Testing for rational bubbles in the presence of structural breaks: Evidence from nonstationary panels

Roy Cerqueti\textsuperscript{a}, Mauro Costantini\textsuperscript{b,*}

\textsuperscript{a} Department of Economics and Financial Institutions, University of Macerata, Italy.
\textsuperscript{b} Department of Economics, University of Vienna, Austria.

This version: 15 February 2011

Abstract

This paper presents new results on the rational bubbles hypothesis for a panel of 18 OECD countries using the model developed by Campbell (2000). We provide an analysis of international data that exploits increased power deriving from the panel unit root and cointegration methodology, together with the flexibility of allowing explicitly for multiple endogenous structural breaks in the individual series. Differently from the time series methodology, the panel data approach allows for a global analysis of the financial crashes that are related to rational bubbles. We find strong evidence in favor of bubbles phenomena.

\textit{JEL Classification:} C23; G15
\textit{Keywords:} Rational bubbles; International financial markets; Panel data; Unit root; Cointegration

\* Corresponding author: Tel.: +43 1 4277 37478; Fax: +43 1 4277-37498.
\textit{Email addresses:} roy.cerqueti@unimc.it (R. Cerqueti), mauro.costantini@univie.ac.at (M. Costantini)
1. Introduction

Since the 1980s there has been considerable interest in studying the phenomena of rational bubbles in stock prices. A number of studies argue that dividends and stock prices data are not consistent with the “market fundamentals hypothesis”. Diba and Grossman (1984) propose an empirical strategy of testing for rational bubbles based on time series unit root and cointegration tests. Time series analysis has become increasingly popular since Diba and Grossman’s paper and empirical evidence on rational bubbles is still mixed due to the use of different unit root and cointegration tests and to the individual stock markets considered.


Although most of the studies are based on the US stock data, empirical evidence using different data is also relevant. Using data for the UK stock market, Brooks and Katsaris (2003) and Capelle-Blanchard and Raymond (2004) find that no rational bubbles phenomena is observed when dividend-price ratio is considered, while McMillan (2007) provides evidence in favour of a unit root process in the log dividend-price ratio. Mixed results are also found when other important European countries such as France and Germany are considered. Some studies also investigate the bubbles phenomena using data for the East Asian and Latin American stock markets (see among other, Sarno and Taylor (1999), Herrera and Perry (2003) and Sarno and Taylor (2003)). Once again mixed results are found depending on which unit root and cointegration tests are applied.
In this paper we extend the time series approach proposed by Herrera and Perry (2003) to panel data. The general idea is to verify or reject the existence of a stable relationship among stock prices, dividends, and returns in the present-value model proposed by Campbell (2000). The analysis consists in two steps. In the first one, we perform unit root tests on log dividend-price ratio and real return variables. If the series have a unit root, the “no bubble” hypothesis is rejected. In the second step, we apply cointegration tests to investigate the existence of a long-run relationship between log dividend-price ratio and stock returns. If a stable relationship is rejected, then the “no bubble” hypothesis is also rejected.

This paper makes an important contribution to the existing literature on rational bubbles using a panel of 18 OECD countries. In this regard, a panel data approach allows for a global analysis of the financial crashes related to rational bubbles since it considers information across different countries. In particular, we apply panel unit root and cointegration tests with breaks under cross-section dependence hypothesis. The use of panel data is generally considered as a mean of generating more powerful tests with respect to the univariate counterpart. In this respect, the small sample econometric problems of testing for bubbles in times series can be avoided. We consider the cross-section dependence hypothesis in order to deal with integrated world stock markets.\(^1\) Since the beginning of 1986, the major stock markets have become increasingly internationalized by deregulation. By 1987, some of the 600 foreign stocks traded in New York market, and the markets in London, Frankfurt and Tokyo have also attracted numerous foreign listings. The simultaneous price collapse around the world after the US Stock Market crash in October 1987 as well as the contagion effect occurred after Mexican Peso crisis (1994) and East Asian crisis (1997) has shown a strong evidence of the linkages between national stock markets.

We also address the issues of nonstationarity of the log dividend-price ratio and real returns and of cointegration relationship between these variables by allowing for structural breaks in the data. Such issues to be investigated are motivated by economic and statistical reasons. Fair (2004) shows that there is only one major structural change in US economy in the second half of the

\(^1\)Bali and Cakici (2010) also point out the importance of the cross-correlation for integrated markets.
1990s, namely the huge increase in stock prices relative to earnings, and Capelle-Blanchard and Raymond (2004) report that the increase of stock markets prices between 1995 and 2000 is amazing compared to the growth of the dividends for French, German, UK and US stock markets. Perron (1989) shows that the ability to reject the unit root null hypothesis can decrease substantially when the stationarity under the alternative is true but existing structural breaks are ignored. This is important because the way in which traditional unit root testing is carried out typically involves employing time series that span extended periods of time, which obviously increases the probability of a structural break. The implication is that the inability of many empirical studies to reject the unit root null hypothesis may well be due to an erroneous omission of structural breaks. Im et al. (2005) argue that the same problem exists in panel unit root tests. Banerjee and Carrion-i-Silvestre (2006) show that an analogous problem exists for cointegration tests. In fact, standard cointegration tests not allowing for structural changes might lead to biases when testing for the null hypothesis of no cointegration in favor of acceptance.

The paper is organized as follows. Section 2 describes the present-value model. Section 3 presents the econometric methodology. Section 4 contains the data and presents the empirical results. The last section concludes.

