

Fractional integration and cointegration in US financial time series data

Guglielmo Maria Caporale · Luis A. Gil-Alana

Received: 18 May 2012 / Accepted: 25 October 2013 / Published online: 17 December 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract This paper examines several US monthly financial time series using fractional integration and cointegration techniques. The univariate analysis based on fractional integration aims to determine whether the series are $I(1)$ (in which case markets might be efficient) or alternatively $I(d)$ with $d < 1$, which implies mean reversion. The multivariate framework exploiting recent developments in fractional cointegration allows to investigate in greater depth the relationships between financial series. We show that there might exist many (fractionally) cointegrated bivariate relationships among the variables examined, for some of which only standard cointegration tests had previously been carried out.

Keywords Fractional integration · Long-range dependence · Fractional cointegration · Financial data

JEL Classification C22 · G10

1 Introduction

This paper re-examines the statistical properties of a number of US financial series (such as stock market prices, dividends, earnings, consumer prices and long-term interest rates) contained in the well-known dataset which can be downloaded from

G. M. Caporale (✉)
Centre for Empirical Finance, Brunel University, London UB8 3PH, UK
e-mail: guglielmo-maria.caporale@brunel.ac.uk

L. A. Gil-Alana
Navarra Center for International Development (NCID), ICS, University of Navarra, Pamplona, Spain

Robert Shiller's homepage, and which also are described in chapter 26 of Shiller's (Shiller 1989) book on 'Market Volatility'.

In the existing literature, the efficient markets hypothesis (EMH) has recently been tested using the present value (PV) model of stock prices, since if stock market returns are not predictable, as implied by the EMH, stock prices should equal the present value of expected future dividends, and therefore, stock prices and dividends should be cointegrated, as pointed out by Campbell and Shiller (1987). In their seminal paper, they tested the PV model of stock prices adopting Engle and Granger (1987) cointegration procedure, an approach which is valid provided stock prices and dividends are stationary in first differences rather than in levels.¹ They used the standard and poor's (S&P's) dividends and value-weighted and equally weighted New York stock exchange (NYSE) 1926–1986 datasets. In the case of the S&P series they rejected the unit root hypothesis for dividends, but not for stock prices, whilst they could not reject it for either when using the NYSE data. As for cointegration, their results were also mixed, some test statistics rejecting the null hypothesis of no-cointegration, other failing to reject it. These inconclusive results may be a consequence of assuming integer orders of differentiation as in the case of standard integration and cointegration models not allowing for non-integer values. Other empirical papers analysing cointegration in stock markets are Hakkio and Rush (1989), Baillie and Bollerslev (1989), Richards (1995), Crowder (1996) and Rangvid (2001).

However, as already mentioned, the discrete options $I(1)$ and $I(0)$ of classical cointegration analysis are rather restrictive: the equilibrium errors might in fact be a fractionally integrated $I(d)$ -type process, with stock and dividends being fractionally cointegrated. This is stressed by Caporale and Gil-Alana (2004), who propose a simple two-step residuals-based strategy for fractional cointegration based on the approach of Robinson (1994): first the order of integration of the individual series is tested, and then the degree of integration of the estimated residuals from the cointegrating regression. They find that the cointegrating relationship between stock prices and dividends possesses long memory, implying that the adjustment to equilibrium takes a long time and that PV models of stock prices are valid only over a long horizon.

The present study makes the following twofold contribution. Firstly, it applies univariate tests based on long memory in order to establish the order of integration of the individual series, extending the analysis from the $I(1)/I(0)$ cases to the more general case of fractional integration. Noting that the results can vary substantially depending on the methodology used we employ a battery of non-parametric, semiparametric and parametric techniques. Secondly, it examines bivariate relationships among the variables using the most recent fractional cointegration techniques, which also allows for slow adjustment to equilibrium. To our knowledge, although numberless studies exist analysing such relationships, ours is the first to do so within such a framework. The implications of the findings are also discussed. In particular, we argue that it is the presence of long memory in the cointegrating relationships (already documented in Caporale and Gil-Alana (2004) that can explain the inconclusiveness of the results

¹ A constant discount rate is assumed in that study. In a subsequent paper (Campbell and Shiller 1988) this assumption is relaxed to allow for time-varying discount rates in the PV model.

of other studies only allowing for integer degrees of differentiation. The layout of the paper is the following. Section 2 reviews the concepts of fractional integration and cointegration and the methods applied in this study. Section 3 describes the data and reports the empirical results. Section 4 offers some concluding remarks.

2 Methodology

The methodology employed in this study is based on the concept of long memory or long-range dependence. Given a zero-mean covariance stationary process $\{x_t, t = 0, \pm 1, \dots\}$ with autocovariance function $\gamma_u = E(x_t x_{t+u})$, in the time domain, long memory is defined such that:

$$\sum_{u=-\infty}^{\infty} |\gamma_u| = \infty.$$

Now, assuming that x_t has an absolutely continuous spectral distribution function, with a spectral density function given by:

$$f(\lambda) = \frac{1}{2\pi} \left(\gamma_0 + 2 \sum_{u=1}^{\infty} \gamma_u \cos(\lambda u) \right),$$

according to the frequency domain definition of long memory the spectral density function is unbounded at some frequency λ in the interval $[0, \pi)$, i.e.

$$f(\lambda) \rightarrow \infty, \text{ as } \lambda \rightarrow \lambda^*, \lambda^* \in [0, \pi].$$

Most of the empirical literature in the last twenty years has focused on the case where the singularity or pole in the spectrum occurs at the 0 frequency, i.e.

$$f(\lambda) \rightarrow \infty, \text{ as } \lambda \rightarrow 0^+.$$

This is the standard case of I(d) models of the form:

$$(1 - L)^d x_t = u_t, \quad t = 1, 2, \dots \tag{1}$$

with $x_t = u_t = 0$ for $t \leq 0$, where L is the lag-operator ($Lx_t = x_{t-1}$) and u_t is I(0), which is defined as a covariance stationary process with a spectral density function that is positive and bounded at all frequencies. Note that we adopt the Type II definition of fractional integration (see [Marinucci and Robinson 1999](#)). Finally, one should also note that fractional integration may also occur at some other frequencies away from 0, as in the case of seasonal/cyclical models (see [Arteche 2002](#); [Arteche and Robinson 2000](#); [Hassler et al. 2009](#) among others).

In the multivariate case, the natural extension of fractional integration is the concept of fractional cointegration. Though the original idea of cointegration, as in [Engle and Granger \(1987\)](#), allows for fractional orders of integration, all the empirical work carried out during the 1990s were restricted to the case of integer degrees of differencing.

Only in recent years have fractional values also been considered. In what follows, we briefly describe the methodology used in this paper for testing fractional integration and cointegration in the case of Shiller's financial time series data.

2.1 Fractional integration

There exist several methods for estimating and testing the fractional differencing parameter d . Some of them are parametric, while others are semiparametric or even non-parametric, and can be specified in the time or in the frequency domain. In this paper, we use first a parametric approach developed by [Robinson \(1994\)](#). This is a testing procedure based on the Lagrange multiplier (LM) principle that uses the Whittle function in the frequency domain. It tests the null hypothesis:

$$H_0 : d = d_0, \quad (2)$$

for any real value d_0 , in a model given by the equation (1), where x_t can be the errors in a regression model of the form:

$$y_t = \beta^T z_t + x_t, \quad t = 1, 2, \dots, \quad (3)$$

where y_t is the observed time series, β is a $(k \times 1)$ vector of unknown coefficients and z_t is a set of deterministic terms that might include an intercept (i.e. $z_t = 1$), an intercept with a linear time trend ($z_t = (1, t)^T$), or any other type of deterministic processes. [Robinson \(1994\)](#) showed that, under certain very mild regularity conditions, the LM-based statistic (\hat{r}):

$$\hat{r} \rightarrow_d N(0, 1) \quad \text{as } n \rightarrow \infty,$$

where $\hat{r} \rightarrow_d$ stands for convergence in distribution, and this limit behaviour holds independently of the regressors z_t used in (3) and the specific model for the $I(0)$ disturbances u_t in (1). Other parametric approaches ([Sowell 1992](#); [Beran 1995](#)) were also employed in the empirical analysis and produced very similar results to those obtained using the method of [Robinson \(1994\)](#).

