

A Model of Fractional Cointegration, and Tests for Cointegration Using the Bootstrap*

James Davidson[†]
Cardiff University

Revised, April 2001

Abstract

The paper proposes a framework for modelling cointegration in fractionally integrated processes, and considers methods for testing the existence of cointegrating relationships using the parametric bootstrap. In these procedures, ARFIMA models are fitted to the data, and the estimates used to simulate the null hypothesis of non-cointegration in a vector autoregressive modelling framework. The simulations are used to estimate p -values for alternative regression-based test statistics, including the F goodness-of-fit statistic, the Durbin-Watson statistic and estimates of the residual d . The bootstrap distributions are economical to compute, being conditioned on the actual sample values of all but the dependent variable in the regression. The procedures are easily adapted to test stronger null hypotheses, such as statistical independence. The tests are not in general asymptotically pivotal, but implemented by the bootstrap, are shown to be consistent against alternatives with both stationary and nonstationary cointegrating residuals. As an example, the tests are applied to the series for UK consumption and disposable income. The power properties of the tests are studied by simulations of artificial cointegrating relationships based on the sample data. The F test performs better in these experiments than the residual-based tests, although the Durbin-Watson in turn dominates the test based on the residual d .

1 Introduction

Cointegration methods are well established as a basis for testing relationships amongst nonstationary time series exhibiting stochastic trends. These methods invoke the ‘I(1) paradigm’ to represent nonstationary series. The crucial implication is that the ordinary differences of a series are stationary and weakly dependent, having either a stable and invertible ARMA form or, at worst, a representation with comparable asymptotic properties. To validate the asymptotic analysis customarily used in these methods it is necessary, at least, that the partial sums of the difference series, normalised by $n^{-1/2}$ where n denotes sample size, should converge to Brownian motion.¹ In routine applications of the method this assumption is rarely tested, beyond making the decision whether to treat a series as I(0) or I(1).

The assumption in cointegration analysis that unity is a ‘special’ value for the autoregressive root has often been criticised, especially by Bayesians (e.g. Sims and Uhlig, 1991). An alternative

*Paper presented at the Cardiff Conference on Long Memory and Nonlinear Time Series, July 9th-11th 2000. Research supported by the ESRC under award L138251025. I thank Uwe Hassler and two anonymous referees for valuable comments.

[†]Email: davidsonje@cf.ac.uk

¹Davidson (2001a) analyses conditions for convergence in a variety of time series models.

approach is to embed $I(1)$ processes in the wider class of fractionally integrated processes, or $I(d)$ for real $d > 0$. Baillie (1996) provides a survey. Recent papers looking at fractional cointegration include Cheung and Lai (1993), Baillie and Bollerslev (1994), Robinson and Marinucci (1998), Jeganathan (1999), Kim and Phillips (2000), Marinucci (2000), Davidson (2001b). For applications see among others Booth and Tse (1995), Masih and Masih (1995), Sephton (1996), and Duecker and Startz (1998).²

In this context, none of the usual inference procedures are appropriate. On the one hand, the usual central limit theorems and associated limit results for $I(0)$ data do not apply when the data series are fractional, and the standard t and F tests are not generally valid. Spurious correlations can be expected, for reasons closely related to the spurious regressions analysis of $I(1)$ series (Phillips 1986). This case has been analysed by Marmol (1996, 1998). On the other hand, the conventional cointegration analysis is equally inapplicable. The usual tests for (non-)cointegration depend on the statistics converging to known functionals of Brownian motion. In the $I(d)$ case with $d \neq 1$, the corresponding functionals are related to the limit processes called fractional Brownian motion. These are distinguished from regular Brownian motion by exhibiting correlated increments (see e.g. Davidson and de Jong, 2000) and the usual tabulations don't apply. Thus, neither of the conventional approaches to testing for a time series relationship is available. Moreover, the statistics are asymptotically non-pivotal, meaning that their limiting distributions depend on d and other nuisance parameters. When d is arbitrary there is no prospect of tabulating them for general use, even as first-order asymptotic approximations.³

In these circumstances, simulation is a natural approach to the testing problem, and this paper investigates the use of parametric bootstrap tests. Section 2 sets the scene by introducing a modelling framework for fractional cointegration. Section 3 discusses the issues relating to the design and implementation of bootstrap tests. Section 4 then introduces the test statistics to be used in this framework. It derives the limiting null distributions and consistency properties of the Durbin Watson and F statistics, in this context. Section 5 applies the methods to a well-known data set, UK nondurable consumption and personal disposable income. The properties of the tests are studied by Monte Carlo experiments, and while these results are specific to the empirical application, such evaluations should be viewed as a routine feature of the bootstrap approach. Section 6 contains concluding remarks. Additional technical results are given in the Appendix.

2 Fractional Cointegration

2.1 Fractionally Integrated Processes

Following Granger (1986), a set of $I(d)$ variables are said to be cointegrated, or $CI(d, b)$, if there exists a linear combination that is $I(d - b)$ for $b > 0$. In the usual applications of the concept, $d = b = 1$, but even with integer values of d and b , it is not difficult to envisage generalisations. The multicointegration model of Granger and Lee (1990), for example, can be fitted into this framework by formally admitting the integrals of $I(1)$ variables to the data set. However, $I(2)$ is a

²Be careful to distinguish the case of 'fractional cointegration' where the data are fractionally integrated, from the case where they are $I(1)$ but have a fractionally integrated linear combination. While we do not address the latter problem specifically because the testing problem can in principle be treated as a standard one, it is an important special case of the setup examined here. See for example Dittmann (2000).

³Breitung and Hassler (2002) show how to construct asymptotically pivotal tests of non-cointegration and cointegrating rank, by fractionally differencing the data, so that the hypothesis becomes one of $I(0)$ against $I(-b)$ for $b > 0$. Their statistic is chi-squared under H_0 , and so can also be viewed as an alternative to the conventional cointegrating rank tests, using standard asymptotics. Their approach still depends on estimation of d , and a comparison with the present approach is an interesting topic for future work.

relatively unusual feature of macroeconomic series. A more commonly arising issue is for data to have bounds on their range of variation (e.g. interest or employment rates) in spite of exhibiting persistent local trends. Such series may be modelled as $I(d)$ where d is a real number between 0 and 1. Alternatively, a process might have long memory albeit stationary increments, and hence be $I(d)$ for $1 < d < 3/2$. Modelling dynamic relationships between processes of this type requires a generalization of the conventional cointegration setup.

2.2 Fractional VECMs

Granger (1986) suggests generalising the vector ECM model by an equation that can be recast in the form⁴

$$\left[\mathbf{B}(L) + ((1-L)^{-b} - 1)\boldsymbol{\alpha}\boldsymbol{\beta}' \right] (1-L)^d \mathbf{x}_t = \mathbf{D}(L)\boldsymbol{\varepsilon}_t \quad (2.1)$$

where \mathbf{x}_t and $\boldsymbol{\varepsilon}_t$ are $N \times 1$, $\boldsymbol{\varepsilon}_t \sim i.i.d.(\mathbf{0}, \boldsymbol{\Sigma})$, $\mathbf{B}(L)$ and $\mathbf{D}(L)$ are finite-order matrix polynomials in the lag operator with all roots outside the unit circle. The constant matrices $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $N \times r$, having rank r , representing the error correction and cointegrating coefficients respectively. Note that the zero-order term of the expansion of $(1-L)^{-b}$ is unity, and therefore model (2.1) has the conventional error-correction representation such that current-dated values of the variables occur only in the leading term. However, also note that the condition

$$(1-L)^{d-b}\boldsymbol{\beta}'\mathbf{x}_t \sim I(0) \quad (2.2)$$

is required so that the equation balances, having both sides $I(0)$. Setting $d = b = 1$ yields the usual Johansen (1988, 1991) style VECM, but d and b can be real values with $d > 0$ and $0 < b \leq d$. In this model, all the elements of \mathbf{x}_t exhibit the same order of integration, not necessarily unity, and similarly the cointegrating residuals $\boldsymbol{\beta}'\mathbf{x}_t$ are all of order $d - b$. Duecker and Startz (1998), for example, estimate a bivariate model of US and Canadian interest rates having this general form.

Once the possibility of fractional cointegration has been admitted, however, a number of difficult issues arise. If d has no special value, there is no logical necessity for the assumption that d is the same for all the series under consideration, especially when the empirical facts do not support this restriction. Therefore, a dynamic modelling framework that can formally admit the generation of series with different d values needs to be established. Consider the following generalisation. Letting

$$\boldsymbol{\Delta}(L) = \text{diag}\{(1-L)^{d_1}, \dots, (1-L)^{d_N}\} \quad (2.3)$$

where d_1, \dots, d_N are any nonnegative reals, and similarly

$$\mathbf{K}(L) = \text{diag}\{(1-L)^{b_1}, \dots, (1-L)^{b_N}\} \quad (2.4)$$

where $0 \leq b_i \leq d_i$, write

$$\left[\mathbf{B}(L) + \boldsymbol{\alpha}\boldsymbol{\beta}'(\mathbf{K}(L)^{-1} - \mathbf{I}) \right] \boldsymbol{\Delta}(L)\mathbf{x}_t = \mathbf{D}(L)\boldsymbol{\varepsilon}_t. \quad (2.5)$$

Here, there is an understanding that the system generates N series integrated to orders d_1, \dots, d_N , such that

$$\boldsymbol{\Delta}(L)\mathbf{x}_t = \mathbf{w}_t \sim I(0) \quad (2.6)$$

(defining \mathbf{w}_t). To balance the equation, it is required that

$$\boldsymbol{\beta}'\mathbf{K}(L)^{-1}\mathbf{w}_t \sim I(0) \quad (2.7)$$

⁴See Granger (1986), equation (4.3). We adopt here the more standard VECM notation, similar to Johansen (1988, 1991).

and in this sense, the system exhibits cointegration. Letting $b_i = b$ and $d_i = d$ for all i , it reproduces (2.1). We refer to this class as fractional vector error correction models, or FVECMs.