2. The model

In this section we describe the present-value model used to study the long-run relations between prices, dividends and returns. If an asset has a constant expected return, then its price is a linear function of its expected future payoffs. The definition of returns provides the following relationship at time $t > 0$ between returns $R_t$, dividends $D_t$ and prices $P_t$

$$1 + R_{t+1} = \frac{P_{t+1} + D_{t+1}}{P_t}.$$
Moreover, if the expected return is a constant $R$, then we write the main equation of the martingale model for stock prices as (see Samuelson (1965)):

$$P_t = \frac{E_t[P_{t+1} + D_{t+1}]}{1 + R},$$  \hspace{1cm} (1)

where $E_t$ indicates the conditional expectation, given the information available up to time $t$. Hence, (1) is an expectational difference equation and we can solve it forward. Let us assume that the expected discounted future price admits limit that is zero, i.e.

$$\lim_{k \to +\infty} E_t[P_{t+k}] = 0.$$  \hspace{1cm} (2)

Then we get

$$P_t = E_t \left[ \sum_{j=1}^{\infty} \frac{D_{t+j}}{(1 + R)^j} \right].$$

The right-hand side of equation (2) is the “fundamental value” of an asset price, and this expression holds only under constant discounted rate condition. Using the transversality condition

$$\lim_{j \to +\infty} E_t \left[ \frac{D_{t+j}}{(1 + R)^j} \right] = 0,$$

equation (2) assures an unique price. If the present value model holds, then real stock price and dividends are cointegrated (see Campbell and Shiller (1987)). The cointegrating vector is $(1, 1/R)$, and we have:

$$P_t - \frac{D_t}{R} = \frac{1}{R} \cdot E_t \left[ \sum_{i=0}^{+\infty} \frac{\Delta D_{i+t+1}}{(1 + R)^i} \right].$$  \hspace{1cm} (3)

Equation (3) suggests that the present value model has to be tested using cointegration between real prices and real dividends. However, evidence for cointegration is mixed (see Campbell and Shiller (1987); Diba and Grossman (1988); Han (1996); Chang et al. (2007), among others). It seems to be more realistic to relax the strong hypothesis of constant discount rate, and prefer rather to allow for time-variation in the discount rate. In this respect, Campbell and Shiller (1988) extend
the linear present-value model to allow for log-linear dividend processes and time-varying discount
rates. They define log return $r$ by approximation, since

$$r_{t+1} = \log(P_{t+1} + D_{t+1}) - \log P_t, \quad t > 0.$$  \hspace{1cm} (4)

Using a first-order Taylor expansion around the mean log dividend-price ratio, $\bar{d}_t - \bar{p}_t$, they obtain

$$r_{t+1} \sim k + \rho p_{t+1} + (1 - \rho)d_{t+1} - p_t,$$

where $\rho$ and $k$ are real parameters defined by

$$\rho = \frac{1}{1 + \exp(\bar{d}_t - \bar{p}_t)},$$

and

$$k = -\log \rho - (1 - \rho) \log \left[ \frac{1}{\rho - 1} \right].$$

When the dividend-price ratio is constant, then we have

$$\rho = \frac{P}{P + D}.$$  

Solving forward, imposing the “no-bubbles” terminal condition that

$$\lim_{j \to +\infty} \rho^j (d_{t+j} - p_{t+j}) = 0,$$

taking expectations, and subtracting the current dividend, one gets

$$d_t - p_t = -\frac{k}{1 - \rho} + \mathbb{E}_t \sum_{j=0}^{\infty} \rho^j [-\Delta d_{t+j+1} + r_{t+j+1}],$$  \hspace{1cm} (5)

where $p_t = \log$ prices, $d_t = \log$ dividends and $r_t = \log$ returns.

If $d_t$ and $p_t$ are both generated by I(1) processes, then equation (5) implies that the log dividend-
price ratio is stationary, $I(0)$, if and only if the return series $r_t$ are $I(0)$ process. Campbell et al. (1997) point out that $r_t$ can be in practice generated by a highly persistence process which is difficult to distinguish from a nonstationary $I(1)$ process. Therefore, testing for stationarity of the log dividend-price ratio may be problematic in the varying-returns model. Rearranging equation (5), we obtain

$$d_t - p_t - \frac{1}{1-\rho} r_t = -\frac{k}{1-\rho} + E_t \sum_{j=0}^{\infty} \rho^j [-\Delta d_{t+j+1} + \Delta r_{t+j+1}]$$

which implies that the problem can be ameliorated by testing for cointegration between the log dividend-price ratio and stock returns and, even if $r_t$ can be only highly persistent rather then a strictly $I(1)$ process, the left-hand side of equation (6) should be a stationary process in absence of bubbles. Therefore, tests on rational bubbles are oriented towards investigating the stationarity of the log dividend-price ratio and the stock returns and the existence of a stable relationship among the log dividend-price ratio and the stock returns (see Sarno and Taylor (1999) and Herrera and Perry (2003)). Accordingly, we perform two types of tests. First, we investigate for unit roots in the log dividend-price ratio and in the real return series using the panel unit root tests of Bai and Carrion-i-Silvestre (2009). If the series follow a unit root process, the “no bubble” hypothesis is rejected. Second, we check for cointegration between log dividend-price ratio and returns by applying the panel cointegration tests developed by Westerlund and Edgerton (2008). If a stable (equilibrium) relationship is rejected, then the “no bubble” hypothesis is also rejected.

3. **Econometric methodology**

In this section, we discuss the econometric techniques used in the empirical analysis. We first present the panel unit root tests developed by Bai and Carrion-i-Silvestre (2009) and then the panel cointegration tests proposed by Westerlund and Edgerton (2008).
3.1. Panel unit root test

Bai and Carrion-i-Silvestre (2009) propose panel unit root statistics that pools the modified Sargan-Bhargava (hereafter MSB) tests for individual series taking into account structural breaks and cross-dependence through a common factors model proposed by Bai and Ng (2004). Structural breaks are assumed to affect the level or the slope or both the level and the slope of a time series and the break points can be located at different dates across individual series. The common factors may be non-stationary processes, stationary processes or a combination of both. \( I(0) \) common factors have the interpretation of common shocks, while \( I(1) \) common factors represent unobservable global stochastic trends.

Bai and Carrion-i-Silvestre (2009) show that for the case where the structural breaks only occur in the mean, the procedure of Bai and Ng (2004) can be applied without using any modification. When structural breaks affect the slope, the method of Bai and Ng (2004) cannot be applied. In this regard, Bai and Carrion-i-Silvestre (2009) develop an iterative estimation procedure that is suitable for handling heterogenous breaks in the deterministic components.

