Journal of Financial Econometrics, 2012, Vol. 10, No. 1, 198–231

Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods

ULRICH HOMM
University of Bonn

JÖRG BREITUNG
University of Bonn

ABSTRACT
We propose several tests for rational bubbles and investigate their power properties. The focus lies on the case where bubble detection is reduced to testing for a unknown change from a random walk to an explosive process. In simulations, a Chow-type break test exhibits the highest power and performs well relative to the power envelope. The Chow-type procedure also provides a reliable estimator for the break date. Furthermore, we suggest monitoring procedures for detecting speculative bubbles in real time. As applications, we analyze the Nasdaq composite index and various other financial time series. (*JEL:* G10, C22)

KEYWORDS: speculative bubbles, structural break, unit root test

Phenomena of speculative excesses have long been present in economic history. Galbraith (1993) starts his account with the famous *Tulipomania*, which took place in the Netherlands in the 17th century. More recently, the so-called *dot.com* or *IT* bubble came to fame during the end of the 1990s. Enormous increases in stock prices followed by crashes have led many researchers to test for the presence of speculative bubbles. Among these are Shiller (1981) and LeRoy and Porter (1981), who proposed variance bounds tests, West (1987), who designed a two-step test for bubbles, and Froot and Obstfeld (1991), who considered intrinsic bubbles. Moreover, Cuñado, Gil-Alana, and de Gracia (2005) and Frömmel and Kruse (2011) employ methods based on fractionally integrated models, while Phillips, Wu,

The authors are grateful to Norbert Christopeit, Uwe Hassler, an associate editor, and two anonymous referees for valuable comments and suggestions. Address correspondence to Ulrich Homm, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany, or e-mail: uhomm@uni-bonn.de.

doi: 10.1093/jjfinec/nbr009

and Yu (2011) use sequential unit root tests. This list is by no means complete. Gürkaynak (2008) provides an overview of different empirical tests on rational bubbles. Here, we adopt the theoretical framework of Phillips, Wu, and Yu (2011) and propose several other test procedures aiming to improve on the testing power. The forward recursive unit root test of Phillips, Wu, and Yu (2011) is an attempt to overcome the weaknesses of the approach of Diba and Grossman (1988), who argue against the existence of bubbles in the S&P 500. Evans (1991) demonstrates that Diba and Grossman's (1988) tests do not have sufficient power to effectively detect bubbles that collapse periodically. Phillips, Wu, and Yu (2011) also propose to use sequences of Dickey–Fuller (DF) statistics to estimate the date of the emergence of a bubble, that is, to estimate the date of a regime switch from a random walk to an explosive process.

A main objective of this paper was to provide alternative tests that are more powerful in detecting a change from a random walk to an explosive process. To this end, we modify several tests that have been proposed in a different context and transfer them to the bubble testing framework. Our Monte Carlo simulations suggest that two of the alternative test procedures outperform the recursive unit root test of Phillips, Wu, and Yu (2011). Moreover, the empirical power of these procedures is quite close to the power envelope. A second objective of this paper was to suggest a reliable estimator for the break date, that is, the starting date of the bubble. We also look at the problem from a practitioners' perspective and suggest a real-time monitoring approach to detect emerging bubbles.

The tests that we adapt to the bubble framework originate from the literature on tests for a change in persistence. Kim (2000), Kim, Belaire-Franch, and Amador (2002), and Busetti and Taylor (2004) proposed procedures to test the null hypothesis that the time series is I(0) throughout the entire sample against the alternative of a change from I(0) to I(1) or vice versa.[1] We adapt these test procedures to the context of bubble detection and study their power properties by means of Monte Carlo simulations. Additionally, two other tests will be included in the study. One is based on Bhargava (1986), who tested whether a time series is a random walk against explosive alternatives. Bhargava (1986) did not construct his test as a test for a structural break. However, we will adjust this test to accommodate regime switches and apply it sequentially to different subsamples. The other test is a version of the classical Chow test.

When there is only a single regime switch in the sample, the proposed sequential Chow test and our modified version of Busetti and Taylor's (2004) procedure exhibit the highest power. This does no longer hold, however, if there are multiple regime changes due to bubble crashes. In that case, it is more appropriate to make a slight change of perspective and apply statistical monitoring procedures. In principle, all tests can be redesigned as monitoring procedures, but here, we will focus on the sequential DF t-statistic and a simple CUSUM procedure.

[1] In Section 2, we will briefly discuss other tests for a change in persistence and the reason why we focus on the tests of Kim (2000), Kim, Belaire-Franch, and Amador (2002), and Busetti and Taylor (2004).