In addition, we employ a semiparametric method ([Robinson 1995](#)) which is essentially a local 'Whittle estimator' in the frequency domain, using a band of frequencies that degenerates to zero. The estimator is implicitly defined by:

$$\hat{d} = \arg \min_d \left(\log \overline{C(d)} - 2d \frac{1}{m} \sum_{s=1}^m \log \lambda_s \right), \quad (4)$$

$$\overline{C(d)} = \frac{1}{m} \sum_{s=1}^m I(\lambda_s) \lambda_s^{2d}, \quad \lambda_s = \frac{2\pi s}{n}, \quad \frac{1}{m} + \frac{m}{n} \rightarrow 0,$$

where $I(\lambda_s)$ is the periodogram of the raw series, x_t , and $d \in (-0.5, 0.5)$. Under finiteness of the fourth moment and other mild conditions, [Robinson \(1995\)](#) proved that:

$$\sqrt{m}(\hat{d} - d^*) \rightarrow_d N(0, 1/4) \text{ as } n \rightarrow \infty,$$

where d^* is the true value of d . This estimator is robust to a certain degree of conditional heteroscedasticity and is more efficient than other semiparametric competitors.²

2.2 Fractional cointegration

Engle and Granger (1987) suggested that, if two processes x_t and y_t are both $I(d)$, then it is generally true that for a certain scalar $a \neq 0$, a linear combination $w_t = y_t - ax_t$, will also be $I(d)$, although it is possible that w_t be $I(d - b)$ with $b > 0$. Given two real numbers d, b , the components of the vector c_t are said to be cointegrated of order d, b , denoted $c_t \sim CI(d, b)$ if:

- (i) all the components of c_t are $I(d)$,
- (ii) there exists a vector $\alpha \neq 0$ such that $s_t = \alpha'c_t \sim I(\gamma) = I(d - b), b > 0$.

Here, α and s_t are called the cointegrating vector and error, respectively. This prompts consideration of an extension of Phillips' (1991) triangular system, which for a very simple bivariate case is:

$$y_t = \nu x_t + u_{1t}(-\gamma), \tag{5}$$

$$x_t = u_{2t}(-d), \tag{6}$$

for $t > 0$, where for any vector or scalar sequence w_t , and any ζ , we introduce the notation $w_t(\zeta) = (1 - L)^\zeta w_t$. Note that $u_t = (u_{1t}, u_{2t})^T$ is now a bivariate zero mean covariance stationary $I(0)$ unobservable process and $\nu \neq 0, \gamma < d$. Under (5) and (6), x_t is $I(d)$, as is y_t by construction, while the cointegrating error $y_t - \nu x_t$ is $I(\gamma)$. Model (5) and (6) reduce to the bivariate version of Phillips' (1991) triangular form when $\gamma = 0$ and $d = 1$, which is one of the most popular models displaying $CI(1, 1)$ cointegration considered in both the empirical and theoretical literature.

Next, we focus on the estimation of the cointegrating relationship, specifically on the estimation of ν in (5). The simplest approach is to estimate it using the well-known ordinary least squares (OLS) estimator

$$\hat{\nu}_{ols}^t = \frac{\sum_{t=1}^n x_t y_t}{\sum_{t=1}^n x_t^2}, \tag{7}$$

where the superscript t indicates time domain estimation. Here, in the standard cointegrating setting, with $\gamma = 0$ and $d = 1$, it has been shown that in general $\hat{\nu}_{ols}^t$ is n -consistent with non-standard asymptotic distribution. In fractional settings, the properties of ols could be different from those within this framework. When the observables are purely non-stationary (so that $d \geq 0.5$), consistency of $\hat{\nu}_{ols}^t$ is retained, but its rate of convergence and asymptotic distribution depend crucially on γ and d .

² See Shao and Wu (2007) for the issue of heteroscedasticity in the context of the local Whittle estimator of the differencing parameter.

An alternative method of estimating v is in the frequency domain. Consider the estimator

$$\hat{v}_{ols}^f = \frac{\sum_{j=0}^{n-1} I_{xy}(\lambda_j)}{\sum_{j=0}^{n-1} I_x(\lambda_j)}, \tag{8}$$

where $\lambda_j = 2\pi j/n, j = 1, \dots, n$, are the Fourier frequencies, and for arbitrary sequences ξ_t, ζ_t , we define the discrete Fourier transform and (cross)-periodogram

$$w_\xi(\lambda) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n \xi_t e^{it\lambda}, I_{\xi\zeta}(\lambda) = w_\xi(\lambda) w'_\zeta(-\lambda), I_\xi(\lambda) = I_{\xi\xi}(\lambda).$$

Here, the discrete Fourier transform at a given frequency captures the components of the series related to this particular frequency. Robinson (1994) proposed the narrow band least squares (NBLS) estimator

$$\hat{v}_{NBLS} = \frac{\sum_{j=0}^m s_j \text{Re } I_{xy}(\lambda_j)}{\sum_{j=0}^m s_j I_x(\lambda_j)}, \tag{9}$$

where $1 \leq m \leq n/2; s_j = 1$ for $j = 0, n/2$ and 2 , otherwise; and $(1/m) + (m/n) \rightarrow 0$ as $n \rightarrow \infty$. He showed the consistency of this estimator even under stationary cointegration. As in the case of ols, in general NBLS has a non-standard limiting distribution.

Assuming that the process u_t in (5) and (6) has a parametric spectral density $f(\lambda) = f(\lambda; \theta)$, where θ is an unknown vector of short-memory parameters, Robinson and Hualde (2003), based on generalized least squares (GLS)-type corrections, propose methods to estimate optimally (under Gaussianity) v when $d - \gamma > 0.5$ (named strong cointegration). Denoting

$$\begin{aligned} z_t(c, d) &= (y_t(c), x_t(d))', \zeta = (1, 0)', p(\lambda; h) = \zeta' f(\lambda; h)^{-1}, \\ a(c, d, h) &= \sum_{j=1}^n p(\lambda_j; h) w_{x(c)}(-\lambda_j) w_{z(c,d)}(\lambda_j), q(\lambda; h) = \zeta' f(\lambda; h)^{-1} \zeta, \\ b(c, d) &= \sum_{j=1}^n q(\lambda_j; h) I_{x(c)}(\lambda_j), \text{ and defining } \hat{v}(c, d, h) = \frac{a(c, d, h)}{b(c, h)}, \end{aligned}$$

they considered five different estimators given by:

$$\hat{v}(\gamma, d, \theta), \hat{v}(\gamma, d, \hat{\theta}), \hat{v}(\gamma, \hat{d}, \hat{\theta}), \hat{v}(\hat{\gamma}, \hat{d}, \hat{\theta}), \hat{v}(\hat{\gamma}, \hat{d}, \hat{\theta}), \tag{10}$$

where $\hat{\gamma}, \hat{d}, \hat{\theta}$, are corresponding estimators of the nuisance parameters γ, d and θ . The estimators in (10) reflect different knowledge about the structure of the model, the first being in general unfeasible, the second only assuming knowledge of the integration

orders (as was done previously in the standard cointegrating literature), whereas the last estimator represents the most realistic case. Under regularity conditions, Robinson and Hualde (2003) showed that any of the estimators in (10) is $n^{d-\gamma}$ -consistent with identical mixed-Gaussian asymptotic distributions, leading to Wald tests on the parameter ν ,

$$W(\gamma, \delta, \theta), W(\gamma, \delta, \hat{\theta}), W(\gamma, \hat{\delta}, \hat{\theta}), W(\hat{\gamma}, \delta, \hat{\theta}), W(\hat{\gamma}, \hat{\delta}, \hat{\theta}), \tag{11}$$

where $W(c, d, h) = b(c, h)\{\hat{\nu}(c, d, h) - 1\}^2$, with a chi-squared limit distribution for the values of d and γ . Hualde and Robinson (2007) propose an estimator of ν in (5) and (6) in the case when $d - \gamma < 0.5$ (named weak cointegration). As in Robinson and Hualde (2003), this method is based on a GLS-type correction.