Bai and Carrion-i-Silvestre (2009) consider the following panel data model:

\[
X_{it} = D_{it} + \mathbf{F}_t' \pi_i + e_{it} \quad (7)
\]

\[
(1 - L) \mathbf{F}_t = C(L) u_t \quad (8)
\]

\[
(1 - \rho_i L) e_{it} = H_i(L) \varepsilon_{it}, \quad (9)
\]

\( t = 1, \ldots, T \) and \( i = 1, \ldots, N \), where \( C(L) = \sum_{j=0}^{\infty} C_j L^j \), \( H_i(L) = \sum_{j=0}^{\infty} C_{ij} L^j \), \( L \) is lag operator, and \( \rho_i \) is the autoregressive parameter in the univariate model (see parameter \( \rho \) in equation (24) in the Appendix). The component \( D_{it} \) indicates the deterministic part of the model, \( \mathbf{F}_t \) is an \((r \times 1)\) vector that accounts for the common factors of the panel, \( e_{it} \) is the idiosyncratic error term, \( \mu_t \sim i.i.d.(0, \Sigma_\mu) \) and \( \varepsilon_{it} \sim i.i.d.(0, \Sigma_{\epsilon_i}) \). Despite the operator \((1 - L)\) in equation (8), \( \mathbf{F}_t \) does not have to be \( I(1) \). In this regard, \( \mathbf{F}_t \) can be \( I(0) \), \( I(1) \), or a combination of both, depending on the rank of \( C(1) \). If \( C(1) = 0 \), then \( \mathbf{F}_t \) is \( I(0) \). If \( C(1) \) is of full rank, then each component of \( \mathbf{F}_t \) is \( I(1) \). If
C(1) = 0 but not full rank, then some components of \( \mathbf{F}_t \) can be \( I(1) \) and others \( I(0) \).\(^2\) As regards the deterministic component, \( D_{it} \), Bai and Carrion-i-Silvestre (2009) propose two specifications:

\[
\text{Model 1: } D_{it} = \mu_i + \sum_{j=1}^{l_i} \theta_{ij} D_{ijt} \\
\text{Model 2: } D_{it} = \mu_i + B_i t + \sum_{j=1}^{l_i} \theta_{ij} D_{ijt} + \sum_{k=1}^{m_i} \gamma_{ik} D_{ikt}
\]

(10)

(11)

where \( l_i \) and \( m_i \) denote the structural breaks affecting the mean and the trend of a time series, respectively, and \( l_i \) is not necessarily equal to \( m_i \). The dummy variables are defined as follows:

\( D_{ijt} = 1 \) for \( t > T^i_{aj} \) and 0 elsewhere, and \( D_{ikt} = (t - T^i_{bk}) \) for \( t > T^i_{bk} \) and 0 elsewhere. \( T^i_{aj} \) and \( T^i_{bk} \) indicate the \( j-th \) and \( k-th \) dates of the break in the level and in the trend, respectively, for the \( i-th \) individual, with \( j = 1, \ldots, l_i \) and \( k = 1, \ldots, m_i \). Bai and Carrion-i-Silvestre (2009) propose to combine individual MSB test statistics to test the null hypothesis of \( \rho_i = 1 \) for all \( i = 1, \ldots, N \) against the alternative \( |\rho_i| < 1 \) for some \( i \). This approach is suitable since \( e_{it} \) are cross-sectionally independent (the individual statistics are free from the common factors). Bai and Carrion-i-Silvestre (2009) provide two approaches for pooling individual statistics. The first approach is based on the use of the average of the individual statistics:

\[
Z = \sqrt{N} \frac{\overline{\text{MSB}}(\lambda)}{\xi} \overset{d}{\sim} N(0, 1)
\]

(12)

with \( \overline{\text{MSB}}(\lambda) = N^{-1} \sum_{i=1}^{N} \text{MSB}_i(\lambda_i) \), \( \overline{\xi} = N^{-1} \sum_{i=1}^{N} \xi_i \) and \( \overline{\xi^2} = N^{-1} \sum_{i=1}^{N} \xi_i^2 \), where \( \xi_i \) and \( \xi_i^2 \) denote the mean and the variance of the individual modified Sargan-Bhargava (\( \text{MSB}_i(\lambda_i) \)) statistics respectively and \( \lambda_i = T^i_{b}/T \) the break fraction parameter.\(^3\) In order to purge the break fraction parameter in the limiting distributions in the case of Model 2, Bai and Carrion-i-Silvestre (2009)

\(^2\)For further details on assumptions regarding the panel data model see Bai and Carrion-i-Silvestre (2009).

\(^3\)See the Appendix for a description of the individual MSB statistics.
propose another test based on the simplified MSB statistics:

\[ Z^* = \sqrt{N} \frac{\overline{MSB}^*(\lambda) - \bar{\xi}^*}{\bar{\varsigma}^*} \xrightarrow{d} N(0, 1) \] (13)

with \( \overline{MSB}^*(\lambda) = N^{-1} \sum_{i=1}^{N} MSB^*_i(\lambda_i) \), \( \bar{\xi}^* = N^{-1} \sum_{i=1}^{N} \xi^*_i \) and \( \bar{\varsigma}^* = N^{-1} \sum_{i=1}^{N} \varsigma^*_i \), where \( \xi^*_i \) and \( \varsigma^*_i \) denote the the mean and the variance of the individual \( MSB^*_i(\lambda_i) \) statistics, respectively.\(^4\)

The second approach is based on the method developed by Maddala and Wu (1999) and Choi (2001) that pools the p-values associated with the individual tests:

\[ P = -2 \sum_{i=1}^{N} \ln p_i \xrightarrow{d} \chi^2_N \] (14)

\[ P_m = -2 \sum_{i=1}^{N} \ln p_i - 2N \frac{d}{\sqrt{4N}} \xrightarrow{d} N(0, 1) \] (15)

where \( p_i \) denotes the individual p-value. Bai and Carrion-i-Silvestre (2009) also propose a version of \( P \) and \( P_m \) tests based on the p-values of the individual simplified MSB. They are denoted as \( P^* \) and \( P_m^* \), respectively.\(^5\)

### 3.2. Panel Cointegration tests

Westerlund and Edgerton (2008) propose two versions of a simple test for the null hypothesis of no cointegration that can be used under very general condition (heteroskedastic and correlated errors, individual-specific intercepts and time trend, cross-section dependence and unknown breaks both in the intercept and slope of the cointegrated regression). The test derives from the Lagrange multiplier (LM)-based unit root tests (see i.e. Schmidt and Phillips (1992)). Westerlund and Edger-
ton (2008) consider the following model:

\[ y_{it} = \alpha_i + \eta_{it} + \delta_i D_{it} + x_{it}' \beta_i + (D_{it} x_{it})' \gamma_i + z_{it}, \]  
(16)

\[ x_{it} = x_{it-1} + \omega_{it}, \quad t = 1, \ldots, T; \quad i = 1, \ldots, N. \]  
(17)

