asset Wealth	Labour Income
ARFIMA	(0,d,0)	(0,d,3)	(0,d,0)	(0,d,0)
d (s.e.)	1.29 (0.058)	1.091 (0.102)	1.045 (0.061)	1.094 (0.043)
Q(12)	23.74	14.71	3.7	14.11
SBC	1143.72	1139.31	859	1012

Table 4
US consumption, income and wealth: tests of non-cointegration.

Consumption Model	Leads	Lags	Bootstrap Method	F statistic	DW statistic
				p values	p values
1	2	2	regular	0.032	0
			double	0.042	0
			FDB1	0.034	0
	3	3	regular	0.054	0.014
			double	0.066	0.024
			FDB1	0.05	0.026
2	2	2	regular	0.024	0
			double	0.018	0
			FDB1	0.018	0
	3	3	regular	0.05	0.002
			double	0.062	0.002
			FDB1	0.048	0

Table 5.
US consumption, income and wealth: Shin tests of cointegration.

Leads	Lags	Bootstrap Method	p -values	
			ECM method ($\mu \leq -0.5$)	I(0) Regression Method
5	2	regular	0.015	0
		double	0.015	0.003
		FDB1	0.015	0.001
6	3	regular	0.016	0
		double	0.020	0.002
		FDB1	0.044	0.002