3 Data and empirical results

The monthly series analysed have been collected by Robert Shiller and his associates, and are available on <http://www.econ.yale.edu/~shiller/>. The sample period goes from 1871m1 to 2010m6. They are described in chapter 26 of Shiller’s (1989) book on ‘Market Volatility’, where further details can be found, and are constantly updated and revised. Specifically, they are the following series: stock market prices (monthly averages of daily closing S&P prices, computed from the S&P four-quarter tools for the quarter, since 1926, with linear interpolation to monthly figures), dividends (an index), earnings (also an index), a consumer price index (Consumer Price Index—All Urban Consumers) used for computing real values of the previous variables, a long-term interest rate (GS10, which is the yield on the 10-year Treasury bonds) and also a cyclically adjusted price-earnings ratio.

3.1 Univariate analysis: fractional integration

We first employ the parametric approach of Robinson (1994) described in Sect. 2, assuming that the disturbances are white noise. Thus, time dependence is exclusively modelled through the fractional differencing parameter d . In particular, we consider the set-up in (3) and (1), with $z^T = (1, t)^T$, testing H_0 given by equation (2), i.e. $d = d_0$, for d_0 in $[0, 0.001, 0.002, \dots, 2]$. In other words, the model under the null becomes:

$$y_t = \beta_0 + \beta_1 t + x_t; (1 - L)^{d_0} x_t = u_t \quad t = 1, 2, \dots, \tag{12}$$

and white noise u_t .

Table 1 displays the estimates of d (obtained as the values of d_0 that produce the lowest \hat{r} — statistics in absolute value) along with the 95 % confidence band of the non-rejection values of d_0 using Robinson’s (1994) parametric approach. For each series, we display the three cases commonly examined in the literature, i.e. the cases of no deterministic terms (i.e. imposing $\beta_0 = \beta_1 = 0$ a priori), an intercept (β_0 unknown and $\beta_1 = 0$ a priori) and an intercept with a linear time trend (β_0 and β_1 unknown). The

Table 1 Estimates of d in a model with white noise disturbances

	No regressors	An intercept	A linear time trend
Stock prices	1.169 (1.134, 1.208)	1.169 (1.135, 1.209)	1.170 (1.135, 1.209)
Dividends	1.906 (1.874, 1.941)	1.951 (1.916, 1.988)	1.951 (1.916, 1.988)
Earnings	1.855 (1.806, 1.910)	1.856 (1.806, 1.911)	1.856 (1.806, 1.911)
Consumer price index	1.210 (1.185, 1.241)	1.396 (1.350, 1.454)	1.401 (1.354, 1.456)
Long interest rate	1.111 (1.070, 1.157)	1.111 (1.070, 1.157)	1.110 (1.070, 1.156)
Real stock prices	1.156 (1.121, 1.195)	1.161 (1.126, 1.201)	1.161 (1.126, 1.201)
Real dividends	1.311 (1.279, 1.346)	1.505 (1.470, 1.544)	1.505 (1.470, 1.544)
Real earnings	1.756 (1.713, 1.803)	1.825 (1.779, 1.877)	1.825 (1.779, 1.877)
Price/earning ratio	1.237 (1.198, 1.282)	1.494 (1.449, 1.542)	1.494 (1.449, 1.542)

The values in parentheses refer to the 95 % confidence band of the non-rejection values of d using Robinson's (1994) parametric tests

inclusion of a time trend may appear unrealistic in the context of financial variables. Note, however, that in the case of fractional (or integer) differentiation the time trend disappears in the long run.³

The first noticeable feature in this table is that all the estimated values of d are above 1 and the unit root null hypothesis (i.e. $d = 1$) is rejected in all cases at the 5 % level. In general, the values are very similar for the three cases with deterministic terms, although the results change substantially from one series to another. Specifically, values of d above 1.5 are found in the case of dividends, earnings and real earnings. For the remaining series, the values are slightly above 1, but still significantly different from 1. However, these results are based on a model characterised by the lack of (weak)-autocorrelation for the error term. Therefore, in what follows we assume that the disturbances are weakly autocorrelated and model them first using the exponential spectral model of Bloomfield (1973). This is an approach to modelling the $I(0)$ error term that produces autocorrelations decaying exponentially as in the AR(MA) case. Therefore, it approximates ARMA structures with a small number of parameters, and performs extremely well in the context of Robinson's (1994) tests (see Gil-Alana 2004). The results using this approach are displayed in Table 2.

It can be seen that the values are much smaller than in the previous case of white noise disturbances. One series (long-term interest rates) has values which are strictly below 1, implying mean-reverting behaviour; for dividends and real stock prices the unit root null cannot be rejected. It is slightly rejected (at the 5 % level, but not at the 1 % level) for stock prices, consumer price index and price/earning ratio, and it is decisively rejected in favour of higher orders of integration for the remaining two series (earnings and real earnings). As a final specification, given the monthly frequency of the data, we assume that the error term follows a seasonal AR(1) process. The inclusion of seasonal dummies produced insignificant coefficients in all cases. The results (displayed in Table 3) are very similar to those based on white noise disturbances, with estimates of

³ For example, if $d_0 = 1$ in (12) with white noise u_t , the model becomes, for $t > 1$, a random walk with a drift. Also, $(1 - L)^{d_0}$ disappears in the long run for $0 < d_0 < 1$, and tends to a constant for $1 = d_0 < 2$.

Table 2 Estimates of d in a model with bloomfield-type disturbances

	No regressors	An intercept	A linear time trend
Stock prices	1.052 (1.000, 1.102)	1.052 (1.001, 1.102)	1.052 (1.001, 1.103)
Dividends	1.033 (0.987, 1.083)	1.037 (0.986, 1.087)	1.037 (0.985, 1.088)
Earnings	1.568 (1.499, 1.649)	1.569 (1.492, 1.653)	1.569 (1.492, 1.653)
Consumer price index	1.175 (1.148, 1.206)	1.187 (1.160, 1.211)	1.195 (1.173, 1.224)
Long interest rate	0.909 (0.864, 0.964)	0.908 (0.863, 0.964)	0.909 (0.864, 0.964)
Real stock prices	1.033 (0.981, 1.083)	1.037 (0.991, 1.087)	1.037 (0.991, 1.088)
Real dividends	1.339 (1.272, 1.419)	1.448 (1.388, 1.521)	1.448 (1.388, 1.521)
Real earnings	1.599 (1.510, 1.671)	1.600 (1.517, 1.681)	1.600 (1.517, 1.681)
Price/earning ratio	1.135 (1.060, 1.231)	1.269 (1.171, 1.398)	1.269 (1.171, 1.398)

The values in parentheses refer to the 95 % confidence band of the non-rejection values of d using Robinson's (1994) parametric tests

Table 3 Estimates of d in a model with seasonal AR(1) disturbances

	No regressors	An intercept	A linear time trend
Stock prices	1.169 (1.135, 1.208)	1.170 (1.135, 1.210)	1.170 (1.135, 1.210)
Dividends	1.902 (1.873, 1.935)	1.953 (1.921, 1.988)	1.953 (1.921, 1.988)
Earnings	1.875 (1.830, 1.926)	1.878 (1.832, 1.928)	1.878 (1.833, 1.928)
Consumer price index	1.188 (1.161, 1.220)	1.374 (1.326, 1.431)	1.378 (1.332, 1.434)
Long interest rate	1.111 (1.071, 1.157)	1.111 (1.071, 1.157)	1.110 (1.071, 1.157)
Real stock prices	1.155 (1.119, 1.194)	1.161 (1.125, 1.201)	1.161 (1.125, 1.201)
Real dividends	1.311 (1.279, 1.346)	1.505 (1.469, 1.544)	1.505 (1.469, 1.544)
Real earnings	1.742 (1.704, 1.787)	1.836 (1.793, 1.883)	1.838 (1.795, 1.885)
Price/earning ratio	1.234 (1.194, 1.278)	1.491 (1.447, 1.539)	1.491 (1.447, 1.540)

The values in parentheses refer to the 95 % confidence band of the non-rejection values of d using Robinson's (Robinson 1994) parametric tests

d which are all strictly above 1. Deeper inspection indicates that time trends are not required in any case, the intercept being sufficient for the deterministic component. Moreover, LR tests and other residuals-based tests suggest that the d -differenced series may all be weakly (non-seasonally) autocorrelated, implying that the model with Bloomfield disturbances may approximate accurately the order of integration of the series. Nevertheless, in view of the sensitiveness of the results to the specification of the error term, we also apply a semiparametric method that does not specify a functional form for the $I(0)$ disturbance term.