\( x_{it} \) contains the regressors and is modelled as a pure random walk. The variable \( D_{it} \) is a scalar break dummy such that \( D_{it} = 1 \) if \( t > T_i^b \) and zero otherwise. \( \alpha_i \) and \( \beta_i \) represent the intercept and slope before the break, while \( \delta_i \) and \( \gamma_i \) represent the change in these parameters at the time of the shift. \( \omega_{it} \) is an error process with mean zero and independent across \( i \). In equation (16), the error term \( z_{it} \) is generated by the following model:

\[ z_{it} = \lambda_i' F_t + \nu_{it}, \]  
(18)

\[ F_{jt} = \rho_j F_{jt-1} + u_{jt} \]  
(19)

\[ \phi_i(L) \Delta \nu_{it} = \phi_i \nu_{it-1} + e_{it}, \]  
(20)

where \( \phi_i(L) = 1 - \sum_{j=1}^{p_i} \phi_{ij} L^j \) is a scalar polynomial in the lag operator \( L \), \( F_t \) is an \( r \)-dimensional vector of unobservable common factors \( F_{jt} \) with \( j = 1, \ldots, r \), \( \lambda_i \) is a conformable vector of loading parameters, \( u_{jt} \) is independent of \( e_{it} \) and \( \omega_{it} \) for all \( i, j \) and \( t \) and \( e_{it} \) is an error term with mean zero and independent across both \( i \) and \( t \). By assuming that \( \rho_j < 1 \) for all \( j \), we ensure that \( F_t \) is strictly stationary, which implies that the order of integration of the composite regression error \( z_{it} \) depends only on the integration of the idiosyncratic disturbance \( \nu_{it} \). Thus, in this data generating process, the relationship in (16) is cointegrated if \( \phi_i < 0 \) and it is spurious if \( \phi_i = 0 \). The null hypothesis to be tested is that all \( N \) units are spurious, while the alternative hypothesis is that the first \( N_1 \) units are cointegrated while the remaining \( N_0 = N - N_1 \) units are spurious. In other words, we test the null of \( H_0 : N_1 = 0 \) against the alternative of \( H_1 : N_1 > 0 \).\(^6\) The hypothesis of \( H_0 \) vs. \( H_1 \) can be

\(^6\)Westerlund and Edgerton (2008) argue that the assumption that the cointegrated units lie first is only for notational simplicity, and is by no means restrictive.
tested using the LM principle that the score vector has zero mean when evaluated at the vector of true parameters under the null. Westerlund and Edgerton (2008) propose the following tests:

\[ Z_\phi(N) = \sqrt{N} \left( \text{LM}_\phi(N) - E(B_\phi) \right) \]
\[ Z_\tau(N) = \sqrt{N} \left( \text{LM}_\tau(N) - E(B_\tau) \right), \]

where \( \text{LM}_\phi = \frac{1}{N} \sum_i^N \text{LM}_\phi(i) \), \( \text{LM}_\tau = \frac{1}{N} \sum_i^N \text{LM}_\tau(i) \), \( \text{LM}_\phi(i) = T \hat{\phi}_i(\hat{\omega}_i/\hat{\sigma}_i) \), \( \text{LM}_\tau(i) = \frac{\hat{\delta}_i}{SE(\hat{\phi}_i)} \), \( \hat{\phi}_i \) is the least square estimates of \( \phi_i \) in the equation (9) in Westerlund and Edgerton (2008), \( \hat{\sigma}_i \) and \( SE(\hat{\phi}_i) \) are the estimated standard errors of the same regression (9).\(^7\)

4. Data and empirical results

We take monthly data on stock prices indexes and corresponding dividends over the period 1992:1-2010:6 for a panel of 18 OECD countries from Bloomberg. In particular, we collect data for the G7 countries (Canada, France, Germany, Italy, Japan, UK and US), eight European countries (Austria, Belgium, Finland, Greece, Netherlands, Spain, Sweden and Switzerland) and Hong Kong, New Zealand and Mexico.\(^8\) Bloomberg data are suitable for the analysis since they are defined as broad indexes of national stock markets, covering also medium and small companies weighted by their market capitalization rate. As such, they are more likely to proxy the whole equity market as opposed to indexes based on high-capitalization companies. In addition, Bloomberg data are expected to be more homogeneous across markets than local stock price indexes, making our empirical panel results consistent. We compute the real returns and the log dividend-price ratio using formulas (4) and (5), respectively.\(^9\) Table 1 reports the descriptive statistics of the data.

We consider the first four moments of dividend-price ratio and returns. Across the stock markets, the average of the dividend-price ratio varies widely, ranging from a minimum of 0.823 (Japan) to a maximum of 4.763 (New Zealand), and the average of returns varies slightly, ranging

---

\(^7\)For reasons of space, we do not report the procedure for developing the tests in (21). See Section 3 in Westerlund and Edgerton (2008).

\(^8\)We exclude some countries due to the shorter dividend data sample period.

\(^9\)We calculate the dividend-price ratio using the Market Convention Trailing 12-month dividends, which are computed by taking the sum of all members’ last 12-month dividends times shares, and the last price.
Table 1
Descriptive statistics. Dividend-price ratio and returns.

<table>
<thead>
<tr>
<th>Countries</th>
<th>Dividend-price ratio</th>
<th>Returns</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std. Dev.</td>
</tr>
<tr>
<td>Austria</td>
<td>1.418</td>
<td>1.092</td>
</tr>
<tr>
<td>Belgium</td>
<td>2.918</td>
<td>1.119</td>
</tr>
<tr>
<td>Canada</td>
<td>1.722</td>
<td>0.708</td>
</tr>
<tr>
<td>Finland</td>
<td>2.150</td>
<td>1.573</td>
</tr>
<tr>
<td>France</td>
<td>2.209</td>
<td>1.066</td>
</tr>
<tr>
<td>Germany</td>
<td>2.018</td>
<td>1.029</td>
</tr>
<tr>
<td>Greece</td>
<td>2.133</td>
<td>1.117</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>2.539</td>
<td>0.895</td>
</tr>
<tr>
<td>Ireland</td>
<td>1.956</td>
<td>1.085</td>
</tr>
<tr>
<td>Japan</td>
<td>0.823</td>
<td>0.500</td>
</tr>
<tr>
<td>Mexico</td>
<td>1.229</td>
<td>0.508</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2.767</td>
<td>1.013</td>
</tr>
<tr>
<td>New Zealand</td>
<td>4.763</td>
<td>1.492</td>
</tr>
<tr>
<td>Spain</td>
<td>2.585</td>
<td>1.240</td>
</tr>
<tr>
<td>Sweden</td>
<td>2.045</td>
<td>1.520</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1.494</td>
<td>0.655</td>
</tr>
<tr>
<td>UK</td>
<td>2.871</td>
<td>0.781</td>
</tr>
<tr>
<td>USA</td>
<td>2.001</td>
<td>0.595</td>
</tr>
</tbody>
</table>