The FVECM provides a very general modelling framework, but the case closest to the standard VECM is that where $d_i - b_i = a \geq 0$ for all i , such that

$$\mathbf{K}(L)^{-1}\mathbf{\Delta}(L) = (1 - L)^a \mathbf{I}.$$

For clarity, and without loss of generality, consider the case $a = 0$. The question of interest is whether this model has the usual property that

$$\boldsymbol{\beta}'\mathbf{x}_t \sim \text{I}(0). \quad (2.8)$$

It is clear that to have this property, the model requires further restrictions. Let the variables be indexed in decreasing order of magnitude of d_i . If $d_1 > d_2$, then (2.8) can hold only if the first row of $\boldsymbol{\beta}$ is zero, so that x_{1t} does not appear in any cointegrating relation. However, also note that provided $d_1 = d_2$ there is no reason why variables integrated to lower order should not appear in the same cointegrating relation. To fix ideas, let $N = 3$, and $r = 1$. It is clearly possible that

$$\beta_1 x_{1t} + \beta_2 x_{2t} \sim \text{I}(d_3) \quad (2.9)$$

for $d_3 < d_1$, and hence if $x_{3t} \sim \text{I}(d_3)$ that

$$\beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} \sim \text{I}(0) \quad (2.10)$$

A necessary condition for cointegration is that the two largest orders of integration of variables in the relation are equal. However, no other restrictions are implied.

In the remainder of the paper, we develop simple regression-based tests for the existence of such relationships. For the sake of simplicity, the asymptotic analysis is carried out on models of the form (2.1), with the same d assumed for all the variables, but the generalization to (2.5), subject to the cited restriction, represents a relatively minor extension apart from the notational overhead. A variant of this type is examined in Davidson (2000b).

3 Bootstrap Tests for Non-cointegration

The bootstrap technique in econometrics can play two distinct roles: to improve the first-order approximations of asymptotically pivotal statistics, and to yield first-order asymptotic approximations in cases where tests are not asymptotically pivotal. The second of these roles is played in the present application. The application of bootstrap methods to time series data is discussed in, for example, Shao and Tu (1995), Li and Maddala (1996, 1997) and Horowitz (1997, 2000). The block resampling method is sometimes advocated for capturing dependence in a time series application, but that requires stationarity. The approach to be applied here is the parametric bootstrap, otherwise known as the recursive bootstrap. For an application of a similar method, to a test of restrictions on the long-run parameters in a standard VECM, see Fachin (2000).

3.1 The Parametric Bootstrap

In the test of fractional cointegration, the bootstrap test entails the following steps.

1. ARFIMA(p, d, q) processes are fitted to the variables in the model.
2. The series are fractionally differenced according to the estimates from step 1, to estimate \mathbf{w}_t in (2.6), and a dynamic model of the differences is constructed. The object is to estimate model (2.1) or (2.5) subject to the restrictions of the null hypothesis, $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$. There are several different approaches to this step, to be detailed in Section 3.4 below.

3. The residuals from stage 2, estimating ε_t , are re-sampled with replacement and used to generate series according to (2.1) or (2.5), under H_0 . Any suitable statistics to test for a cointegrating relationship can be computed from these. The values of these statistics in the observed data are located in the bootstrap distributions to yield an estimated p -value.

These methods should yield valid first-order approximations to the true p -values. The application of the bootstrap to long memory models has not been much researched to date, although a recent application to tests of integration order (not cointegration) is Andersson and Gredenhoff (1998). Some well-known issues point to the need for caution. Basawa et. al. (1991a,b) show that bootstrap tests for unit roots, with the stable AR(1) as alternative, are inconsistent. The problem arises because of the well-known discontinuity in the family of distributions of statistics based on AR processes, at the point where the root is unity. Here, the order of magnitude of the error-of-estimate jumps from $n^{-1/2}$ to n^{-1} , and bootstrapping with an estimated autoregressive root may place it the wrong side of the discontinuity.

However, unlike the case of the AR coefficient there is no discontinuity in the distributions of statistics based on $I(d)$ processes, either at the point $d = 1$ or elsewhere.⁵ Dufour (2000, Proposition 6.3) demonstrates the asymptotic validity of bootstrap p -values based on asymptotically non-pivotal statistics, under an equicontinuity condition on the family of distribution functions. The distributions of our statistics should satisfy these conditions everywhere. There are no known reasons to doubt the asymptotic validity of bootstrap testing of the value of d , although a number of important issues of test design remain to be considered.

3.2 Treatment of the Nuisance Parameters

Dufour (2000) points out that exact α -level tests can be performed by Monte Carlo, even in the presence of nuisance parameters under the null, if these parameters are treated appropriately. Denoting these parameters by $\theta \in \Omega_0$ where Ω_0 denotes the region of the parameter space corresponding to H_0 , a test of level α is performed, by definition, by rejecting only if

$$\sup_{\theta \in \Omega_0} G(\theta) \leq \alpha \tag{3.1}$$

where G represents the p -value, or probability of the test statistic falling in the critical region under H_0 . In other words, the unknown parameters should be replaced by their ‘worst-case’ values. This implies estimating them by maximizers of G , according to the criterion in (3.1). Dufour calls these maximised Monte Carlo (MMC) tests.

By contrast, the bootstrap method can be identified with a Monte Carlo test in which θ is replaced by a consistent estimate, and these tests are only asymptotically correct α -level tests. The asymptotic approximation of the raw bootstrap can be poor, although it can be much improved by methods such as pre-pivoting (Beran 1988). It is however worth remarking that the hazard due to approximation error is asymmetric. In the event of rejection at the chosen level, one would wish to know whether the result might be reversed by a different choice of θ , yielding a larger G and hence a more conservative test. If the hypotheses of interest are not rejected at the chosen significance level, the test outcome is the same as the MMC test outcome. The bootstrap approximation can produce, at worst, ‘false positives’ at the α level.

⁵The normalized partial sums switch to a.s. continuity in the limit at $d = 1/2$, but note that the distributions of functionals of these processes do not experience a discontinuous shift comparable to the AR-unit-root model. For example, the OLS estimator with $I(d)$ regressors and $I(0)$ disturbances converges at $n^{1/2}$ (independent of d) for $d < 1/2$, and at rate n^d for $d \geq 1/2$ (see Davidson 2001b).

3.3 ARFIMA estimation

Successful implementation of these methods depends on being able to compute good estimates of the fractional integration parameters. Maximum likelihood estimation in both the time domain and the frequency domain has been extensively studied, see Sowell (1992), Cheung and Diebold (1994), Hauser (1999) among other references. The Whittle frequency domain likelihood (see Fox and Taqqu 1986, Velasco and Robinson 2000) has been adopted for this study. It is simple to compute and has been found both fast and reliable, generally locating a point close to the Sowell (1992) time domain estimate when both are computed.

Model selection is guided by the Schwarz (1978) criterion supplemented by a measure of residual autocorrelation. The ARFIMA model has generally good characteristics in samples of reasonable size, with well determined estimates of the fractional integration parameter d that are not too sensitive to the choice of ARMA specification. The well-known exception to this generalization is a tendency for the ARMA(p, q) with an autoregressive root near unity to mimic the ARFIMA(p, d, q) with d close to unity. In certain cases, adding an autoregressive term can dramatically reduce the estimate of d . This is an identification problem, the ARIMA($p, 1, q$) being nested within both the ARMA($p + 1, q$) and the ARFIMA(p, d, q), and suppressing the extra AR term may resolve it.

Experimenting with a range of different macroeconomic series of lengths in the 100-500 range, two broad lessons emerge. First, it is notable how often the ‘pure-fractional’ case, with $p = q = 0$, provides an adequate representation, in terms of both maximizing the Schwarz criterion and showing no significant residual autocorrelation. ARFIMAs are frequently more parsimonious in this sense than their ARIMA counterparts. Second, some series are much more amenable to fractional modelling than others, for no reason that is immediately apparent from eyeballing the time plot. By ‘amenable’ may be meant either the existence of an adequate low-order specification, the robustness of the d -estimate to alternative ARMA specifications, or both of these features. Possibly, these differences can be attributed to the presence of nonlinear dependence features that can be handled with varying degrees of success by the ARFIMA class in different realizations. This is an issue that merits further study.

3.4 Representation of the Null Hypothesis

Setting $b = 0$ in (2.1), or $\mathbf{K}(L) = \mathbf{I}$ in (2.5), gives

$$\mathbf{B}(L)\mathbf{w}_t = \mathbf{D}(L)\boldsymbol{\varepsilon}_t \quad (3.2)$$

where $\mathbf{w}_t \sim I(0)$ is the vector of fractional differences. Step 2 of the bootstrap procedure entails, in principle, estimating this model assuming H_0 to be true, subject to the caveats on choice of specification discussed below. Fractional differencing must be approximated in finite samples since it involves lags of infinite order. After integer differencing as necessary to give $|d| < 1/2$, this is performed using the approximation

$$(1 - L)^d x_t \approx \sum_{j=0}^{t-1} b_j x_{t-j} \quad t = 1, \dots, n \quad (3.3)$$

where $b_0 = 1$ and

$$b_j = \frac{-d\Gamma(j-d)}{\Gamma(1-d)\Gamma(j+1)} \quad (3.4)$$

for $j > 0$. These coefficients are easily computed by the recursion $b_j = (j - d - 1)b_{j-1}/j$.

The bootstrap replications could then, in principle, be performed by randomly resampling the residuals $\hat{\boldsymbol{\varepsilon}}_t$ (say), and passing the series so generated back through the filter (3.2) to yield

a resampled series $\{\mathbf{w}_t^*\}_1^n = \mathbf{w}_1^*, \dots, \mathbf{w}_n^*$. This would be integrated back to get the resampled series $\{\mathbf{x}_t^*\}_1^n$, non-cointegrated by construction but having, asymptotically, the same short-run dynamics as the observed data.