Figure 1 displays for each series the estimates of d based on the semiparametric method of Robinson (1995), i.e. \hat{d} as given by (4). The estimates of d are shown for a whole range of values of the bandwidth parameter $m = 1, 2, \dots, n/2$ (on the horizontal axis). Alternatively, we could have chosen an optimal bandwidth parameter as suggested in Henry (2001). Note that the choice of the bandwidth is crucial in view of the trade-off between bias and variance: the asymptotic variance is decreasing with m , while the bias is growing with m ; the 95 % confidence bands corresponding to

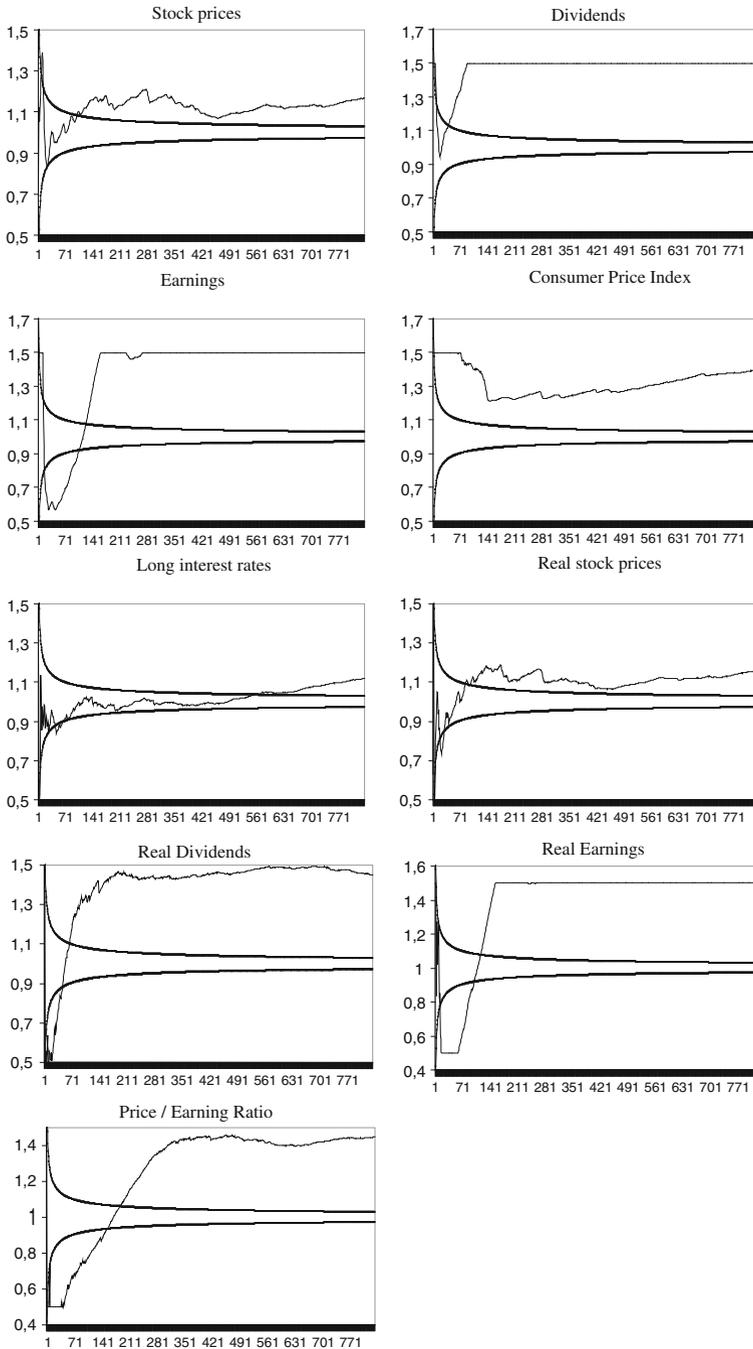


Fig. 1 Estimates of d based on the semiparametric estimate of Robinson (1995). The horizontal axis refers to the bandwidth parameter, while the vertical one corresponds to the estimated values of d . We report the estimates of d along with the 95 % confidence band of the $I(1)$ hypothesis

Table 4 Estimates of d using Robinson’s (1995) semiparametric method for different bandwidth numbers

	25	41	100	200	300	500
Stock prices	0.850*	0.953*	1.004*	1.121	1.158	1.092
Dividends	1.021*	1.105*	1.500	1.500	1.500	1.500
Earnings	0.589	0.580	0.875	1.500	1.500	1.500
Consumer price index	1.500	1.500	1.417	1.228	1.235	1.278
Long interest rate	0.893*	0.895*	0.983*	0.958*	0.990*	1.013*
Real stock prices	0.768	0.888*	1.071*	1.107	1.099	1.086
Real dividends	0.538	0.896*	1.326	1.455	1.438	1.464
Real earnings	0.500	0.500	0.933*	1.500	1.500	1.500
Price/earning ratio	0.500	0.500	0.745	1.041*	1.377	1.431
95 % confidence interval	(0.835, 1.164)	(0.871, 1.128)	(0.917, 1.082)	(0.941, 1.058)	(0.952, 1.047)	(0.963, 1.036)

* Indicates that the $I(1)$ hypothesis cannot be rejected at the 5 % level

the $I(1)$ hypothesis are also displayed. It can be seen that, for small values of m , the unit root null is rejected in favour of mean reversion ($d < 1$) in the case of earnings, real dividends, real earning and price-earning ratio. For the remaining series (still with a small m), the estimated values of d are within the $I(1)$ interval, except for the CPI series for which d is found to be strictly above 1. However, when the bandwidth parameter is large, the estimates are clearly above 1 in all cases, the only exception being long-term interest rates, with many values in the $I(1)$ interval. Table 4 reports the numerical values for different bandwidth parameters, $m = 25, 41 (= n^{0.5}), 100, 200, 300$ and 500 : at the 5 % level, there are several cases where the unit root null cannot be rejected. Specifically, in the case of $m = (n)^{0.5}$, which has been widely considered in the empirical literature, the unit root null hypothesis cannot be rejected for stock prices, dividends, long-term interest rates, real stock prices and real dividends, while it is rejected in favour of mean reversion (i.e. $d < 1$) for earnings and real earnings, and in favour of $d > 1$ for the consumer price index.

Overall, the univariate results provide no evidence of mean reversion: all series appear to be $I(1)$ or $I(d)$ with $d > 1$, implying permanent effects of shocks, with evidence of long memory in many cases for the first-differenced series as well.

3.2 Multivariate analysis: fractional cointegration

A number of cointegrating (bivariate) relationships might exist between the individual variables examined in the previous subsection, in particular between:

- (a) Stock prices and dividends
- (b) Real stock prices and real dividends
- (c) Price/earning ratio and long-term interest rates and
- (d) Real stock prices and real earnings.

Some of these relationships have been extensively analysed in the literature. Campbell and Shiller (1987) and DeJong (1992) tested a present value model of the stock

market using time series data for real US annual stock prices and dividends from 1871 to 1986. In the first of these studies, they carried out ADF tests, with and without a time trend, on both individual series, and their results suggested that both series were integrated of order 1. When using the DF and ADF tests on the residuals from the cointegrating regressions, their results were mixed: the former test rejected the null hypothesis of no cointegration at the 5 % level, while the latter narrowly failed to reject it at the 10 % level. DeJong (1992) used a Bayesian approach to model these two variables and found evidence in favour of trend-stationary representations. Similarly, Koop (1991), using a different dataset, came to the same conclusion that both variables are stationary around a linear trend, and, even when assuming unit roots, he found little evidence of cointegration with $I(0)$ errors.

Pereira-Garmendia (2010) finds that real stock prices and real earnings are related through inflation. The relationship among stock prices, earnings and bond yield is analysed by Durre and Giot (2007). Papers examining long-run linkages between the price/earnings ratio and interest rates include Philips (1999), Campbell and Shiller (1998), Campbell and Shiller (2001) and Asness (2003) inter alia.