from 0.138 (Mexico) to 1.487 (New Zealand). Even in the case of the standard deviation, we observe a more pronounced variation for dividend-price ratio than returns. Specifically, the standard deviation of the dividend-price ratio varies from 0.500 (Japan) to 1.573 (Finland) whereas that of returns varies from 0.267 (UK) to 0.933 (Finland). The asymmetry of the data is described by the skewness. In this respect, we observe that there are positive values in most of the cases for the dividend-price ratio and mixed positive and negative values for returns. The kurtosis is substantially higher than the normal case for most of the countries, implying a leptokurtic distribution for the data. In other terms, the distribution of the data is generally “fat-tailed”. Since we consider the hypothesis of cross-section dependence, we first check for cross-correlation in the data. To this

10 The standard deviation can be informative on the benefit of the diversification in the stock market when it is used as a measure of risk. On this point, see You and Daigler (2010).
end, we apply the CD test developed by Pesaran (2004):

\[
CD = \sqrt{\frac{2T}{N(N-1)}} \left( \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \hat{\rho}_{ij} \right),
\]

where

\[
\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T} e_{it}e_{jt}}{(\sum_{t=1}^{T} e_{it}^2)^{1/2}(\sum_{t=1}^{T} e_{jt}^2)^{1/2}}
\]

denotes the sample estimate of the pair-wise correlation of the residuals \(e_{it}\) from the regression of any variable of interest on an intercept, a linear trend and a lagged dependent variable for each country \(i\). Pesaran (2004) shows that under the null hypothesis of cross-section independence \(CD \overset{d}{\rightarrow} N(0,1)\). With respect to the CD test results, we find evidence of cross-section dependence as the findings for log dividend-price ratio and returns are 63.437 (0.000) and 37.693 (0.000) respectively (p-values are in parenthesis).

In order to further support the need of allowing for cross-section dependence, we also compute the long-run cross-section correlation matrix of the residuals of the cointegration equation (5) (see Table 2). Results show that all the correlations lie between 0.12 and 0.99, with an overall average of 0.76, suggesting the violation of the hypothesis of cross-section independence.

Thus, we must consider a panel nonstationary analysis which allows for cross-section dependence. In order to investigate the nonstationary properties of log dividend-price ratio and returns, we apply the panel unit root tests developed by Bai and Carrion-i-Silvestre (2009). We consider both models in (10) and (11). Table 3 reports the results.

The null hypothesis of unit root cannot be rejected with all tests for both log dividend-price ratio and returns, implying occurrence of bubbles for all countries since under the null hypothesis all the series are generated by a unit root process.\(^{11}\) It is noteworthy that our panel results give stronger evidence of bubbles in the OECD countries than the most of evidence from the univariate

\(^{11}\)Chortareas and Kapetanios (2009) point out that an attractive feature of panel unit root tests is the ability to exploit coefficient homogeneity under the null hypothesis of a unit root for all series considered in order to obtain more powerful tests of the unit root hypothesis.
Table 2
Estimated long-run cross-section correlations of the residuals of cointegration equation.

<table>
<thead>
<tr>
<th>No</th>
<th>Country</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Austria</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Belgium</td>
<td>0.86</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Canada</td>
<td>0.80</td>
<td>0.96</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Finland</td>
<td>0.55</td>
<td>0.49</td>
<td>0.48</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>France</td>
<td>0.73</td>
<td>0.67</td>
<td>0.56</td>
<td>0.29</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Germany</td>
<td>0.12</td>
<td>0.42</td>
<td>0.60</td>
<td>0.12</td>
<td>0.18</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Greece</td>
<td>0.68</td>
<td>0.75</td>
<td>0.88</td>
<td>0.50</td>
<td>0.36</td>
<td>0.69</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Hong Kong</td>
<td>0.83</td>
<td>0.98</td>
<td>0.97</td>
<td>0.44</td>
<td>0.67</td>
<td>0.47</td>
<td>0.79</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Ireland</td>
<td>0.94</td>
<td>0.93</td>
<td>0.87</td>
<td>0.59</td>
<td>0.78</td>
<td>0.20</td>
<td>0.72</td>
<td>0.91</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Japan</td>
<td>0.72</td>
<td>0.91</td>
<td>0.88</td>
<td>0.35</td>
<td>0.60</td>
<td>0.39</td>
<td>0.66</td>
<td>0.90</td>
<td>0.74</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Mexico</td>
<td>0.87</td>
<td>0.95</td>
<td>0.96</td>
<td>0.62</td>
<td>0.63</td>
<td>0.47</td>
<td>0.86</td>
<td>0.94</td>
<td>0.92</td>
<td>0.84</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Netherlands</td>
<td>0.86</td>
<td>0.99</td>
<td>0.95</td>
<td>0.46</td>
<td>0.74</td>
<td>0.35</td>
<td>0.74</td>
<td>0.98</td>
<td>0.94</td>
<td>0.91</td>
<td>0.94</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>New Zealand</td>
<td>0.24</td>
<td>0.61</td>
<td>0.55</td>
<td>0.13</td>
<td>0.15</td>
<td>0.38</td>
<td>0.24</td>
<td>0.59</td>
<td>0.32</td>
<td>0.73</td>
<td>0.42</td>
<td>0.58</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Spain</td>
<td>0.75</td>
<td>0.90</td>
<td>0.93</td>
<td>0.56</td>
<td>0.36</td>
<td>0.65</td>
<td>0.86</td>
<td>0.89</td>
<td>0.83</td>
<td>0.73</td>
<td>0.90</td>
<td>0.85</td>
<td>0.49</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Sweden</td>
<td>0.80</td>
<td>0.97</td>
<td>0.95</td>
<td>0.44</td>
<td>0.69</td>
<td>0.41</td>
<td>0.78</td>
<td>0.97</td>
<td>0.90</td>
<td>0.91</td>
<td>0.93</td>
<td>0.99</td>
<td>0.59</td>
<td>0.86</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Switzerland</td>
<td>0.84</td>
<td>0.86</td>
<td>0.87</td>
<td>0.57</td>
<td>0.59</td>
<td>0.46</td>
<td>0.80</td>
<td>0.89</td>
<td>0.89</td>
<td>0.66</td>
<td>0.89</td>
<td>0.83</td>
<td>0.50</td>
<td>0.86</td>
<td>0.79</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>UK</td>
<td>0.85</td>
<td>0.99</td>
<td>0.96</td>
<td>0.47</td>
<td>0.71</td>
<td>0.39</td>
<td>0.77</td>
<td>0.98</td>
<td>0.93</td>
<td>0.92</td>
<td>0.94</td>
<td>0.78</td>
<td>0.97</td>
<td>0.87</td>
<td>0.99</td>
<td>0.83</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>USA</td>
<td>0.80</td>
<td>0.96</td>
<td>0.99</td>
<td>0.48</td>
<td>0.53</td>
<td>0.58</td>
<td>0.88</td>
<td>0.96</td>
<td>0.86</td>
<td>0.88</td>
<td>0.95</td>
<td>0.95</td>
<td>0.94</td>
<td>0.96</td>
<td>0.85</td>
<td>0.96</td>
<td>1.00</td>
<td></td>
</tr>
</tbody>
</table>