For a model involving several variables, such a procedure threatens to be computationally burdensome. However, the tests considered in this paper are simple regression-based statistics, implying the partition $\mathbf{x}_t = (y_t, \mathbf{z}_t)'$ where y_t is the scalar ‘explained’ variable, and \mathbf{z}_t the explanatory variables. An important feature of the setup is that the whole procedure may be conditioned on \mathbf{z}_t . In other words, the observed regressors are used in the bootstrap replications, and y_t alone is resampled. In this case, the VAR in \mathbf{w}_t should be replaced by a dynamic regression, used to generate a residual series with the schematic form

$$\varepsilon_{1t} = \Delta^d y_t - E(\Delta^d y_t | \dots \Delta^d \mathbf{z}_{t+1}, \Delta^d \mathbf{z}_t, \Delta^d \mathbf{z}_{t-1}, \Delta^d y_{t-1}, \dots). \quad (3.5)$$

Note the inclusion in this regression of both leads and lags of the conditionally fixed variables. The null hypothesis does not rule out the possibility that $\Delta^d y_t$ may Granger-cause $\Delta^d \mathbf{z}_t$, and it is necessary to purge the residuals of this potential dependence.⁶ Once this is done, however, they can be randomly resampled and back-filtered through the fitted form of (3.5) to generate the bootstrap series $\{\Delta^d y_t\}_1^n$. Asymptotically, these series exhibit the same pattern of cross-autocorrelation with the observed $\{\Delta^d \mathbf{z}_t\}_1^n$ as the original series. Finally, re-integrating yields a resampled series $\{y_t\}_1^n$ conforming to the null hypothesis, such that $\Delta^d y_t$ is generally dynamically correlated with $\Delta^d \mathbf{z}_t$, but (y_t, \mathbf{z}_t) is not a cointegrated set. This approach has clear advantages over resampling the whole data set. It is computationally much more tractable, and should also yield a more powerful test, since the conditional null distributions of the test statistics have lower dispersion, in general, than the unconditional distributions.⁷

It should be emphasised that the dynamic modelling exercise at Step 2 is not intended to capture the DGP of the observed data. Its object is to allow the correlation structure of $\{\Delta^d y_t, \Delta^d \mathbf{z}_t\}$ *under the null hypothesis* to be recreated from the resampled residuals. There is a major penalty to be incurred by overfitting it. This is best illustrated by considering a leading case of the alternative hypothesis, say

$$y_t = \gamma' \mathbf{z}_t + u_t \quad u_t \sim \text{i.i.d., independent of } \mathbf{z}_t. \quad (3.6)$$

Clearly, in this case ε_{1t} in (3.5) corresponds to u_t . For clarity let $d = 1$, and note that if (3.6) is true the criterion of choosing the best-fitting representation *in the observed data* leads inevitably to

$$\varepsilon_{1t} = \Delta y_t + \Delta y_{t-1} + \Delta y_{t-2} + \dots - \gamma' \Delta \mathbf{z}_t - \gamma' \Delta \mathbf{z}_{t-1} - \gamma' \Delta \mathbf{z}_{t-2} - \dots$$

Obviously, backsolving this equation generates y_t not Δy_t . Note that the test based on the best-fitting representation is asymptotically correctly *sized*, since if H_0 is true the appropriate model is chosen, based on the data. However, this test would have little power against alternatives such as (3.6), since the short-run dynamics would tend to be contaminated by the long-run dynamics in that case. There is a balance of advantages here, and the optimal choice of lag length will depend on sample size, but as a rule of thumb, one or two lags and leads should often suffice to represent the dynamics adequately if H_0 is true, yielding a correctly sized asymptotic test.

⁶Compare the analyses of e.g. Phillips and Loretan (1991), Saikkonen (1991) and Stock and Watson (1993). These authors point out that the inclusion of both leads and lags in a least squares cointegrating regression is necessary to purge short-run dynamic effects. The present procedure implements this strategy under the null hypothesis of non-cointegration, with the specific object of rendering the residuals orthogonal to the conditionally fixed variables in the bootstrap.

⁷Although the full VAR specification of (2.5) is only implicit in this framework, it plays the role of formally justifying the linear modelling of the short-run dynamics in (3.5).

This approach also points to the option of testing stronger hypotheses than just non-cointegration. Thus, if $\Delta^d z_t$ and its leads and lags are omitted from the conditioning set in (3.5), which is equivalent to modelling y_t solely using its univariate ARFIMA representation, this corresponds to the null of statistical independence between y_t and z_t . The additional restrictions can be viewed, in terms of (2.1) or (2.5), as corresponding to the block diagonality of $\mathbf{B}(L)$, $\mathbf{D}(L)$ and $\mathbf{\Sigma}$ in respect of their first row/column. Below we call this the strong null, in contrast to the weak null of simple noncointegration. Further, by suppressing the leads but not the lags from (3.5), a ‘semi-weak’ null can be specified, under which y_t is neither cointegrated with nor Granger-causes z_t . Testing each of this hierarchy of non-dependence hypotheses may be a useful exercise, since if one can be rejected and not another, this can provide a new view of the evidence for a relationship contained in the sample.

4 The Choice of Test

4.1 Residual-based Tests

In this section, we assume for simplicity that the DGP has the form (2.1). Consider the regression

$$y_t = \boldsymbol{\gamma}' z_t + u_t \quad (4.1)$$

where $\mathbf{x}_t = (y_t, z_t)'$ ($(k+1) \times 1$) is an $I(d)$ vector, for $d > 1/2$.⁸ \mathbf{x}_t is a nonstationary fractional process that can be thought of as possibly a subvector from a model with structure (2.1). It is assumed to satisfy the weak convergence property,⁹

$$n^{1/2-d} \mathbf{x}_{[n\delta]} \xrightarrow{d} \mathbf{X}(\delta) \quad 0 \leq \delta \leq 1$$

where $\mathbf{X} = (Y, \mathbf{Z})$ is a fractional Brownian motion with parameter d .¹⁰ The null hypothesis under test is $H_0 : b = 0$, implying that $u_t \sim I(d)$ for any choice of $\boldsymbol{\gamma}$. Under the alternative hypothesis $b > 0$, there exists $\boldsymbol{\gamma}$ such that $u_t \sim I(d-b)$.

Let $\boldsymbol{\Psi}$ ($(k+1) \times (k+1)$) represent the covariance matrix of \mathbf{X} , defined by

$$E\mathbf{X}(\delta)\mathbf{X}(\delta)' = \delta^{2d-1}\boldsymbol{\Psi} \quad 0 \leq \delta \leq 1. \quad (4.2)$$

It is shown in the Appendix that when H_0 is true, $\boldsymbol{\Psi}$ has full rank, and if $\hat{\boldsymbol{\gamma}}$ is the OLS estimator of $\boldsymbol{\gamma}$ then

$$\hat{\boldsymbol{\gamma}} \xrightarrow{d} \boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21} + \boldsymbol{\zeta} \quad (4.3)$$

where $\boldsymbol{\Psi}_{22}$ ($k \times k$) and $\boldsymbol{\Psi}_{21}$ ($k \times 1$) are derived from the obvious partition of $\boldsymbol{\Psi}$, and $\boldsymbol{\zeta}$ is a zero-mean random vector. On the other hand, when $b > 0$, $\boldsymbol{\Psi}$ is singular, the vector $(1, -\boldsymbol{\gamma}')$ lies in its null space, and it can be shown (see Davidson 2001b, Table 1) that

$$\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} = O_p(n^{\max\{-b, 1-2d\}}). \quad (4.4)$$

With the exception of Duecker and Startz (1998) who estimate the FVECM directly by maximum likelihood, the applied studies referenced in the introduction proceed by estimating d from the OLS regression residuals. Except for Baillie and Bollerslev (1994), who use approximate

⁸Note that the case $d = 1/2$ is nonstationary but requires special treatment, and is not considered here.

⁹ \xrightarrow{d} denotes convergence in distribution.

¹⁰See Davidson and de Jong (2000) for details. Note that this result establishes the limit as fractional Brownian motion of type I, in the typology of Marinucci and Robinson (1999).

maximum likelihood but do not attempt to test for non-cointegration, these studies base their tests on the Geweke and Porter-Hudak (1983) (GPH) estimator of $d - b$, applied to the OLS residuals.

If the least squares residuals are

$$\hat{u}_t = u_t + (\hat{\gamma} - \gamma)' \mathbf{z}_t \quad (4.5)$$

where $u_t \sim \text{I}(d-b)$ and $\mathbf{z}_t \sim \text{I}(d)$, consider the cases $d-b < 1/2$ and $d-b > 1/2$.¹¹ In the former case, note that with $d > 1/2$,

$$(\hat{\gamma} - \gamma)' \mathbf{z}_t = O_p(n^{\max\{1/2-d, d-b-1/2\}}) = o_p(1)$$

and hence, the residuals can proxy for the true disturbances asymptotically. On the other hand, if $d-b > 1/2$ then u_t is nonstationary but satisfies a functional central limit theorem for fractional processes (see Davidson and de Jong 2000). Letting $\boldsymbol{\xi}$ ($k \times 1$) denote the limit in distribution of $n^{\min\{b, 2d-1\}}(\hat{\gamma} - \gamma)$ we have the weak convergence

$$n^{1/2-d+b} \hat{u}_{[n\delta]} \xrightarrow{d} U(\delta) + \boldsymbol{\xi}' \mathbf{Z}(\delta) \quad 0 \leq \delta \leq 1 \quad (4.6)$$

where U is a fractional Brownian motion of order $d-b$, so interestingly enough, the residual process does not share the memory characteristics of the true disturbances, even in the limit. However,

$$\Delta \hat{u}_t = \Delta u_t + (\hat{\gamma} - \gamma)' \Delta \mathbf{z}_t \quad (4.7)$$

where the first right-hand side term is $O_p(1)$ and the second is $O_p(n^{-b})$, provided $d < 3/2$. Therefore, as shown by Hassler, Marmol and Velasco (2000), log-periodogram regression can be based either on \hat{u}_t to estimate $d-b$, or on $\Delta \hat{u}_t$ to estimate $d-b-1$, depending on which of these quantities falls in $(-1/2, 1/2)$. In Section 5 we consider as a candidate for bootstrapping a procedure closely related to the residual GPH test, based on the Robinson (1994) nonparametric estimator of d .