In all cases, we follow the same strategy. We first estimate individually the orders of integration of the series using the log-periodogram-type estimator devised by Robinson (1995). This is defined as:

$$\hat{d}(l) = \sum_{j=l+1}^m (a_j - \bar{a}) \log I(\lambda_j)/S_l, \quad (13)$$

where

$$a_j = -\log \left(4 \sin^2 \left(\frac{\lambda_j}{2} \right) \right), \quad \bar{a} = \frac{1}{m-l} \sum_{j=1}^p a_j u_{t-j} + \varepsilon_t,$$

$$S_l = \sum_{j=l+1}^m (a_j - \bar{a})^2, \quad \lambda_j = \frac{2\pi j}{n},$$

and $0 \leq l < m < n$. The results for the individual series possibly involved in cointegration relationships are displayed in Table 5 (for $m = n^{0.5}$ and $l = 0, 1, \dots, 5$).⁴

Next, we test the homogeneity of the orders of integration in the bivariate systems (i.e. $H_0 : d_x = d_y$), where d_x and d_y are now the orders of integration of the two individual series, by using an adaptation of Robinson and Yajima (2002) statistic \hat{T}_{xy} to log-periodogram estimation. The statistic is:

$$\hat{T}_{xy} = \frac{m^{1/2} (\hat{d}_x - \hat{d}_y)}{\left(\frac{1}{2} \left(1 - \hat{G}_{xy}/(\hat{G}_{xx}\hat{G}_{yy}) \right) \right)^{1/2} + h(n)}, \quad (14)$$

where $h(n) > 0$ and \hat{G}_{xy} is the (xy) th element of

⁴ We will examine later these tables in detail for each of the potential cointegrating relationships.

Table 5 Estimates of d using Robinson’s (1995) log-periodogram semiparametric method for different values of l and fixed $m = (n)^{0.5}$

$m = n^{0.5}$	$l = 0$	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$
Stock prices	1.080	1.070	1.061	1.041	1.018	1.017
Dividends	1.095	1.048	1.026	1.037	1.134	1.222
Long interest rate	0.956	1.010	0.972	0.914	0.832	0.803
Real stock prices	0.972	1.000	1.085	1.077	1.018	1.016
Real dividends	0.822	0.851	0.981	0.975	0.996	0.997
Real earnings	0.970	1.009	1.073	1.099	1.129	1.162
Real earnings	0.279	0.128	0.059	-0.078	-0.142	-0.082
Price/earning ratio (*)	0.913	0.931	0.945	0.920	1.010	1.127
Price/earning ratio (**)	0.484	0.606	0.576	0.589	0.637	0.645

(*) and (**) indicates that the results are based on the original and first-differenced data respectively

Table 6 Testing the homogeneity in the order of integration (Robinson and Yajima 2002)

$m = n^{0.5}$	$l = 0$	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$
Stock prices/dividends	-0.145	0.214	0.340	0.038	-1.125	-1.970
Real stock prices/real dividends	1.471	1.455	1.009	0.990	0.215	0.185
P.E.R./long interest rates	-0.425	-0.776	-0.813	-1.617	-1.527	-0.675
Real stock prices/real earnings	0.580	0.676	1.356	1.520	0.078	-1.077

In all cases we employ $h(n)$ chosen as b^{-5-2i} , $i=1, 2, 3, 4$ and 5

$$\hat{G} = \frac{1}{m} \sum_{j=1}^m \text{Re} \left[\hat{\Lambda}(\lambda_j)^{-1} I(\lambda_j) \hat{\Lambda}(\lambda_j)^{-1*} \right], \hat{\Lambda}(\lambda_j) = \text{diag} \left\{ e^{i\pi \hat{d}_x / 2} \lambda^{-\hat{d}_x}, e^{i\pi \hat{d}_y / 2} \lambda^{-\hat{d}_y} \right\},$$

(see Gil-Alana and Hualde 2009 for evidence on the finite sample performance of this procedure). The results using this approach are displayed in Table 6. In general, we cannot reject the null hypothesis of equal orders of integration.⁵ In the following step, we perform the Hausman test for no cointegration of Marinucci and Robinson (2001) comparing the estimate \hat{d}_x of d_x with the more efficient bivariate one of Robinson (1995), which uses the information that $d_x = d_y = d_*$. Marinucci and Robinson (2001) show that

$$H_{im} = 8m \left(\hat{d}_* - \hat{d}_i \right)^2 \rightarrow_d \chi_1^2 \text{ as } \frac{1}{m} + \frac{m}{n} \rightarrow 0, \tag{15}$$

with $i = x, y$, and where $m < [n/2]$ is again a bandwidth parameter, analogous to that introduced earlier; \hat{d}_i are univariate estimates of the parent series and \hat{d}_* is a restricted estimate obtained in the bivariate context under the assumption that $d_x = d_y$. In particular,

⁵ As in the case of the previous table, the comments for the specific series will be presented later.

Table 7 Hausman test for no cointegration (Marinucci and Robinson 2001)

$m = n^{0.5}$	H	$s = 25$	$s = 41$	$s = 50$
Stock prices/ dividends	H_{as}	4.205*	0.239	0.102
	H_{bs}	2.420	1.260	1.040
	\hat{d}_*	0.916	1.088	1.077
Real stock prices/ real dividends	H_{as}	26.499*	11.469*	8.880*
	H_{bs}	13.520*	2.259	0.810
	\hat{d}_*	0.721	0.898	0.936
P.E.R./long interest rates	H_{as}	24.780*	16.457*	13.209*
	H_{bs}	28.728*	20.664*	32.262*
	\hat{d}_*	0.593	0.721	0.688
Real stock prices/ real earnings	H_{as}	71.520*	68.502*	74.649*
	H_{bs}	68.679*	64.952*	70.560*
	\hat{d}_*	0.487	0.628	0.653

H_{as} refers to the hypothesis with respect to the first of the two parent series, while H_{bs} to the second one $\chi^2_1(5\%) = 3.84$. * Indicates rejection of the null hypothesis of no cointegration at the 5% level.

$$\hat{d}_* = - \frac{\sum_{j=1}^s 1'_2 \hat{\Omega}^{-1} Y_j v_j}{21'_2 \hat{\Omega}^{-1} 1_2 \sum_{j=1}^s v_j^2}, \tag{16}$$

with 1_2 indicates a (2×1) vector of 1s, $\hat{\Omega}$ refers to the variance covariance matrix of $Y_j = [\log I_{xx}(\lambda_j), \log I_{yy}(\lambda_j)]^T$, and $v_j = \log j - \frac{1}{s} \sum_{j=1}^s \log j$. The limiting distribution above is presented heuristically, but the authors argue that it seems sufficiently convincing for the test to warrant serious considerations. The results using this approach are displayed in Table 7, and although when using the Hausman-type tests like those employed here a rejection of the null of no cointegration does not necessarily imply that the alternative (in our case, fractional cointegration) holds, we provide in the following pages supportive evidence of cointegration in many of the cases examined.

In the final part of the analysis, we apply the methods of Robinson and Hualde (2003) and Hualde and Robinson (2007). We identify parametric models for $f(\lambda)$ with u_t in (5) and (6) having the form,

$$u_t = A(L)\varepsilon_t, \tag{17}$$

where ε_t is supposed to be an i.i.d. process, and $A(L)$ is initially assumed to be diagonal, thus treating u_{1t} and u_{2t} separately. We approximate the two series as

$$\tilde{u}_{1t} = (1 - L)^{\tilde{\gamma}} [y_t - \hat{v}_{ols} x_t], \tag{18}$$

and

$$\tilde{u}_{2t} = (1 - L)^{\tilde{d}} x_t, \tag{19}$$

to obtain estimates of γ and d previously estimated using other methods, and follow Box–Jenkins-type procedures to identify the models within the ARMA class. The results based on this method are displayed in Table 10.

Next we examine each of the bivariate relationships.

3.2.1 Stock market prices and dividends

Figure 2a displays the plots of the two series. Both of them are relatively stable until the end of World War II, when they start increasing and also exhibit a higher degree of volatility.