Notes: The correlations are based on the estimated long-run covariance matrix. Bartlett kernel with window length four is used according to Newey and West (1994).
Table 3
Panel unit root results.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Test</th>
<th>dividend-price ratio</th>
<th>returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>Z</td>
<td>-0.819</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>0.538</td>
<td>0.764</td>
</tr>
<tr>
<td></td>
<td>Pm</td>
<td>40.569</td>
<td>42.484</td>
</tr>
<tr>
<td>Constant and trend</td>
<td>Z</td>
<td>-0.071</td>
<td>1.329</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>1.617</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>Pm</td>
<td>49.721</td>
<td>36.803</td>
</tr>
<tr>
<td>Break in the mean</td>
<td>Z</td>
<td>-0.546</td>
<td>-0.234</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>0.783</td>
<td>0.567</td>
</tr>
<tr>
<td></td>
<td>Pm</td>
<td>40.084</td>
<td>41.028</td>
</tr>
<tr>
<td>Break in the trend</td>
<td>Z</td>
<td>-1.839</td>
<td>-1.469</td>
</tr>
<tr>
<td></td>
<td>P</td>
<td>1.010</td>
<td>1.215</td>
</tr>
<tr>
<td></td>
<td>Pm</td>
<td>45.368</td>
<td>46.314</td>
</tr>
<tr>
<td></td>
<td>Z*</td>
<td>-1.820</td>
<td>-1.565</td>
</tr>
<tr>
<td></td>
<td>P*</td>
<td>1.160</td>
<td>1.149</td>
</tr>
<tr>
<td></td>
<td>Pm*</td>
<td>45.846</td>
<td>45.756</td>
</tr>
</tbody>
</table>

Notes: $P^*$ denotes the corresponding $P$ statistics that is computed using the p-values of the simplified MSB statistics. $P_{m}^*$ denotes the corresponding $P_{m}$ statistics obtained using the p-values of the simplified MSB statistics. The 5% critical values for Z, P and Pm tests are 1.645, -1.645 and 50.998, respectively. The number of common factors is estimated using the Panel Bayesian criterion information in Bai and Ng (2002) with $r_{max} = 6$ the maximum number of factors as in Bai and Carrion-i-Silvestre (2009). No simplified test for model 1 is provided by Bai and Carrion-i-Silvestre (2009).
analysis: Aburachis and Kish (1999) find evidence of stationarity in the dividend yields and stock returns for 8 OECD countries, namely Canada, France, Germany, Japan, Netherlands, Switzerland, UK and US; Ryan (2006) finds that the log dividend-price ratio variable is stationary for 16 OECD countries; McMillan (2007) reports that the dividend-price ratio variable shows a unit root process for 9 out of 13 OECD countries when applying six different unit root tests; Park (2010) shows that the unit root null hypothesis cannot be rejected for the dividend-price ratio in 6 out of 29 markets at the 10% level, while it can be largely rejected for stock returns in all markets.

As far as the break dates are concerned, we find different break points for log dividend-price ratio and total returns series (see Table 4). When considering Model 1 (see equation (10)) and log dividend-price ratio, we find a break for six countries: Belgium (1999:6), Germany (1999:1), Greece (1998:3), Japan (1999:2), UK (1999:5) and USA (1999:1); in the case of returns series, four countries experience a break: Germany (1998:8), Japan (1998:8), UK (1999:6) and USA (2000:4). When considering Model 2 (see equation (11)), we find a break for eight countries in the case of log dividend-price ratio series, namely Belgium (1999:9), Germany (1999:1), Greece (1998:3), Japan (1999:3), Netherlands (1999:1), Sweden (1999:4), UK (1999:5) and USA (1999:1), and a break for five countries in the case of returns series, namely Germany (1998:11), Japan (1998:7), Spain (1998:2), UK (2000:3) and USA (2000:4). The breaks seem to cluster around specific dates for most of the countries both for log dividend-price ratio and returns variables, implying “common” breaks. Brooks and Del Negro (2006) show that global shocks are a more important source of variation for returns than the country-specific shocks. A large rise in the importance of global factors seems to drive changes in returns and the economic growth seems to be the most important global factor that affects stock returns.

After studying stationarity properties of the data by using panel unit root tests, we investigate the presence of bubbles phenomena in the long-run relationship between log dividend-price ratio

---

12 We apply the Bai and Perron (1998) dynamic programming algorithm to estimate the number and the location of the breaks.
13 Beltratti and Morana (2010) also show the relative importance of the influence of global shocks on stock markets.
Table 4
Estimates of break dates for unit root.