4.2 The Dickey-Fuller/Durbin-Watson Approach

Other candidate test statistics based on the regression residuals are the (unaugmented) Dickey-Fuller (DF) and Durbin-Watson (DW) statistics. In fact, tests based on $\text{DF} = n\hat{\phi}$ and on $\text{DW} \approx -2\hat{\phi}$ where

$$\hat{\phi} = \frac{\sum_{t=2}^n \hat{u}_{t-1} \Delta \hat{u}_t}{\sum_{t=2}^n \hat{u}_{t-1}^2} \quad (4.8)$$

are virtually equivalent from the point of view of computing bootstrap p -values. Note that the weak-dependence correction (augmentation) of the DF is inappropriate when $\Delta \hat{u}_t$ is long memory, and these statistics are not asymptotically pivotal in any case, due to dependence on d . In a bootstrap exercise, either form of the statistic, or $\hat{\phi}$ itself, will serve equally well. In the example of Section 5 the DW is cited for no better reason than its familiarity.

Letting $\hat{\boldsymbol{\theta}} = (1, -\hat{\gamma})$, note that

$$\hat{\phi} = \frac{\hat{\boldsymbol{\theta}}' \sum_{t=2}^n \mathbf{x}_{t-1} \Delta \mathbf{x}_t \hat{\boldsymbol{\theta}}}{\hat{\boldsymbol{\theta}}' \sum_{t=2}^n \mathbf{x}_{t-1} \mathbf{x}_{t-1}' \hat{\boldsymbol{\theta}}} \quad (4.9)$$

¹¹As noted above, the case $d-b = 1/2$ requires separate treatment.

Assume H_0 is true. It is shown in the Appendix that for $1/2 < d < 3/2$, and subject to the additional assumption $E(\varepsilon_{jt}^4) < \infty$ for $j = 1, \dots, k+1$ in the case $d > 1$, that

$$n^{\min\{1, 2d-1\}} \hat{\phi} \xrightarrow{d} \frac{\boldsymbol{\theta}' \mathbf{J} \boldsymbol{\theta}}{\boldsymbol{\theta}' \mathbf{H} \boldsymbol{\theta}} \quad (4.10)$$

where $\boldsymbol{\theta} = (1, -(\boldsymbol{\Psi}_{22}^{-1} \boldsymbol{\Psi}_{21} + \boldsymbol{\zeta})')'$, $\mathbf{H} = \int_0^1 \mathbf{X} \mathbf{X}' ds$, and

$$\mathbf{J} = \begin{cases} \frac{1}{\Gamma(d-1)} \int_0^1 \xi^{d-2} \int_0^{1-\xi} \mathbf{X} d\mathbf{W}' d\xi + \boldsymbol{\Lambda}, & d > 1 \\ \int_0^1 \mathbf{X} d\mathbf{X}' + \boldsymbol{\Lambda}^*, & d = 1 \\ \boldsymbol{\Lambda}^*, & d < 1. \end{cases} \quad (4.11)$$

Thus, there are three distinct cases to consider. In the formula corresponding to $d > 1$, \mathbf{W} is a *regular* Brownian motion defined by the weak convergence

$$n^{-1/2} \sum_{t=1}^{[n\delta]} \mathbf{w}_t \xrightarrow{d} \mathbf{W}(\delta) \quad 0 \leq \delta \leq 1 \quad (4.12)$$

where $\mathbf{w}_t = (1-L)^d \mathbf{x}_t$, and

$$\boldsymbol{\Lambda} = \lim_{n \rightarrow \infty} n^{1-2d} \sum_{t=2}^n \sum_{j=1}^{t-1} E(\Delta \mathbf{x}_j \Delta \mathbf{x}_t'). \quad (4.13)$$

This matrix satisfies the relation

$$\boldsymbol{\Psi} = \boldsymbol{\Lambda} + \boldsymbol{\Lambda}' \quad (4.14)$$

since when $\Delta \mathbf{x}_t$ is a long-memory process, its contemporaneous covariances are asymptotically negligible relative to the summed autocovariances.

The case $d = 1$ is the ‘standard’ result, in which $\mathbf{X} = \mathbf{W}$. In this case the decomposition

$$\boldsymbol{\Psi} = \boldsymbol{\Sigma} + \boldsymbol{\Lambda}^* + \boldsymbol{\Lambda}^{*'} \quad (4.15)$$

holds where $\boldsymbol{\Sigma} = E(\Delta \mathbf{x}_t \Delta \mathbf{x}_t')$, and $\boldsymbol{\Lambda}^* = \sum_{j=1}^{\infty} E(\Delta \mathbf{x}_{t-j} \Delta \mathbf{x}_t') < \infty$.

However, if $d < 1$ then

$$\boldsymbol{\Sigma} + \boldsymbol{\Lambda}^* + \boldsymbol{\Lambda}^{*'} = \mathbf{0}. \quad (4.16)$$

In this case the stochastic integral (the random component of \mathbf{J}) is of smaller order of magnitude than n . $\boldsymbol{\Psi}$ is still defined by (4.2), but it does not have a decomposition either of the form (4.14) or (4.15). Note that $\boldsymbol{\Lambda}$ and $\boldsymbol{\Lambda}^*$ are different matrices, being different functions of d and the other model parameters as well as having different normalizations.

Since these distributions depend both on d and on the parameters of the short-run dynamics in (2.1), the object of modelling the fractional differences in the simulation procedure described in Section 3.4 can be seen as ensuring that the bootstrap distribution of $\hat{\phi}$ is that of (4.10), to a first-order approximation. The case $\boldsymbol{\Psi}_{21} = \mathbf{0}$, such that the distribution of $\hat{\gamma}$ has zero mean in the limit, corresponds to the ‘strong null’ hypothesis of independence between y_t and \mathbf{z}_t .

The test based on this distribution is consistent against alternatives in which u_t is a mean reverting process. It is shown in the Appendix that with $b > 0$,

$$n^{\min\{1, 2d-1\}} |\hat{\phi}| = O_p(n^\theta) \quad (4.17)$$

where¹²

$$\theta = \begin{cases} \begin{cases} 1, & 0 < d - b < 1/2 \\ 2(b - d + 1), & 1/2 < d - b < 1 \\ 0, & 1 < d - b < d \end{cases}, & d > 1 \\ \begin{cases} 2d - 1, & 0 < d - b < 1/2 \\ 2b, & 1/2 < d - b < d \end{cases}, & d \leq 1. \end{cases} \quad (4.18)$$

4.3 The Goodness of Fit Test

The third test proposed is based on the usual F statistic for the joint ‘significance’ of the regressors. This is the statistic whose asymptotic form is¹³

$$F = n \frac{\sum_{t=1}^n y_t^2 - \sum_{t=1}^n \hat{u}_t^2}{\sum_{t=1}^n \hat{u}_t^2}. \quad (4.19)$$

Like the GPH and DW/DF, the statistic is not asymptotically pivotal, but since it is based on the goodness of fit of the regression, and not merely on the properties of the residuals, it promises additional sensitivity in detecting a relationship. The main problem is to tabulate the distribution under the null hypothesis. The ‘spurious regression’ phenomenon implies that $F = O_p(n)$ under the null hypothesis $b = 0$. It is shown in the Appendix that on H_0 ,

$$\frac{F}{n} \xrightarrow{d} \frac{\left(\Psi_{22}^{-1/2} \Psi_{21} + \sqrt{\psi_{11.2}} \mathbf{G}_{22}^{-1} \mathbf{G}_{21} \right)' \mathbf{G}_{22} \left(\Psi_{22}^{-1/2} \Psi_{21} + \sqrt{\psi_{11.2}} \mathbf{G}_{22}^{-1} \mathbf{G}_{21} \right)}{\psi_{11.2} (g_{11} - \mathbf{G}_{12} \mathbf{G}_{22}^{-1} \mathbf{G}_{21})}. \quad (4.20)$$

Here, the \mathbf{G}_{ij} are the partitioned blocks of a full-rank matrix $\mathbf{G} = \int_0^1 \mathbf{W}_d \mathbf{W}_d' dr$ where \mathbf{W}_d is a $(k+1)$ -vector of independent standard fractional Brownian motions, and the scalar constant $\psi_{11.2}$ is positive under H_0 but zero under the alternative. It is easy to show that under the alternative hypothesis $b > 0$, $F/n = O_p(n^\theta)$ where

$$\theta = \begin{cases} 2d - 1, & 0 < d - b < 1/2 \\ 2b, & 1/2 < d - b < d \end{cases}. \quad (4.21)$$

Comparing this with the divergence rates of the DW test in (4.18), note that the test has power even against alternatives where $d > 1$ and the residuals are non-mean reverting. This suggests that the F test may perform better against nonstationary alternatives than the DW test. In this case, the resampling procedures described in Section 3.4 have as their object the reproduction of the distribution in (4.20), to a first-order approximation. Setting $\Psi_{21} = \mathbf{0}$ in (4.20) gives the ‘strong null’ distribution, and note that increasing the absolute magnitude of the covariances shifts the distribution in (4.20) to the right. This suggests that in tests of the weak null, the F test may show a corresponding diminution of test power against local alternatives.

¹²The case $d - b = 1/2$ will typically involve a slowly varying sequence, converging slower than $n^{-\varepsilon}$ for any $\varepsilon > 0$.

¹³In practice, y_t and z_t in this formula may be regarded as deviations from fitted means and deterministic trends, so that the statistic tests for a relationship net of these effects. The asymptotic distribution then depends on functionals of demeaned or detrended fractional Brownian motion, but is otherwise unchanged. For clarity of exposition these corrections are not shown explicitly.

4.4 Relation to existing tests

These methods can of course be applied to the case of regular cointegration, corresponding to $d = 1$. It is therefore desirable to say how they relate to the existing tests of cointegration. The following points may be noted.