Focusing first on the univariate results using the Whittle semiparametric estimator (Robinson 1995), it can be seen that for small values of m the unit root null cannot be rejected (see Table 4). Specifically, for $m = (n)^{0.5} = 4l$, the estimates are 0.953 and 1.105, respectively, for stock prices and dividends. Similar evidence of unit roots, though with slightly higher values, is obtained with the log-periodogram estimator of Robinson (1995) (see Table 5). For example, for $l = 0, 1, 2, \dots, 5$, and $m = (n)^{0.5}$, the estimates of d for stock prices range between 1.041 and 1.080 and those for dividends between 1.026 and 1.222. Testing now the homogeneity condition with Robinson and Yajima’s (Robinson and Yajima (2002)) procedure (see Table 6), it is found that the two orders of integration are equal. Here $h(n)$ is set equal to b^{-5-2i} , with $i = 1, 2, 3, 4$ and 5 and $b = (n)^{0.5}$, which is the bandwidth used in the estimation. The Hausmann test of no cointegration (Marinucci and Robinson 2001) (see Table 7) indicates that the estimates of d for the individual series using the bivariate representation (\hat{d}_* in (15)) are very close to 1 and not significantly different from 1 (using three different values for s in (15)), but evidence of cointegration is only obtained in one case out of the six considered (H_{as} with $s = 25$ —see Table 7).

3.2.2 Real stock market prices and real dividends

The same relationship as above, but in real terms is examined in this subsection. A time series plot of the two series is displayed in Fig. 2b. They exhibit a similar pattern to the previous case, although with more volatility in the early part of the sample, and may have a common stochastic trend. Starting again with the univariate tests (see Table 4), it is found that, when applying the Whittle semiparametric method of Robinson (1995), for $m = (n)^{0.5} = 4l$, the estimates of d are 0.888 and 0.896, respectively, for real stock prices and real dividends, and the unit root null cannot be rejected for either series. Similar evidence is obtained with the log-periodogram estimator (see Table 5), with values of d ranging from 0.972 and 1.085 for real stock prices and from 0.822 and 0.997 for real dividends. The test of homogeneity of the orders of integration (Table 6) implies equality in the values of d , while testing the null of no cointegration with the Hausman test of Marinucci and Robinson (2001) (in Table 7) suggests that the two series might be cointegrated.

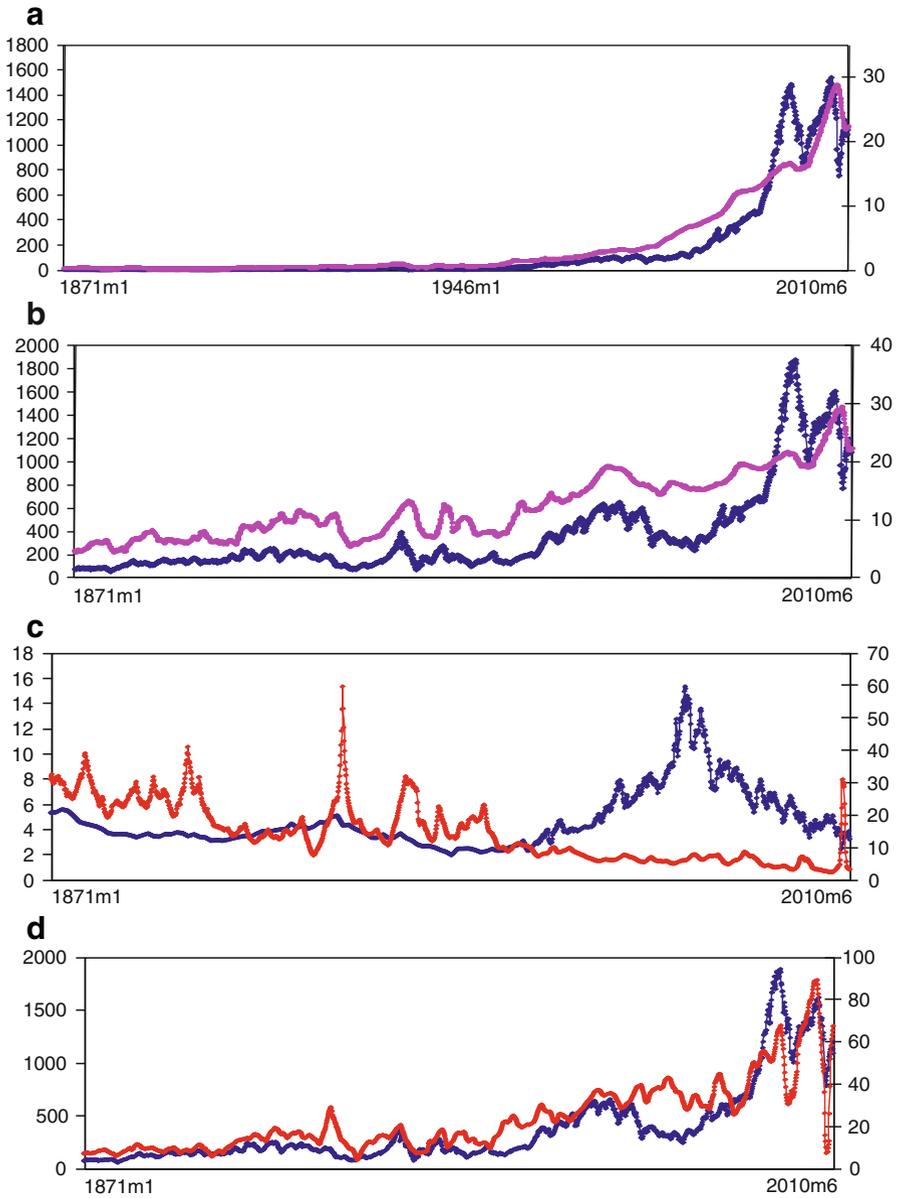


Fig. 2 **a** Stock market prices and dividends. The *thick line* refers to the stock market prices and the *thin one* is for dividends. **b** Real stock market prices and real dividends. The *thick line* refers to real stock market prices and the *thin one* to real dividends. **c** Price-earning ratio and long interest rate. The *thick line* refers to the long-term interest rate and the *thin one* is for the price-earning ratio. **d** Real stock market prices and real earnings. The *thick line* refers to the real stock market prices and the *thin one* is for real earnings

3.2.3 Price/earning ratio and long-term interest rates

These two series are plotted in Fig. 2c. Interest rates appear to be more stable than the price/earning ratio during the first half of the sample; however, during the second half, there is a sharp increase in interest rates, but not in the price/earning ratio. As for the Whittle estimates of d (see Table 4), it is found that for the price/earning ratio the values of d are very sensitive to the bandwidth parameter: for small values (e.g. 25, 41 or 100) the unit root is rejected in favour of values of d below 1; on the contrary, the unit root null cannot be rejected for $m = 200$, and it is rejected in favour of $d > 1$ for $m = 300$ and 500. For the long-term interest rates, the results are more stable and the unit root null cannot be rejected for any bandwidth parameter. These results are corroborated by the log-periodogram estimates, displayed in Table 5. Thus, for the price/earning ratio, different results are obtained depending on whether or not the series is first-differenced, while for long-term interest rates the evidence strongly support the $I(1)$ case. Interestingly, when performing the homogeneity tests of Robinson and Yajima (2002) we cannot reject the null of equal orders of integration, and the Hausman test reject in all cases the null hypothesis of no cointegration in favour of fractional cointegration (Table 7).

3.2.4 Real stock market prices and real earnings

Plots of the two series are displayed in Fig. 2d. They both have a very similar upward trend, which suggests that they may be cointegrated around a common stochastic trend. The estimated values of d using the Whittle method and for $m = (n)^{0.5}$ (see Table 4) are 1.071 for real stocks and 0.933 for real earnings, and in both cases we cannot reject the null of $I(1)$ series. The same evidence in favour of unit roots is obtained with the log-periodogram estimates in Table 5, and the homogeneity restriction cannot be rejected in any single case (see Table 6). The Hausman tests of Marinucci and Robinson (2001) also indicate that the two series might be cointegrated, since the null hypothesis of no cointegration is rejected in all cases in Table 7 in favour of long-memory cointegrating errors.