<table>
<thead>
<tr>
<th>Countries</th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Belgium</td>
<td>1999:6</td>
<td>1999:9</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Greece</td>
<td>1998:3</td>
<td>1998:3</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Netherlands</td>
<td>−</td>
<td>1999:1</td>
<td>−</td>
<td>−</td>
</tr>
<tr>
<td>Spain</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>1998:2</td>
</tr>
<tr>
<td>Sweden</td>
<td>−</td>
<td>1999:4</td>
<td>−</td>
<td>−</td>
</tr>
</tbody>
</table>

Notes: In Model 1 (equation (10)), the structural breaks affect the mean. In Model 2 (equation (11)), the breaks affect the trend; − indicates no break.

and stock returns. If a stable (equilibrium) relationship is rejected, then the “no bubbles” hypothesis is also rejected. When applying both LM tests, $Z_t(N)$ and $Z_\phi(N)$, we allow for three breaks, which may be positioned at different dates for different units. We take into account structural breaks in both the level and the slope of the relationship in (5). We determine the number of lags used in the test regression for the LM tests by applying the sequential procedure proposed by Campbell and Perron (1991). We fix the maximum number of lags to 13. Furthermore, we set the maximum number of common factors to 5 and use a significance level of 5%. Table 5 reports the findings of the cointegration tests.

Table 5
Panel cointegration tests results.

<table>
<thead>
<tr>
<th></th>
<th>$Z_t(N)$</th>
<th>$Z_\phi(N)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>No break</td>
<td>−0.330</td>
<td>−0.640</td>
</tr>
<tr>
<td></td>
<td>(0.371)</td>
<td>(0.261)</td>
</tr>
<tr>
<td>Level break</td>
<td>−0.141</td>
<td>−0.317</td>
</tr>
<tr>
<td></td>
<td>(0.444)</td>
<td>(0.376)</td>
</tr>
<tr>
<td>Regime shift</td>
<td>−0.629</td>
<td>−1.080</td>
</tr>
<tr>
<td></td>
<td>(0.265)</td>
<td>(0.376)</td>
</tr>
</tbody>
</table>

Notes: The number of lags in the test regressions for both LM tests is selected using the procedure of Campbell and Perron (1991). The maximum number of common factors is set to 5. p-values in parentheses are for a one-sided test based on the normal distribution.

The cointegration results suggest that the null of no cointegration cannot be rejected at any significance conventional level for all models and the presence of the bubbles phenomena is con-
firmed. With respect to the break points (see Table 6), we find the same dates for both level breaks and regime shifts for different groups of countries: Greece (2005:3) and Sweden (2005:3); Hong Kong (1998:9) and New Zealand (1998:9); Austria (2008:8), Ireland (2008:8), Japan (2008:8), Netherlands (2008:8), UK (2008:8) and USA (2008:8).

Table 6
Estimates of break dates for cointegrating relationship.

<table>
<thead>
<tr>
<th></th>
<th>level break</th>
<th>regime shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>Austria</td>
<td>2008:8</td>
<td>2008:8</td>
</tr>
<tr>
<td>Belgium</td>
<td>2008:8</td>
<td>2003:4</td>
</tr>
<tr>
<td>Canada</td>
<td>2008:8</td>
<td>2001:1</td>
</tr>
<tr>
<td>Finland</td>
<td>1994:1</td>
<td>2002:11</td>
</tr>
<tr>
<td>France</td>
<td>2006:4</td>
<td>2003:2</td>
</tr>
<tr>
<td>Germany</td>
<td>1999:5</td>
<td>2008:5</td>
</tr>
<tr>
<td>Greece</td>
<td>2005:3</td>
<td>2005:3</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>1998:9</td>
<td>1998:9</td>
</tr>
<tr>
<td>Ireland</td>
<td>2008:8</td>
<td>2008:8</td>
</tr>
<tr>
<td>Japan</td>
<td>2008:8</td>
<td>2008:8</td>
</tr>
<tr>
<td>Netherlands</td>
<td>2008:8</td>
<td>2008:8</td>
</tr>
<tr>
<td>Spain</td>
<td>2002:9</td>
<td>2008:5</td>
</tr>
<tr>
<td>Sweden</td>
<td>2005:3</td>
<td>2005:3</td>
</tr>
<tr>
<td>Switzerland</td>
<td>1994:5</td>
<td>2008:1</td>
</tr>
<tr>
<td>UK</td>
<td>2008:8</td>
<td>2008:8</td>
</tr>
<tr>
<td>USA</td>
<td>2008:8</td>
<td>2008:8</td>
</tr>
</tbody>
</table>

Notes: The break dates are selected using a grid search (see Westerlund and Edgerton (2008)).

5. Conclusions

In this paper we present empirical evidence of the bubbles phenomena in the international stock markets over the period 1992:1-2010:6 for a panel of 18 OECD countries. Using the log-linear present value model of Campbell (2000), we investigate the presence of rational bubbles in the log dividend-price ratio and total returns. This paper makes an important contribution to the existing literature by providing an analysis of international data that exploits increased power deriving from the panel unit root and cointegration methodology, together with the flexibility of allowing explicitly for multiple endogenous structural breaks in the individual series. Differently from the
time series methodology, the panel data approach allows for a global analysis of the financial crashes that are related to the speculative bubbles. Our empirical results suggest the existence of the bubbles. These findings are relevant for bubbles literature for several reasons. First, it is noteworthy that our results provide stronger evidence of bubble behavior in the OECD countries than the most of the evidence from the univariate analysis. Second, our empirical analysis regarding structural breaks in the dividends and returns data shows that breaks occur around the same dates for most of the countries, in particular around the “tech-stock” bubbles period. These findings confirm that global shocks are a more important source of variation for returns than the country-specific shocks (see Brooks and Del Negro (2006)). Third, our results provide various economic and financial implications to private agents, investors, policy makers and financial authorities at international level as the panel data approach uses information across different countries: i) changes in asset prices from cyclical movements in firms’ net worth tend to have strong effects on agency costs and credit conditions, thereby affecting the firms’ investment policy at international level; ii) even if the bubbles allow investors to earn abnormal profits in the international stock markets, the existence of bubbles makes the investors conscious of the size of a bubble and then assists them in identifying early signals prior to crashes (Markwat et al. (2009) point out that the domino effect can be also used to improve early-warning systems), which may enable the investors to rationally perform by selling the assets and adjusting the share price toward its fair value, as well as making the markets to be efficient; iii) efforts to adjust accounting standards and practices to promote long term participation in the markets and improve and diversify the risk-management tools used should be produced by policy makers; iv) financial institutions can better identify and prevent bubbles if they improve the public access to accurate information.
Acknowledgement

We are grateful to the Editor, Ike Mathur, and an anonymous referee for their helpful and extensive comments which greatly improved the paper. We also thank Marco Ligios for many useful discussions. The usual disclaimer applies.