1. In conventional tests of cointegration, an asymptotically pivotal statistic is compared with a tabulated distribution. This is achieved in general by replacing unknown parameters by consistent estimates. In the present tests, unknown parameters are also estimated, but in this case are used to construct a DGP to simulate the distribution of the data under the null. These two approaches to first-order approximation may have different characteristics in given samples, and it is of interest to compare them
2. Although the tests described employ a least squares regression, they are not directly comparable with the Engle-Granger (1987)-type test of cointegration, based on the regression residuals. Being based implicitly on modelling of the full system dynamics, they are not susceptible to the critique of the latter tests by Kremers et. al (1992), that the dynamic corrections are subject to invalid restrictions. Note that the specification in (3.5) embodies joint determination of all the variables, notwithstanding that we condition on the regressors. The choice of regression-based statistics is merely convenient, and other features of the simulated distribution, such as the eigenvalues of the VAR, might also be tabulated.
3. As specified, the test is of the hypothesis of zero cointegrating rank, against the alternative of positive rank. However, note the fact that in a collection of variables of cointegrating rank 2, one of them can always be dropped, while retaining cointegration amongst the remaining set (compare Davidson 1998, Theorem 2). Therefore, in the event that the null hypothesis of zero rank is rejected, a test of rank 1 against the alternative of higher rank could be implemented (taking relevance of the ‘dependent’ variable as a maintained hypothesis) by dropping regressors in turn. The natural statistic to bootstrap in this case would be the maximum of the test statistics over the possible deletions. The same procedure might be iterated for tests of still higher cointegrating rank.
4. A major advantage over the usual tests is that, even if the ‘explained’ variable is $I(1)$, the regressors can be validly integrated to any order compatible with a cointegrating relationship, including $I(0)$. By contrast, all the variables in the Engle-Granger regression, or the Johansen VAR, must be $I(1)$ to validate the usual tabulations. Since we condition on them, the distribution of the regressors is arbitrary. Even cointegration amongst the regressors is allowable in principle, although note that for simplicity, the asymptotic analysis given here rules out this possibility. The test of higher cointegrating rank described in 3 above could distinguish this case.
5. The ability to test for relationships in the weak and strong senses defined in Section 3.4 may yield valuable extra information. In the former case, the test determines whether cointegration exists after correlation between the changes in the series has been taken into account. As noted, the Engle-Granger type of test corrects for such correlations in a fairly crude manner.
6. That the goodness of fit test may prove a more powerful test of cointegration than the purely residual-based tests has been argued above, and is confirmed experimentally in the examples examined in the next section. These may prove a useful addition to the test armoury, once the difficulty of deriving asymptotically pivotal versions has been circumvented by the simulation approach.

5 Application: UK Consumption

5.1 Cointegration Tests

It is useful to look the performance of these procedures with a familiar data set, albeit one that, as it turns out, is amenable to the conventional testing approach. This permits a direct comparison of methods. One of the best known studies to have been interpreted (retrospectively) in the cointegration framework is that of UK real nondurable consumption and real personal disposable income by Davidson et. al. (1978). The data examined here are the logarithmic, quarterly, seasonally adjusted series for 1955 Quarter 1 to 1996 Quarter 4. The data are shown in Figure 1. After removing the deterministic trend,¹⁴ a parsimonious representation of the log income series y_t is

$$(1 - L)^{0.977}(1 - .123L)(y_t - 10.49 - 0.0065t) = \hat{u}_{1t} \quad (5.1)$$

The residual Box-Pierce Q statistic for this model with 12 lags is 14.55. The log-consumption series c_t , proves more difficult to model with an autoregressive component for the reasons outlined in Section 3.3, but the equation

$$\begin{aligned} (1 - L)^{1.028}(c_t - 10.43 - .0062t) \\ = (1 - 0.069L + 0.141L^2 + .094L^3 - 0.131L^4)\hat{u}_{2t} \end{aligned} \quad (5.2)$$

exhibits a residual Q for 12 lags of 13.54. These are the representations used in the test of cointegration.

There is a strong theoretical presumption that these series should be integrated to the same order, although that this common value is 1 is conventionally tested only with the $I(0)$ autoregressive model as the alternative. The robust standard errors for the fractional coefficients of income and consumption are respectively 0.132 and 0.194, so that the hypotheses of unit ds (and hence of equal ds) are not rejected under fractional alternatives, although the confidence intervals are wide. Whilst in practical applications one might wish to impose the restriction, the object of the present exercise requires us to assume that no d value is ‘special’. Likewise, we should not pre-judge the issue of equality of the coefficients, under the null hypothesis of non-cointegration. Note that (5.1) is not used for bootstrap resampling, although it is used for estimating the regression in differences.

The bootstrap replications for the strong null are performed by randomly resampling the residuals from (5.2), with replacement. For the weak null hypothesis, the bootstrap samples are constructed using the fitted equation

$$\hat{\varepsilon}_{2t} = \hat{u}_{2t} - 0.23\hat{u}_{2,t-1} - 0.16\hat{u}_{1,t+1} - 0.28\hat{u}_{1t} - 0.21\hat{u}_{1,t-1} \quad (5.3)$$

where \hat{u}_{1t} and \hat{u}_{2t} are from (5.1) and (5.2) respectively. In other words, the short-run dynamic model is generalized, as parsimoniously as possible, by specifying the interactions between the individually pre-whitened series. The series $\{\hat{\varepsilon}_{2t}^*\}_1^n$ are obtained by random resampling with replacement from the distribution of $\{\hat{\varepsilon}_{2t}\}_1^n$ obtained from (5.3). The resampled ‘consumption’ series $\{c_t^*\}_1^n$ are then obtained by passing $\hat{\varepsilon}_{2t}^*$ back through (5.3) and (5.2). That is to say, the series \hat{u}_{2t}^* used to obtain c_t^* from (5.2) is generated from

$$\hat{u}_{2t}^* = 0.23\hat{u}_{2,t-1}^* + 0.16\hat{u}_{1,t+1} + 0.28\hat{u}_{1t} + 0.21\hat{u}_{1,t-1} + \hat{\varepsilon}_{2t}^*.$$

¹⁴The detrending is done by preliminary regression, although note that the fractional coefficient is estimated after differencing, so that the trend coefficient corresponds to the mean of the series. In recent work, Velasco and Robinson (2000) report that both nonstationarity and deterministic components can be allowed for in Whittle estimation by suitable trimming of the periodogram, and this approach should be considered for future work.

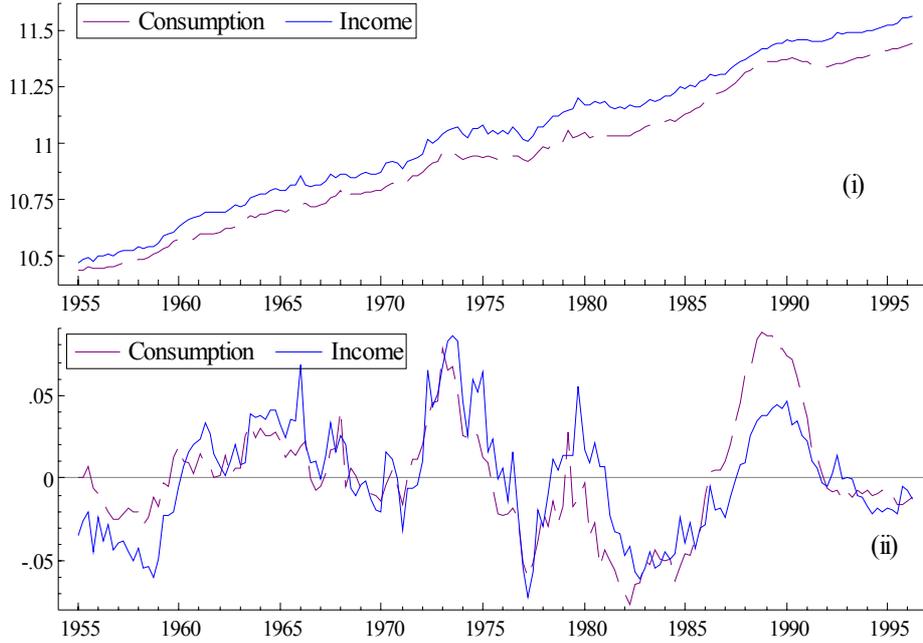


Figure 1: UK log real consumption and log real PDI: (i) Actual (ii) Deviations from linear trend.

	F	DW	res- d	Δ res- d
test statistics	229.6	0.397	0.393	-.608
p -value, strong null	0	0	0.03	0
p -value, semi-weak null (1 lag)	0.03	0.05	0.15	0.07
p -value, weak null (1 lag, 1 lead)	0.13	0.21	0.37	0.39

Table 1: Cointegration Tests, 1955Q1-1996Q2

The procedure for the semi-weak hypothesis is similar, but using the fitted equation

$$\hat{\varepsilon}_{2t} = \hat{u}_{2t} - 0.17\hat{u}_{2,t-1} - 0.27\hat{u}_{1t} - 0.19\hat{u}_{1,t-1}$$

in place of (5.3).

Table 1 shows the results of bootstrap tests using 10,000 bootstrap replications.¹⁵ The F statistic is the test for significance of the regressors excluding constant and trend. Hence, in this case it is just the squared t value for y_t from the fitted cointegrating regression with trend dummy, which is

$$\hat{c}_t = 0.795y_t + 2.0781 + 0.00103t.$$

The last two columns of the table give the Robinson (1994) nonparametric estimator of d for the levels and differences of the residuals, respectively. The latter should have power against nonstationary alternatives, as discussed in Section 4.1.

It has been pointed out that the simple DHSY model is found to predict the UK data well up until 1984, but breaks down thereafter. Since this fact points to a possible shift in the long-run

¹⁵This number of replications represents a degree of overkill, but the experiments run very quickly and were extended to check the experimental error. In practice, 500 or so should give ample experimental precision, relative the other approximation errors involved.