We estimated the cointegrating coefficients for each of the four relations using various methods such as OLS' (Eq. 7); OLS^f ((8)) and NBLs with different bandwidths (9); the results were similar for all procedures. On the basis of these coefficients, we estimated the orders of integration in the residuals of the cointegrating regression. First, we used the parametric approach of Robinson (1994). However, the results varied considerably depending on the specification of the error term. Owing to this disparity, we estimate d with semiparametric methods.

Table 8 displays the estimates of d based on the log-periodogram regression estimator of Robinson (1995) for $m = n^{0.5}$ and $l = 0$ and $l = 2$. In many cases, the estimates are significantly smaller than l , especially for the price/earning ratio—long-term interest rates and real stock prices—real earning relationships. Table 9 reports the results from the semiparametric Whittle method of Robinson (1995), again applied to the estimated residuals from the cointegrating relationships. Two different bandwidth parameters, $m = 25$ and $m = n^{0.5} = 4l$ are considered. Virtually all estimated values are strictly below 1. For the first two relationships (stock prices and dividends and their

Table 8 Estimates of d on the residuals using the estimate of Robinson (1995)

$m = n^{0.5}$ $l = 0$	Stock prices/ dividends		Real stock prices/ real dividends		P.E.R./Long interest rates		Real stock prices/ real earnings	
	$l = 0$	$l = 2$	$l = 0$	$l = 2$	$l = 0$	$l = 2$	$l = 0$	$l = 2$
\tilde{v} (OLS) time domain	0.775	0.991	1.082	1.273	0.811	0.650	0.723	0.805
\tilde{v} (OLS) freq. domain	0.633	0.617	0.976	1.127	0.828	0.697	0.526	0.430
\tilde{v} (NBLS) ($m = 25$)	1.012	1.260	1.112	1.304	0.837	0.725	1.129	1.283
\tilde{v} (NBLS) ($m = 41$)	1.126	1.386	1.132	1.326	0.860	0.781	1.042	1.190
\tilde{v} (NBLS) ($m = 100$)	1.083	1.337	1.128	1.322	0.878	0.827	0.600	0.623
\tilde{v} (NBLS) ($m = 200$)	0.942	1.186	1.103	1.294	0.855	0.766	0.598	0.620
\tilde{v} (NBLS) ($m = 300$)	0.832	1.066	1.087	1.277	0.859	0.777	0.560	0.557
\tilde{v} (NBLS) ($m = 400$)	0.754	0.956	1.077	1.269	0.853	0.763	0.551	0.535
\tilde{v} (NBLS) ($m = 500$)	0.708	0.853	1.070	1.263	0.846	0.744	0.532	0.486

Table 9 Estimates of d on the residuals using the estimate of Robinson (1995)

$m = 25$ m	Stock prices/ dividends		Real stock prices/ real dividends		P.E.R. / Long interest rates		Real stock prices/ real earnings	
	25	41	25	41	25	41	25	41
\tilde{v} (OLS) time domain	0.606	0.778	0.733	0.803	0.500	0.612	0.610	0.715
\tilde{v} (OLS) freq. domain	0.608	0.781	0.709	0.777	0.500	0.581	0.500	0.500
\tilde{v} (NBLS) ($m = 25$)	0.663	0.827	0.741	0.814	0.500	0.567	0.878	1.047
\tilde{v} (NBLS) ($m = 41$)	0.683	0.815	0.744	0.819	0.500	0.543	0.843	0.987
\tilde{v} (NBLS) ($m = 100$)	0.676	0.811	0.743	0.817	0.500	0.526	0.515	0.623
\tilde{v} (NBLS) ($m = 200$)	0.648	0.796	0.739	0.811	0.500	0.549	0.514	0.621
\tilde{v} (NBLS) ($m = 300$)	0.622	0.784	0.735	0.805	0.500	0.545	0.500	0.562
\tilde{v} (NBLS) ($m = 400$)	0.599	0.775	0.730	0.800	0.500	0.551	0.500	0.521
\tilde{v} (NBLS) ($m = 500$)	0.582	0.770	0.726	0.794	0.500	0.559	0.500	0.500

real terms), the values for the order of integration in the residuals range between 0.6 and 0.8. Smaller values are obtained for the price/earning ratio—long-term interest rate relationship: if $m = 41$, the estimated value of d is about 0.55, however, using $m = 25$, the values are in all cases 0.50 suggesting that the residual series may be stationary. There is a wider range of values in the case of the real stock prices—real earnings relationship, although most of them are also in the interval (0.5, 1).

Finally, we identify parametric models for $f(\lambda)$ with u_t in (5) and (6) on the basis of Eqs. (17–19), using wide-ranging values for the orders of integration from the previous tables. Here, we employ both the Robinson and Hualde (2003) and Hualde and Robinson (2007) approaches based on the approximate difference between the order of integration of the parent series and the estimated residuals. Using a Box–Jenkins-type methodology we identified at most AR(1) structures in all cases. Therefore, we simply consider combinations of white noises and AR(1) processes in each bivariate

Table 10 Estimates of the orders of integration of the individual series and of the cointegrating regression

Model	$\tilde{\nu}$	\tilde{d}	$\tilde{\gamma}$
(a) Stock market prices and dividends			
u_{1t} is white noise	52.754	1.161	0.773
u_{2T} is white noise			
u_{1t} is AR(1)	48.829	1.166	0.795
u_{2T} is white noise			
u_{1t} is white noise	48.829	1.188	0.609
u_{2T} is AR(1)			
u_{1t} is AR(1)	48.792	1.151	0.631
u_{2T} is AR(1)			
(b) Real stock market prices and real dividends			
u_{1t} is white noise	52.754	1.161	0.773
u_{2T} is white noise			
u_{1t} is AR(1)	48.829	1.166	0.795
u_{2T} is white noise			
u_{1t} is white noise	48.829	1.188	0.609
u_{2T} is AR(1)			
u_{1t} is AR(1)	48.792	1.151	0.631
u_{2T} is AR(1)			
(c) Price-earning ratio and long interest rate			
u_{1t} is white noise	57.435	1.047	0.780
u_{2T} is white noise			
u_{1t} is AR(1)	52.251	1.036	0.763
u_{1T} is white noise			
u_{1t} is white noise	52.249	0.996	0.526
u_{2T} is AR(1)			
u_{1t} is AR(1)	52.208	1.159	0.878
u_{2T} is AR(1)			
(d) Real stock market prices and real earning			
u_{1t} is white noise	-1.566	1.165	0.779
u_{2T} is white noise			
u_{1t} is AR(1)	0.874	1.053	0.763
u_{2T} is white noise			
u_{1t} is white noise	0.876	1.115	0.527
u_{2T} is AR(1)			
u_{1t} is AR(1)	0.874	1.153	0.877
u_{2T} is AR(1)			

relation. For each model, we apply the univariate Whittle procedure of [Velasco and Robinson \(2000\)](#), using untapered versions, and, as usual, the first-differenced data, then adding 1 to the estimated value. The results for the four bivariate relationships are summarised in [Table 10](#) and are fairly similar for the different types of $I(0)$ errors.

Although we do not report it, we also estimated a multivariate version of the [Bloomfield \(1973\)](#) model for $I(0)$ autocorrelation, with fairly similar results to those presented in [Table 10](#). In general, there is a statistically significant reduction in the order of integration in all cases of about 0.3/0.4 from the original series to the cointegrating relationship. The orders of integration in the latter are about 0.7 for three of these relations: stock prices/dividends; real prices/real dividends and real prices/real earnings. For the price-earning ratio/interest rates relationship, the reduction is slightly bigger, and the order of integration of the cointegrating relationship seems to be slightly above 0.5.⁶

Overall, the four relationships examined in this paper appear to be fractionally cointegrated, with orders of integration for the individual series equal or slightly above 1, and being in the interval $[0.5, 1)$ for the cointegrating regression, which implies a slow mean-reverting behaviour in the long run.

4 Conclusions

In this paper we have examined bivariate relationships among various financial variables using fractional integration and cointegration methods. In particular, we focus on the following bivariate relationships: stock prices and dividends; real stock prices and real dividends; price/earning ratio and long-run interest rates and real stock prices and real earnings, monthly, for the time period 1871m1–2010m6.