This paper is organized as follows. In Section 1, we present the basic theory of rational bubbles. We introduce a simple model for *randomly starting bubbles* and review Evans' (1991) *periodically collapsing bubbles*. In Section 2, the above-mentioned tests and estimation procedures are introduced. Monitoring procedures are considered in Section 3. Furthermore, in Section 4, the performance of the procedures is analyzed via Monte Carlo methods. Finally, in Section 5, the test and estimation procedures are applied to Nasdaq index data and various other financial time series. Section 6 concludes.

1 RATIONAL BUBBLES

Speculative bubbles in stock markets are systematic departures from the fundamental price of an asset. Following Blanchard and Watson (1982) or Campbell, Lo, and MacKinlay (1997), the fundamental price of the asset is derived from the following standard no arbitrage condition:

$$P_t = \frac{E_t\left[P_{t+1} + D_{t+1}\right]}{1+R}, \tag{1}$$

where P_t denotes the stock price at period t, D_{t+1} is the dividend for period t, R is the constant risk-free rate, and $E_t\left[\cdot\right]$ denotes the expectation conditional on the information at time t. Solving Equation (1) by forward iteration yields the fundamental price

$$P_t^f = \sum_{i=1}^{\infty} \frac{1}{(1+R)^i} E_t\left[D_{t+i}\right]. \tag{2}$$

Equation (2) states that the fundamental price is equal to the present value of all expected dividend payments. Imposing the transversality condition

$$\lim_{k\to\infty} E_t\left[\frac{1}{(1+R)^k} P_{t+k}\right] = 0 \tag{3}$$

ensures that $P_t = P_t^f$ is the unique solution of Equation (1) and thereby rules out the existence of a bubble. However, if Equation (3) does not hold, P_t^f is not the only price process that solves Equation (1). Consider a process $\{B_t\}_{t=1}^{\infty}$ with the property

$$E_t\left[B_{t+1}\right] = (1+R)B_t. \tag{4}$$

It can easily be verified that adding B_t to P_t^f will yield another solution to Equation (1). In fact, there are infinitely many solutions. They take the form

$$P_t = P_t^f + B_t, \tag{5}$$

where $\{B_t\}_{t=1}^{\infty}$ is a process that satisfies Equation (4). The last equation decomposes the price into two components: the fundamental component, P_t^f, and a part that is commonly referred to as the bubble component, B_t. If a bubble is present in the stock price, Equation (4) requires that any rational investor, who is willing to buy that stock, must expect the bubble to grow at rate R. If this is the case and if B_t is strictly positive, this sets the stage for speculative investor behavior: a rational investor is willing to buy an "overpriced" stock since she believes that through price increases, she will be sufficiently compensated for the extra payment B_t. If investors expect prices to increase at rate R and buy shares, the stock price will indeed rise and complete the loop of a self-fulfilling prophecy.

The crucial condition for rational bubbles is given by Equation (4). However, this restriction leaves room for a variety of processes. We next present several models for rational bubbles, some of which will also be considered in our Monte Carlo analysis. The simplest example of a process that satisfies Equation (4) is the deterministic bubble, given by $B_t = (1+R)^t B_0$, where B_0 is an initial value. A somewhat more realistic example, in which the bubble does not necessarily grow forever, is taken from Blanchard and Watson (1982). The bubble process is given by

$$B_{t+1} = \begin{cases} \pi^{-1}(1+R)B_t + \mu_{t+1}, & \text{with probability } \pi \\ \mu_{t+1}, & \text{with probability } 1-\pi, \end{cases} \tag{6}$$

where $\{\mu_t\}_{t=1}^{\infty}$ is a sequence of iid random variables with zero mean. In each period, the bubble described in Equation (6) will continue, with probability π, or collapse, with probability $1-\pi$. As long as the bubble does not collapse, the realized return exceeds the risk-free rate R as a compensation for the risk that the bubble bursts.

Not every process that satisfies Equation (4) is consistent with rationality. For instance, given that stock prices cannot be negative, negative bubbles can be excluded: applying the law of iterated expectations to Equation (4) yields $E_t[B_{t+\tau}] = (1+R)^\tau B_t$. If at some time t, B_t was negative, then, as τ goes to infinity, the expected bubble tends to minus infinity, implying a negative stock price at some future time period. Furthermore, Diba and Grossman (1988) argue that a bubble process cannot start from zero. Assume that $B_t = 0$ at some time t. Then, $E_t[B_{t+1}] = 0$, by Equation (4). Assuming nonnegative stock prices, we have $B_{t+1} \geqslant 0$. Together this implies that $B_{t+1} = 0$ almost surely.