Appendix

We first describe the univariate MSB statistics in absence and in presence of structural breaks and then we present the simplified MSB statistics.

MSB statistics for univariate series with no break

Consider the model with a constant:

\[ X_t = \mu + e_t \]  \hspace{1cm} (23)

\[ e_t = \rho e_{t-1} + H(L)e_t \] \hspace{1cm} (24)

where \( t = 1, \ldots, T \), \( H(L) = \sum_{j=0}^{\infty} H_j L^j \), \( e_t \sim i.i.d(0, \Sigma_e) \), \( E |e_t|^{8} \leq A \) with \( A < \infty \) a generic positive number not depending on \( T \). The MSB procedure tests the null hypothesis that \( X_t \) is a non-stationary I(1) process against the alternative hypothesis that \( X_t \) is I(0):

\[ H_0 : \rho = 1 \]

\[ H_1 : |\rho| < 1 \]

where \( \rho \) is the autoregressive parameter in (24). The MSB statistics is defined as:

\[ MSB = T^{-2} \frac{\sum_{t=1}^{T} \hat{e}_{t-1}^2}{\hat{\sigma}^2} \] \hspace{1cm} (25)

where \( \hat{e}_t \) indicates the estimated residuals from equation (23), whereas \( \hat{\sigma}^2 \) is a consistent estimator of the long-run variance of \( e_t - e_{t-1} \). To estimate \( e_t \), Bai and Carrion-i-Silvestre (2009) use a
difference-recumulation procedure. They take the first difference of equation (23), $\Delta X_t = \Delta e_t$ and let $\Delta \hat{e}_t = \Delta X_t$. Then, the cumulative sum of $\Delta \hat{e}_t$ gives $\hat{e}_t = \sum_{s=1}^{t} \Delta X_s = X_t - X_1$.

In the case of linear trend, equation (23) can be written as:

$$X_t = \mu + \beta t + e_t \quad (26)$$

To estimate $e_t$, the first difference of equation (26) is taken, $\Delta X_t = \beta + \Delta e_t$. Then the parameter $\beta$ is estimated using $\hat{\beta} = \Delta X = (X_T - X_1) / T$ and the estimated residuals of the first-difference model are taken: $\Delta \hat{e}_t = \Delta X_t - (X_T - X_1) / T$. Cumulating $\Delta \hat{e}_t$ gives $\hat{e}_t = \sum_{s=1}^{t} \Delta \hat{e}_s = (X_t - X_1) - (X_T - X_1) t / T$.

Using the estimated residuals $\hat{e}_t$, we can compute the MSB statistics as in equation (25) to test the unit root hypothesis.

**MSB statistics for univariate series with multiple breaks**

We extend the framework presented in the previous subsection to the case of multiple structural breaks.

Consider first the case when the multiple structural breaks occur in the intercept:

$$X_t = \mu + \sum_{j=1}^{l} \theta_j D_{U_{jt}} + e_t \quad (27)$$

where $e_t$ follows the specification in (24) and $D_{U_{jt}}$ is a dummy variables with $D_{U_{jt}} = 1$ for $t > T_j$ and 0 elsewhere ($j = 1, \ldots, l$). Taking the first difference of the model (27), we have:

$$\Delta X_t = \sum_{j=1}^{l} \theta_j I(T_j) + e_t \quad (28)$$

where $I(T_j)$ are impulses such that $I(T_j) = 1$ for $t = T_j + 1$ and 0 elsewhere, $j = 1, \ldots, l$. To construct the MSB statistics, we simply use $\hat{e}_t = \sum_{s=1}^{t} \Delta X_s = X_s - X_1$, which is the same as the statistics defined in equation (25).

Consider now the model in which the structural breaks are present in both the intercept and the
time trend:

\[ X_t = \mu + \beta t + \sum_{j=1}^{l} \theta_j DU_{jt} + \sum_{k=1}^{m} \gamma_k DT_{kt} + e_t \]  

(29)

where \( e_t \) is defined as in (24), \( DU_{jt} = 1 \) for \( t > T_{aj} \) and 0 elsewhere, and \( DT_{kt} = (t - T_{bk}) \) for \( t > T_{bk} \) and 0 elsewhere. \( T_{aj} \) and \( T_{bk} \) denote the \( j \)-th break in the intercept and the \( k \)-th break in the trend, respectively (\( j = 1, \ldots, l \), and \( k = 1, \ldots, m \)). Taking the first difference of (29), we have:

\[ \Delta X_t = \beta + \sum_{j=1}^{l} \theta_j I(T_{aj}) + \sum_{k=1}^{m} \gamma_k DU_{kt} + \Delta e_t \]  

(30)

\[ = \beta + \sum_{k=1}^{m} \gamma_k DU_{kt} + \Delta e_t^* \]  

(31)

where \( \Delta e_t^* = \sum_{j=1}^{l} \theta_j I(T_{aj}) + \Delta e_t \), with \( DU_{kt} = 1 \) for \( t > T_{bk} \) and 0 otherwise, and \( I(T_{aj}) = 1 \) for \( t = T_{aj} + 1 \) and 0 otherwise. Let \( \hat{e}_t \) be the estimated residuals obtained from the application of the algorithm proposed by Bai and Perron (1998) to equation (31). Define \( \hat{e}_t = \sum_{s=2}^{t} \Delta e_t^* \). The MSB statistics is based on \( \hat{e}_t \) as in equation (25).

**Simplified MSB statistics**

The simplified statistics is defined as:

\[ MSB^*(\lambda) = \frac{\sum_{k=1}^{m+1} ((\hat{T}_{bk} - \hat{T}_{bk-1})^{-2} - \sum_{t=\hat{T}_{bk-1}+1}^{\hat{T}_{bk}} \hat{e}_{t-1}^2)}{\hat{\sigma}^2} \]  

(32)

where \( T_{b0} = 0 \) and \( T_{bm+1} = T \). We can use the simplified statistics to test the null hypothesis that \( X_t \) is a nonstationary \( I(1) \) stochastic process, (\( \rho = 1 \) in equation (24)), against the alternative hypothesis that \( X_t \) is a stationary \( I(0) \) stochastic process (\( |\rho| < 1 \) in equation (24)).
References


Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70, 191-221.