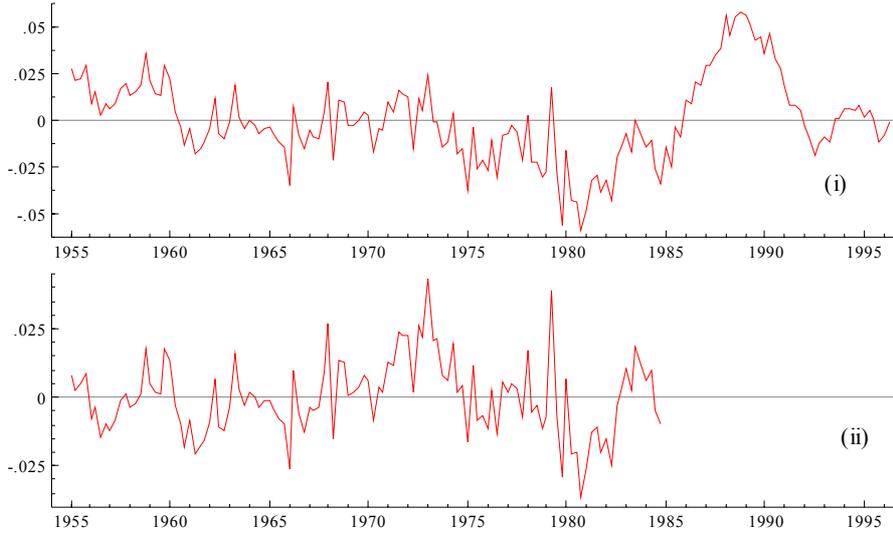


Figure 2: Regression Residuals: (i) Full sample (ii) Sample to 1984Q4.

	F	DW	res- d	Δ res- d
test statistics	369.65	1.102	0.242	-0.492
p -value, strong null	0	0	0	0.04
p -value, weak null (1 lag, 1 lead)	0.01	0.008	0.13	0.72
p -value, weak null (2 lags, 2 leads)	0.11	0.054	0.32	0.78.

Table 2: Cointegration Tests, 1955Q1-1984Q4

relationship, it is also desirable to conduct the test over this shorter span. In this case the fitted regression is

$$\hat{c}_t = 0.668y_t + 3.42 + 0.0015t$$

and the test results are given in Table 2. Here, we check the effect of over-parameterising the null model by extending the number of lags and leads to two, although the one-lead-one-lag test is the one to be relied on, and shows a reasonably clear-cut rejection by the F and DW criteria.

For comparison, the augmented Dickey Fuller statistic, applied to the residuals in Figure 2(ii), is equal to -4.35 with one lag and -3.44 with two lags. MacKinnon's (1991) tables give the approximate 5% and 1% critical values in 120 observations as -3.86 and -4.46 , respectively. It might be found surprising that the evidence for cointegration seems equivocal, but note that this version of the DHSY model includes a linear trend, which the original paper and many subsequent variants have not. There is no doubt that these series share a strong deterministic trend. The match between the stochastic trends is somewhat less clear cut, and the bootstrap tests summarise the evidence quite neatly.

5.2 Size and Power Evaluation

The properties of these tests are studied by Monte Carlo replication of the whole test procedure. Randomly generated ARFIMA(0, d ,0) series are added to $0.795y_t$, plus intercept and trend, to form the artificial regressand. The shocks are independent normal variates, whose variance is set equal to the variance of the differences of the 'predicted part' of the artificial regression, after

		F			DW			residual d		
alternatives		1%	5%	10%	1%	5%	10%	1%	5%	10%
strong null	$d = 1$	0.576	0.706	0.779	0.020	0.052	0.089	0.005	0.021	0.042
	$d = 0.75$	0.967	0.991	0.997	0.183	0.324	0.407	0.044	0.140	0.232
	$d = 0.5$	1	1	1	0.612	0.735	0.782	0.229	0.483	0.621
	$d = 0.25$	1	1	1	0.875	0.926	0.945	0.579	0.811	0.897
	$d = 0$	1	1	1	0.978	0.993	0.995	0.833	0.955	0.975
true null	$d = 1$	0.022	0.067	0.125	0.019	0.049	0.094	0.006	0.024	0.058
	$d = 0.75$	0.269	0.505	0.639	0.177	0.293	0.365	0.032	0.115	0.200
	$d = 0.5$	0.913	0.989	0.998	0.599	0.708	0.763	0.217	0.464	0.604
	$d = 0.25$	1	1	1	0.88	0.93	0.949	0.561	0.806	0.879
	$d = 0$	1	1	1	0.980	0.993	0.996	0.810	0.950	0.979
weak null	$d = 1$	0.009	0.052	0.102	0.013	0.042	0.069	0.017	0.043	0.090
	$d = 0.75$	0.146	0.337	0.467	0.083	0.218	0.319	0.029	0.097	0.186
	$d = 0.5$	0.680	0.890	0.955	0.482	0.743	0.856	0.140	0.325	0.462
	$d = 0.25$	0.982	0.998	1	0.995	0.999	1	0.361	0.621	0.791
	$d = 0$	0.999	1	1	1	1	1	0.672	0.857	0.943

Table 3: Powers of UK consumption cointegration tests.

detrending, as a scaling adjustment. The bootstrap tests are conducted on these series using 500 bootstrap replications at each Monte Carlo replication. The simulation of the test procedure includes the fitting of the artificial regressand's d by maximum likelihood. The only concession of realism to feasibility is the omission of the specification search for the ARFIMA model of the regressand. This was chosen as ARFIMA(0, d ,0) for each replication.

If the bootstrapping process correctly represents the distribution of the test statistics under the null hypothesis, the asymptotic distribution of the p -values is uniform on $[0, 1]$, by construction. Therefore, the proportion of p -values falling below α in the replications approximates the power of a α -level test. In particular, simulating the null hypothesis in this manner yields a check on the size of the test, since the value returned should be converging to α in this case. Table 3 gives these estimated powers for tests of levels 1%, 5% and 10%, based on 1000 Monte Carlo replications.¹⁶

To simulate tests of the weak null hypothesis, two approaches were taken. In the case where the added disturbance has $d = 1$, such that the null of non-cointegration is true, the model of the differences corresponding to (3.5) takes the form

$$\Delta c_t = 0.795\Delta y_t + u_t \quad u_t \sim NID(0, \sigma^2).$$

In other words, the correct representation of the short-run dynamics contains Δy_t alone, with no lags and leads of either variable. We refer to this case in Table 3 as the test of the 'true null', in contrast to the test of the strong null (independence, which is false even when $d = 1$) and of the weak null, where Δc_{t-1} , Δy_{t-1} , Δy_t and Δy_{t+1} are included in the dynamic regression. The interest is to determine how far this over-fitting (which would typically be done in practice) lowers the power of the test

A notable feature of these results is the good performance of the F test, which is best in all cases. It is to be expected that this test would perform best with the strong null hypothesis,

¹⁶Both variants of the d test perform poorly relative to the F and the DW, and for the sake of clarity, results for the residual d in differences are not reported.

since in this case the null is false even if the residuals are $I(1)$, but it also performs better on average against the weak nulls, and against the stationary alternatives. By contrast, the test based on the Robinson residual d estimate performs poorest by a substantial margin, although it does have some power against the stationary alternatives.¹⁷ The performance of the Durbin Watson statistic is creditable, however. Be careful to note that the null is false when $d = 1$ in the strong null test, so the entries in this row of the table are not technically sizes, although only the F test has power against this alternative. In the other cases, the sizes (i.e. rejection frequencies in the cases $d = 1$) are on the whole fairly accurate, to within the expected experimental error, and the asymptotic approximation in 166 observations. Note that in a different application with values of d smaller than 1 (Davidson 2000b), quite substantial size distortion has been observed, although this can be corrected by pre-pivoting. See the cited paper for details.

It should not be overlooked that these power evaluations are specific to the consumption-income example, and are not a substitute for a full-scale Monte Carlo evaluation based on a range of different models. However, they are representative of other evaluations that have been conducted with different data sets.. Note too that, given adequate computing power, such a case-specific evaluation should be a routine part of any practical testing exercise.

6 Concluding Remarks

The role of bootstrap methods in econometrics often emphasized, as for example by Horowitz (1997), is that of providing second-order refinements to existing asymptotic tests. The methods in this paper, by contrast, are chiefly motivated by the idea of obtaining first-order approximations to asymptotically non-pivotal tests. This approach has specific advantages in the fractional cointegration model, but as the consumption example has illustrated, they can be used as an alternative to the usual first-order tests. In addition to being applicable to any order of integration, it provides the means to test both strong and weak forms of the non-cointegration hypothesis, permitting a more sensitive appraisal of the relationship under investigation. Its power in the model of interest can be checked directly by simulation.

An advantage that has not been explored here, but will be the subject of future work, is the ability to base tests on more general regression models. These might contain dummy variables to permit switches of regime, for example, or lags of regressors and regressand, and even nonlinear-in-variables models. The main issue in such implementations will be the adequacy of the ARFIMA framework to represent the distribution of the data, but modifications of the approach can easily be envisaged; for example, in the case of shift dummies, suitable shifts might be applied to the artificial regressands.

The feasibility of the method for routine application should also be emphasized. With 166 observations, and a realistically adequate 500 bootstrap replications, the procedures reported here, coded in Ox 2.20 (Doornik, 1999), take around 4 seconds on a 300 MHz Pentium II. Increasing the number of regressors would not increase this time proportionately, since these are held fixed and much of the manipulation is done only once, not 500 times.

However, some work remains to be done to develop these procedures for general application. Learning more about their properties in different data sets and sample sizes, and possible pitfalls in their use, as well as a head-to-head comparison with conventional asymptotic tests and also the maximized Monte Carlo procedure, are on the agenda for future work.

¹⁷Hassler, Marmol and Velasco (2000) point out that trimming the periodogram should improve the properties of this estimator, although whether such refinements could make this test competitive with the other two is an open question. The improvement would clearly need to be dramatic.