Although rational bubbles cannot start from zero, they can take a constant positive value for some time and start to grow exponentially with some probability π. For instance, consider the following randomly starting bubble:

$$B_t = \begin{cases} B_{t-1} + \frac{RB_{t-1}}{\pi}\theta_t, & \text{if } B_{t-1} = B_0 \\ (1+R)B_{t-1}, & \text{if } B_{t-1} > B_0 \end{cases} \qquad \text{for } t = 1, \dots, T, \tag{7}$$

where $B_0 > 0$ is the initial value of the bubble. R is the risk-free rate, and $\{\theta_t\}_{t=1}^{T}$ is an exogenous iid Bernoulli process with $\text{Prob}(\theta_t = 1) = \pi = 1 - \text{Prob}(\theta_t = 0)$, and $\pi \in (0, 1]$. A process generated according to Equation (7) starts at some strictly positive value B_0 and remains at that level until the Bernoulli process switches to unity. At that point, the process B_t makes a jump of size RB_0/π and from then on grows at rate R. One can easily verify that such a process satisfies the no arbitrage condition for rational bubbles, given in Equation (4). In our Monte Carlo analysis, we will simulate such a bubble process together with a fundamental price P_t^f generated according to Equations (10) and (11). An example of the resulting price processes $P_t = P_t^f + B_t$ is given in Figure 1 (solid line), which also depicts the fundamental price (dotted line). In this example, the parameters for the bubble process are $\pi = 0.05$, $R = 0.05$, and $B_0 = 1$. Following Evans (1991), the dividend process in Equation (10) is simulated with drift $\mu = 0.0373$, initial value $D_0 = 1.3$, and identical normally distributed disturbances with mean zero and variance $\sigma^2 = 0.1574$.

In his critique of Diba and Grossman's (1988) testing approach, Evans (1991) proposed the following model for periodically collapsing bubbles:

$$B_{t+1} = \begin{cases} (1+R)B_t u_{t+1} & \text{if } B_t \leqslant \alpha \\ \left[\delta + \pi^{-1}(1+R)\theta_{t+1}(B_t - (1+R)^{-1}\delta)\right] u_{t+1} & \text{otherwise.} \end{cases} \quad (8)$$

Here, δ and α are parameters satisfying $0 < \delta < (1+R)\alpha$, and $\{u_t\}_{t=1}^{\infty}$ is an iid process with $u_t \geqslant 0$ and $E_t[u_{t+1}] = 1$, for all t. $\{\theta_t\}_{t=1}^{\infty}$ is an iid Bernoulli process, where the probability that $\theta_t = 1$ is π and the probability that $\theta_t = 0$ is $1 - \pi$, with $0 < \pi \leqslant 1$. It is easy to verify that the bubble process defined in Equation (8) satisfies Equation (4). Letting the initial value $B_0 = \delta$, the bubble increases until it exceeds some value α. Thereafter, it is subject to the possibility of collapse with

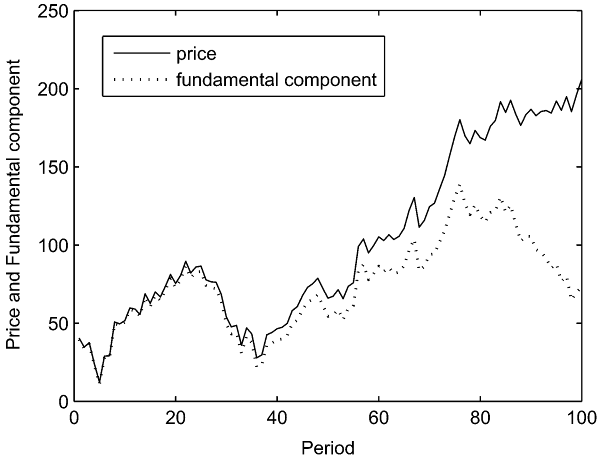

Figure 1 Simulated price series with randomly starting bubble component.

probability $(1-\pi)$, in which case, it will return to δu_{t+1} (i.e., to δ in expectation). For our simulations, we will follow Evans (1991) and specify $u_t = \exp(\xi_t - \frac{1}{2}\tau^2)$, where $\xi_t \sim$ iid $N(0,\tau^2)$. The parameter values are set to $\alpha = 1$, $\delta = 0.5$, $\tau = 0.05$, and $R = 0.05$. Note that such a periodically collapsing bubble never crashes to zero. Thus, it does not violate Diba and Grossman's (1988) finding that a bubble cannot restart from zero. Evans (1991) demonstrated that Diba and Grossman's (1988) tests lack sufficient power to detect periodically collapsing bubbles, even if the probability of collapse is small. A realization of a periodically collapsing bubble is shown in the right panel of Figure 2, where $\pi = 0.85$. The left panel of Figure 2 displays the fundamental price (dotted line), which is generated as above, and the observed price (solid line).