The univariate results strongly support the hypothesis that all individual series are non-stationary with orders of integration equal to or higher than 1 in practically all cases. The multivariate results provide evidence of fractional cointegration for the four bivariate relationships with the orders of integration of the cointegrating regressions being in the interval $[0.5, 1)$ which implies mean-reverting behaviour. The implication is that there exist long-run equilibrium relationships consistent with economic theory and that the effects of shocks are temporary, although the fact that fractional cointegration (rather than standard cointegration) holds means that the adjustment process is much slower, and that, therefore, the overall costs of deviations from equilibrium are bigger than standard cointegration approaches would estimate. This is an important result that should be taken into account when formulating policies and deciding on policy actions. It also provides an explanation for the mixed evidence reported in other papers only allowing for integer degrees of differentiation and therefore not modelling long-memory properties. However, it is important to mention that due to the variety of methods employed, some of them parametric and others semiparametric, along with the sensitivity of the results obtained with the semiparametric methods to the bandwidth parameters, the evidence presented in this study is not entirely conclusive. This is something one has to face when working with fractional models in finite samples due to fact that the differencing parameter has real values.

Other recently developed bivariate or multivariate fractional cointegration testing methods based on co-fractional VAR models (e.g. [Johansen 2011](#); [Nielsen 2010](#);

⁶ Performing the analysis of the bivariate relationships in the opposite direction leads essentially to the same conclusions in all cases.

Nielsen and Frederiksen 2011) could also be applied. Moreover, our analysis does not take into account other possible features of the data, such as structural breaks, non-linearities and other issues. Of course, these are also important issues whose relevance for fractional integration tests has already been investigated (see, e.g. Diebold and Inoue 2001; Granger and Hyung 2004; Caporale and Gil-Alana 2008). Our future research will consider them in the context of fractional cointegration.

Acknowledgments Comments from the Editor and two anonymous referees are gratefully acknowledged. Luis A. Gil-Alana gratefully acknowledges financial support from the Ministerio de Ciencia y Tecnología (ECO2012-2014, No. 28196 ECON Y FINANZAS, Spain) and from a Jeronimo de Ayanz project of the Government of Navarra.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Arteche J (2002) Semiparametric robust tests on seasonal and cyclical long memory series. *J Time Ser Anal* 23:1–35
- Arteche J, Robinson PM (2000) Semiparametric inference in seasonal and cyclical long memory processes. *J Time Ser Anal* 21:1–25
- Asness CS (2003) Fight the fed model. *J Portfolio Manag* 30:11–24
- Baillie R, Bollerslev T (1989) Common stochastic trends in a system of exchange rates. *J Fin* 44:167–181
- Beran J (1995) Maximum likelihood estimation of the differencing parameter for invertible short and long memory ARIMA models. *J R Stat Soc, Ser B* 57:659–672
- Bloomfield P (1973) An exponential model in the spectrum of a scalar time series. *Biometrika* 60:217–226
- Campbell JY, Shiller RJ (1987) Cointegration and tests of present value models. *J Political Econ* 95:1062–1088
- Campbell JY, Shiller RJ (1988) The dividend–price ratio and expectations of future dividends and discount factors. *Rev Fin Stud* 1:195–228
- Campbell JY, Shiller RJ (1998) Valuation ratios and the long-run stock market out-look. *J Portfolio Manag* 24:11–26
- Campbell JY, Shiller RJ (2001) Valuation ratios and the long-run stock market out-look: an update. Cowles Foundation Discussion Paper No. 1295, Yale University
- Caporale GM, Gil-Alana LA (2004) Fractional cointegration and tests of present value models. *Rev Fin Econ* 13(3):245–258
- Caporale GM, Gil-Alana LA (2008) Modelling structural breaks in the US, UK and Japanese unemployment. *Comput Stat Data Anal* 52(11):4998–5013
- Crowder W (1996) A note on cointegration and international capital market efficiency: a reply. A reply, *J Int Money Fin* 15:661–664
- DeJong DN (1992) Co-integration and trend-stationarity in macroeconomic time series: evidence from the likelihood function. *J Econ* 52:347–370
- Diebold FX, Inoue A (2001) Long memory and regime switching. *J Econ* 105:131–159
- Durre A, Giot P (2007) An international analysis of earnings, stock prices and bond yields. *J Bus Fin Account* 34(3–4):613–641
- Engle RF, Granger CWJ (1987) Cointegration and error correction model. Representation, estimation and testing. *Econometrica* 55:251–276
- Gil-Alana LA (2004) The use of the model of Bloomfield (1973) as an approximation to ARMA processes in the context of fractional integration. *Math Comp Model* 39:429–436
- Gil-Alana LA, Hualde J (2009) In: Mills TC, Patterson K (eds) Fractional integration and cointegration. An overview with an empirical application. The Palgrave handbook of applied econometrics, vol 2. MacMillan Publishers, New York, pp 434–472

- Granger CWJ, Hyung N (2004) Occasional structural breaks and long memory with an application to the S&P 500 absolute stock returns. *J Empir Fin* 11:399–421
- Hakkio G, Rush M (1989) Market efficiency and cointegration. the Sterling and Deutschmark exchange markets. *J Int Money Fin* 8:75–88
- Hassler U, Rodrigues PMM, Rubia A (2009) Testing for the general unit root hypothesis in the time domain. *Econ Theory* 25:1793–1828
- Henry M (2001) Robust automatic bandwidth for long memory. *J Time Ser Anal* 22:293–316
- Hualde J, Robinson PM (2007) Root-n-consistent estimation of weak fractional cointegration. *J Econ* 140:450–484
- Johansen S (2011) An extension of cointegration to fractional autoregressive processes, CREATES Research Papers 2011-06. School of Economics and Management, University of Aarhus
- Koop G (1991) Cointegration tests in present value relationships. Stock prices and dividends. *J Econ* 49:105–139
- Marinucci D, Robinson PM (1999) Alternative forms of fractional Brownian motion. *J Stat Plan Inference* 80:111–122
- Marinucci D, Robinson PM (2001) Semiparametric fractional cointegration analysis. *J Econ* 105:225–247
- Marinucci D, Robinson PM (2001) Narrow-band analysis of nonstationary processes. *Ann Stat* 29(4):947–986
- Nielsen MO (2010) Nonparametric cointegration analysis of fractional systems with unknown integration orders. *J Econ* 155:1701–1787
- Nielsen MO, Frederiksen P (2011) Fully modified narrow-band least squares estimation of weak fractional cointegration. *Econ J* 14:77–120
- Pereira-Garmendia D (2010) Inflation, real stock prices and earnings: Friedman was right. Working Paper, University Pompeu Fabra, Barcelona
- Phillips PCB (1991) Optimal inference in cointegrating systems. *Econometrica* 59:283–306
- Philips TK (1999) Why do valuation ratios forecast long-run equity returns? *J Portfolio Manag* 25:39–44
- Rangvid J (2001) Increasing convergence among European stock markets? A recursive common stochastic trends analysis. *Econ Lett* 71:383–389
- Richards AJ (1995) Comovements in national stock market returns. Predictability but not cointegration. *J Money Econ* 36:631–654
- Robinson PM (1994a) Efficient tests of nonstationary hypotheses. *J Am Stat Assoc* 89:1420–1437
- Robinson PM (1994b) Semiparametric analysis of long memory time series. *Ann Stat* 22:515–539
- Robinson PM (1995a) Gaussian semi-parametric estimation of long range dependence. *Ann Stat* 23:1630–1661
- Robinson PM (1995b) Log-periodogram regression of time series with long range dependence. *Ann Stat* 23:1048–1072
- Robinson PM, Hualde J (2003) Cointegration in fractional systems with unknown integration orders. *Econometrica* 71:1727–1766
- Robinson PM, Yajima Y (2002) Determination of cointegrating rank in fractional systems. *J Econ* 106:217–241
- Shao X, Wu WB (2007) Local Whittle estimation of fractional integration for nonlinear processes. *Econ Theory* 23:899–929
- Shiller RJ (1989) *Market volatility*. MIT Press, Cambridge
- Sowell F (1992) Maximum likelihood estimation of stationary univariate fractionally integrated time series models. *J Econ* 53(165):188
- Velasco C, Robinson PM (2000) Whittle pseudo maximum likelihood estimation for nonstationary time series. *J Am Stat Assoc* 95:1229–1243