A Appendix

This section provides details of the derivations given in Sections 4.1-4.3. Let $\mathbf{x}_t = (y_t, \mathbf{z}_t)'$ denote a $\mathbf{I}(d)$ vector for $d > 1/2$. According to the formula in equation (3.12) of Davidson and de Jong (2000), the long-run covariance matrix (4.2) in the case being considered has the form

$$\Psi = \frac{1}{\Gamma(2d-1)} \left(\frac{1}{2d-1} + \int_0^\infty ((1+\tau)^{d-1} - \tau^{d-1})^2 d\tau \right) \Omega \quad (\text{A-1})$$

where

$$\Omega = \lim_{n \rightarrow \infty} n^{-1} E \left(\sum_{t=1}^n \mathbf{w}_t \sum_{s=1}^n \mathbf{w}_s' \right)$$

and $\mathbf{w}_t = (1-L)^d \mathbf{x}_t$.¹⁸ It follows that

$$\sum_{t=1}^n \mathbf{x}_t \mathbf{x}_t' = O_p(n^{2d}). \quad (\text{A-2})$$

Note that the null hypothesis $H_0 : b = 0$ in (2.1) rules out cointegration amongst the \mathbf{z}_t variables. Under this hypothesis, Ψ has full rank. Under the alternative, u_t is either covariance stationary ($d-b < 1/2$), or has long run variance of order $n^{2(d-b)-1}$ for $b > 0$. This implies $|\Psi| = 0$. Let the partition of Ψ into the first 1 and last k rows and columns be

$$\Psi = \begin{bmatrix} \psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}.$$

The vector \mathbf{W}_d of standard fractional Brownian motions satisfies $E \mathbf{W}_d(\delta) \mathbf{W}_d(\delta)' = \delta^{2d-1} \mathbf{I}$ for $0 \leq \delta \leq 1$. One can write $\mathbf{X} = \mathbf{L}' \mathbf{W}_d$ such that $\mathbf{L}' \mathbf{L} = \Psi$, and \mathbf{L} has the partition

$$\mathbf{L} = \begin{bmatrix} \sqrt{\psi_{11.2}} & \mathbf{0}' \\ (\mathbf{L}_{22}^{-1})' \Psi_{21} & \mathbf{L}_{22} \end{bmatrix}$$

where $\psi_{11.2} = \psi_{11} - \Psi_{12} \Psi_{22}^{-1} \Psi_{21}$ and $\mathbf{L}_{22}' \mathbf{L}_{22} = \Psi_{22}$. Thus,

$$\begin{bmatrix} Y \\ \mathbf{Z} \end{bmatrix} = \begin{bmatrix} \sqrt{\psi_{11.2}} W_{d1} + \Psi_{12} \mathbf{L}_{22}^{-1} \mathbf{W}_{d2} \\ \mathbf{L}_{22}' \mathbf{W}_{d2} \end{bmatrix}. \quad (\text{A-3})$$

where $\mathbf{W}_d = (W_{d1}, \mathbf{W}_{d2})'$. Similarly define the partition of $\mathbf{G} = \int_0^1 \mathbf{W}_d \mathbf{W}_d' dr$ into its first 1 and last k rows and columns as

$$\mathbf{G} = \begin{bmatrix} g_{11} & \mathbf{G}_{12} \\ \mathbf{G}_{21} & \mathbf{G}_{22} \end{bmatrix}.$$

Under $H_0 : b = 0$, with $d > 1/2$,

$$n^{1/2-d} y_{[n\delta]} \xrightarrow{d} Y(\delta) = \Psi_{21}' \Psi_{22}^{-1} \mathbf{Z}(\delta) + V(\delta) \quad (\text{A-4})$$

where \mathbf{Z} and $V = \sqrt{\psi_{11.2}} W_{d1}$ are uncorrelated by construction, and being Gaussian, are therefore independent. We have $E \mathbf{Z}(\delta) \mathbf{Z}(\delta)' = \delta^{2d-1} \Psi_{22}$, $EV(\delta)^2 = \delta^{2d-1} \psi_{11.2}$ and $E(\mathbf{Z}(\delta) V(\delta)) = \mathbf{0}$ for $0 \leq \delta \leq 1$. Therefore, on H_0 ,

$$\hat{\gamma} \xrightarrow{d} \left(\int_0^1 \mathbf{Z} \mathbf{Z}' ds \right)^{-1} \int_0^1 \mathbf{Z} Y ds$$

¹⁸If the ds of different elements of the vector are different, the simple scale factor in (A-1) is replaced a more complicated dependence of the matrix on the d values and Ω . See Davidson and de Jong (2000) for the details of this case.

$$\begin{aligned}
&= \Psi_{22}^{-1} \Psi_{21} + \left(\int_0^1 \mathbf{Z} \mathbf{Z}' ds \right)^{-1} \int_0^1 \mathbf{Z} V ds \\
&= \Psi_{22}^{-1} \Psi_{21} + \sqrt{\psi_{11.2}} \mathbf{L}_{22}^{-1} \mathbf{G}_{22}^{-1} \mathbf{G}_{21}.
\end{aligned}$$

This is the formula in (4.3) where $\zeta = \sqrt{\psi_{11.2}} \mathbf{L}_{22}^{-1} \mathbf{G}_{22}^{-1} \mathbf{G}_{21}$.

Since it is somewhat more straightforward, we first derive the distribution of the F/n statistic under H_0 . The formula in (4.19) may also be written in the form

$$\frac{F}{n} = \frac{\sum_{t=1}^n z_t' y_t (\sum_{t=1}^n z_t z_t')^{-1} \sum_{t=1}^n z_t y_t}{\sum_{t=1}^n y_t^2 - \sum_{t=1}^n z_t' y_t (\sum_{t=1}^n z_t z_t')^{-1} \sum_{t=1}^n z_t y_t}.$$

Combining the functional central limit theorem for fractional processes with the continuous mapping theorem, and substituting for the Y process using (A-4) leads straightforwardly to

$$\frac{F}{n} \xrightarrow{d} \frac{\left(\int_0^1 \mathbf{Z} \mathbf{Z}' dr \Psi_{22}^{-1} \Psi_{21} + \int_0^1 \mathbf{Z} V dr \right)' \left(\int_0^1 \mathbf{Z} \mathbf{Z}' dr \right)^{-1} \left(\int_0^1 \mathbf{Z} \mathbf{Z}' dr \Psi_{22}^{-1} \Psi_{21} + \int_0^1 \mathbf{Z} V dr \right)}{\int_0^1 V^2 dr - \int_0^1 \mathbf{Z}' V dr \left(\int_0^1 \mathbf{Z} \mathbf{Z}' dr \right)^{-1} \int_0^1 \mathbf{Z} V dr}.$$

Standard manipulations using (A-3) put this expression into the form of (4.20).

Next, consider the statistic $\hat{\phi}$ in (4.9) under H_0 . For the denominator,

$$\frac{1}{n^{2d}} \hat{\theta}' \sum_{t=2}^n \mathbf{x}_t \mathbf{x}_t' \hat{\theta} \xrightarrow{d} \theta' \int_0^1 \mathbf{X} \mathbf{X}' ds \theta$$

follows by a straightforward teaming of the fractional FCLT with the continuous mapping theorem. For the three cases of the numerator, first note that if $1 < d < 3/2$ in (2.1) and $E(\varepsilon_{jt}^4) < \infty$ for $j = 1, \dots, N$,¹⁹ the result

$$\frac{1}{n^{2d-1}} \sum_{t=2}^n \mathbf{x}_{t-1} \Delta \mathbf{x}_t' \xrightarrow{d} \frac{1}{\Gamma(d-1)} \int_0^1 \xi^{d-2} \int_0^{1-\xi} \mathbf{X} d\mathbf{W}' d\xi + \mathbf{\Lambda}$$

holds by Theorem 4.1 of Davidson (2001b). For $d = 1$ the result

$$\frac{1}{n} \sum_{t=2}^n \mathbf{x}_{t-1} \Delta \mathbf{x}_t' \xrightarrow{d} \int_0^1 \mathbf{X} d\mathbf{X}' + \mathbf{\Lambda}^*$$

is standard (see e.g. Davidson 2000a, Chapter 15). For $1/2 < d < 1$, the result

$$\frac{1}{n} \sum_{t=2}^n \mathbf{x}_{t-1} \Delta \mathbf{x}_t' \xrightarrow{p} \mathbf{\Lambda}^*$$

follows by Theorem 4.3(iv) of Davidson (2001b).

Under the alternative hypothesis $b > 0$, there exists γ such that $u_t = y_t - \gamma' z_t \sim \text{I}(d-b)$. Here, it is convenient to expand the denominator of (4.19) and (4.8) according to

$$\sum_{t=1}^n \hat{u}_t^2 = \sum_{t=1}^n u_t^2 - (\hat{\gamma} - \gamma)' \sum_{t=1}^n z_t z_t' (\hat{\gamma} - \gamma). \tag{A-5}$$

¹⁹It is not yet clear whether the existence of fourth moments can be relaxed by refining the proof of this result. Given that moments of order strictly exceeding 2 are sufficient for the FCLT, according to Davidson and de Jong (2000), the conjecture may not be unreasonable, though a proof is currently lacking.

If $d > 1/2$ and $d - b < 1/2$, so that u_t is stationary, the first term on the right-hand side is $O_p(n)$ and the second term is $O_p(n^{\max\{2(d-b), 2(1-d)\}})$. Note that the first term dominates the second, so that the residual variance consistently estimates the disturbance variance. Also note that if $d - b > 1/2$, both terms are of $O_p(n^{2(d-b)})$, and the estimation error does not vanish in the limit. Our results do not depend on the latter condition, however. Since $\sum_{t=1}^n y_t^2 = O_p(n^{2d})$ continues to apply, the stated rates of divergence in (4.21) follow directly.

The numerator of (4.9) may be expanded similarly to (A-5), as

$$\begin{aligned} \sum_{t=2}^n \hat{u}_{t-1} \Delta \hat{u}_t &= \sum_{t=2}^n u_{t-1} \Delta u_t - (\hat{\gamma} - \gamma)' \sum_{t=2}^n z_{t-1} \Delta u_t - \sum_{t=2}^n u_{t-1} \Delta z'_{t-1} (\hat{\gamma} - \gamma) \\ &\quad + (\hat{\gamma} - \gamma)' \sum_{t=2}^n z_{t-1} \Delta z'_{t-1} (\hat{\gamma} - \gamma). \end{aligned}$$

The facts that $u_t \sim \text{I}(d-b)$ and $z_t \sim \text{I}(d)$, Theorems 4.1 and 4.3 from Davidson (2001b), and the results in Section 3 of that paper, yield the following rates of divergence.

$$\begin{aligned} \sum_{t=2}^n u_{t-1} \Delta u_t &= O_p(n^{\max\{1, 2(d-b)-1\}}) \\ \sum_{t=2}^n z_{t-1} \Delta u_t &= O_p(n^{\max\{1, d, 2d-b-1\}}) \\ \sum_{t=2}^n u_{t-1} \Delta z_t &= O_p(n^{\max\{1, 2d-b-1\}}) \\ \sum_{t=2}^n z_{t-1} \Delta z_t &= O_p(n^{\max\{1, 2d-1\}}). \end{aligned}$$

Also using (4.4), we obtain

$$\sum_{t=2}^n \hat{u}_{t-1} \Delta \hat{u}_t = O_p(n^{\max\{1, 2(d-b)-1\}}).$$

Considering (A-5), combining (4.4) with (A-2) and the analogous

$$\sum_{t=1}^n u_t^2 = O_p(n^{\max\{1, 2(d-b)\}})$$

we obtain

$$\sum_{t=1}^n \hat{u}_t^2 = O_p(n^{\max\{1, 2(d-b)\}}).$$

Combining these results gives the conclusions of (4.17)+(4.18). However, be careful to note that in general, the rates of divergence of the two terms in (A-5) are the same. These results do not depend upon the error in $\hat{\gamma}$ becoming asymptotically negligible.

References

- Andersson, M. K. and Gredenhoff, M. P., 1998, Robust testing for fractional integration using the bootstrap, Working Paper in Economics and Finance 218, Stockholm School of Economics.
- Athreya, K. B., 1987, Bootstrap of the mean in the infinite variance case, *Annals of Statistics* 15, 2, 724–31.
- Baillie, R. T., 1996, Long memory processes and fractional integration in econometrics, *Journal of Econometrics* 73, 5–59
- Baillie, R. T. and Bollerslev, T., 1994, Cointegration, fractional integration and exchange rate dynamics, *Journal of Finance* XLIX, 2, 737–45
- 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.
- Booth, G. B. and Tse, Y., 1995, Long memory in interest rate futures markets: a fractional cointegration analysis, *Journal of Futures Markets* 15, 5, 573–84
- Box, G. E. P., and D. A. Pierce, (1970). The distribution of residual autocorrelations in autoregressive-integrated moving average time series models *Journal of the American Statistical Association* 65, 1509–26.
- Breitung, J., and U. Hassler, 2002, Inference on the cointegration rank in fractionally integrated processes, *Journal of Econometrics*, this issue.
- Cheung, Y.-W. and F. X. Diebold, 1994, On maximum likelihood estimation of the differencing parameter of fractionally integrated noise with unknown mean, *Journal of Econometrics* 62, 301–316
- Cheung, Y.-W. and Lai, K. S., 1993, A fractional cointegration analysis of purchasing power parity *Journal of Business and Economic Statistics* 11, 1, 103–12
- Davidson, J., 1998, Structural relations, cointegration and identification: some simple results and their application *Journal of Econometrics* 87, 87–113
- Davidson, J., 2000a, *Econometric Theory*, Oxford: Blackwell Publishers.
- Davidson, J., 2000b, 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>.
- Davidson, J., 2001a, Establishing conditions for the functional central limit theorem in nonlinear and semiparametric time series processes, Forthcoming in *Journal of Econometrics*, and at <http://www.cf.ac.uk/carbs/davidsonje>
- Davidson, J., 2001b, 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>
- Davidson, J., Hendry, D. F., Srba, F. and Yeo, S., 1978, Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom, *Economic Journal* 88, 661–92.
- 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.

- Dittmann, I., 2000, Residual-based tests for fractional cointegration: a Monte carlo study, *Journal of Time Series Analysis* 21, 6, 615–647.
- Doornik, J. A., 1999, *Object-Oriented Matrix Programming Using Ox*, 3rd Edn. London: (Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik).
- Dueker, M. and Startz, R., 1998, Maximum likelihood estimation of fractional cointegration with an application to US and Canadian bond rates, *Review of Economics and Statistics* 80, 3, 420–6.
- Dufour, J.-M., 2000, Monte Carlo tests with nuisance parameters: a general approach to finite sample inference and nonstandard asymptotics.’ Mimeo, Department of Economic Science, University of Montreal.
- Engle, R. F., and C. W. J. Granger, 1987, Co-integration and error correction: representation, estimation and testing, *Econometrica* 55, 251–76.
- Fachin, S., 2000, Bootstrap and asymptotic tests of long-run relationships of cointegrated systems, *Oxford Bulletin of Economics and Statistics* 62, 4, 543–551.
- Fox, R. and M. S. Taqqu, 1986, Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series, *Annals of Statistics* 14, 517–532.
- Geweke, J. and S. Porter-Hudak, 1983, The estimation and application of long memory time series models, *Journal of Time Series Analysis* 4, 221–238
- Granger, C. W. J., 1986, Developments in the study of cointegrated economic variables, *Oxford Bulletin of Economics and Statistics* 48, 213–228.
- Granger, C. W. J. and T.-H. Lee, 1990, Multicointegration, *Advances in Econometrics* Vol. 8, 71–84, reprinted as Chapter 9 of: *Long-run Economic Relationships*, Engle and Granger, eds. (Oxford University Press 1991).
- Hassler, U., F. Marmol and C. Velasco, 2000, Residual log-periodogram inference for long-run relationships, Working paper, Institute of Statistics and Econometrics, Free University of Berlin
- Hauser, M. A., 1999, Maximum likelihood estimators for ARMA and ARFIMA models: A Monte Carlo study, *Journal of Statistical Inference and Planning* 80, 229–255.
- 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 (Cambridge University Press).
- Horowitz, J. L., 2000, The Bootstrap, forthcoming in: *Handbook of Econometrics* Vol. 5 (North-Holland-Elsevier).
- Jeganathan, 1999, On asymptotic inference in cointegrated time series with fractionally integrated errors, *Econometric Theory* 15, 583–621.
- 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.
- Kim, C. S. and P. C. B. Phillips, 2000, Fully modified estimation of fractional cointegration models, Cowles Foundation Working Paper, Yale University.
- Kremers, J. J. M., N. R. Ericsson, and J. J. Dolado, 1992, The power of cointegration tests, *Oxford Bulletin of Economics and Statistics* 54, 3, 325–348.
- Li, H. and Maddala, G. S., 1996, Bootstrapping time series models, *Econometric Reviews* 15, 2, 1–115

- Li, H. and Maddala, G. S., 1997, Bootstrapping cointegrating regressions, *Journal of Econometrics* 80, 297–318
- MacKinnon, J. G., 1991, Critical values for cointegration tests, Chapter 13 of *Long Run Economic Relationships*, eds. R. F. Engle and C. W. J. Granger, (Oxford University Press).
- McLeod, A. I. and Li, W. K., 1983, Diagnostic checking ARMA time series models using squared-residual autocorrelations, *Journal of Time Series Analysis* 4, 269–273.
- Marinucci, D., 2000, Spectral regression for cointegrated time series with long-memory innovations, *Journal of Time Series Analysis* 21, 6, 685–705
- Marinucci, D. and Robinson, P. M., 1999, Alternative forms of fractional Brownian motion. *Journal of Statistical Inference and Planning* 80, 111–122.
- Marmol, F., 1996, Nonsense regressions between integrated processes of different orders *Oxford Bulletin of Economics and Statistics* 58,3 525–536
- Marmol, F., 1998, Spurious regression theory with nonstationary fractionally integrated processes, *Journal of Econometrics* 84, 233–50
- Masih, R and Masih, A. M. M., 1995, A fractional cointegration approach to empirical tests of PPP: new evidence and methodological implications from an application to the Taiwan/US Dollar relationship, *Weltwirtschaftliches-Archiv*; 131, 4, 673–94
- Phillips, P. C. B., 1986, Understanding spurious regressions in econometrics, *Journal of Econometrics* 33, 311–40.
- Phillips, P. C. B. and M. Loretan, 1991, Estimating long-run economic equilibria, *Review of Economic Studies* 58, 407–37.
- Robinson, P. M., 1994, Semiparametric analysis of long memory time series, *Annals of Statistics* 22, 1, 515–539.
- Robinson, P. M. and D. Marinucci, 1998, Semiparametric Frequency Domain Analysis of Fractional Cointegration, STICERD Working Paper EM/98/348.
- Saikkonen, P., 1991, Asymptotically efficient estimation of cointegration regressions, *Econometric Theory* 7, 1–21.
- Schwarz, G., 1978, Estimating the Dimension of a Model. *Annals of Statistics* 6, 461–4.
- Sephton, P. S., 1996, A note on fractional cointegration, *Applied Economics Letters* 3, 683–5.
- Shao, J. and Tu, D., 1995, *The Jackknife and Bootstrap*, New York: Springer.
- Sims, C. A. and H. Uhlig, 1991, Understanding unit roots: a helicopter tour, *Econometrica* 59, 6, 1591–9.
- Sowell, F., 1992, Maximum likelihood estimation of stationary univariate fractionally integrated time series models, *Journal of Econometrics*, 53, 165–188.
- Stock, J. H. and Mark W. Watson, 1993, A simple estimator of cointegrating vectors in higher-order cointegrated systems, *Econometrica* 61, 783–820
- Velasco, C. and P. M. Robinson, 2000, Whittle pseudo-maximum likelihood estimation for non-stationary time series, STICERD Working Paper EM/00/391, London School of Economics.