The observed price is constructed as

$$P_t = P_t^f + 20B_t. \tag{9}$$

As in Evans (1991), the bubble process is multiplied by a factor of 20. This is to ensure that the variance of the first difference of the bubble component $\Delta(20B_t)$ is large relative to the variance of the first difference of the fundamental price ΔP_t^f.

An obvious problem is that the fundamental component in Equation (5) cannot be directly observed. Therefore, assumptions have to be imposed to characterize the time-series properties of the fundamental price P_t^f. A convenient—and nevertheless empirically plausible—assumption is that dividends follow a random walk with drift

$$D_t = \mu + D_{t-1} + u_t, \tag{10}$$

where u_t is a white noise process. Under this assumption, the fundamental price results as

$$P_t^f = \frac{1+R}{R^2}\mu + \frac{1}{R}D_t, \tag{11}$$

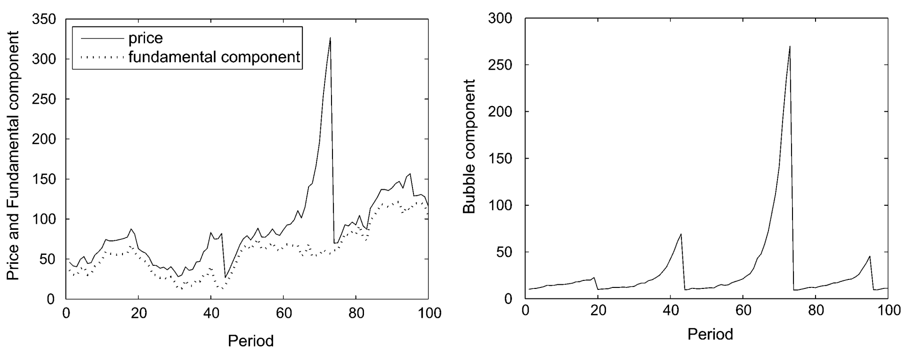

Figure 2 Simulated price with fundamental component (left) and bubble component (right).

(e.g., Evans 1991). Consequently, if D_t follows a random walk with drift, so does P_t^f. This allows us to distinguish the fundamental price from the bubble process that is characterized by an explosive autoregressive process (see also Diba and Grossman 1988).

2 TEST PROCEDURES

The test procedures are based on the time-varying AR(1) model

$$y_t = \rho_t y_{t-1} + \epsilon_t \, , \tag{12}$$

where ε_t is a white noise process with $E(\varepsilon_t) = 0$, $E(\varepsilon_t^2) = \sigma^2$, and $y_0 = c < \infty$. To simplify the exposition, we ignore a possible constant in the autoregression. If the test is applied to a series of daily stock prices, the constant is usually very small and insignificant. To account for a possible constant in Equation (12), the series may be detrended by running a least squares regression on a constant and a linear time trend. All test statistics presented in this section can be computed by using the residuals of this regression instead of the original time series.[2]

Under the null hypothesis, y_t follows a random walk for all time periods, that is,

$$H_0: \quad \rho_t = 1 \quad \text{for } t = 1, 2, \ldots, T. \tag{13}$$

Under the alternative hypothesis, the process starts as a random walk but changes to an explosive process at an unknown time $[\tau^* T]$ (where $\tau^* \in (0, 1)$ and $[\tau^* T]$ denotes the greatest integer smaller than or equal to $\tau^* T$):

$$H_1: \quad \rho_t = \begin{cases} 1 & \text{for } t = 1, \ldots, [\tau^* T] \\ \rho^* > 1 & \text{for } t = [\tau^* T] + 1, \ldots, T. \end{cases} \tag{14}$$

Various statistics have been suggested to test for a structural break in the autoregressive parameter. Most of the work focus on a change from a nonstationary regime (i.e., $\rho_t = 1$) to a stationary regime ($\rho_t < 1$) or vice versa. Since these test statistics can be easily adapted to the situation of a change from an I(1) to an explosive process, we first consider various test statistics suggested in the literature.

2.1 The Bhargava Statistic

To test the null hypothesis of a random walk ($\rho_t = 1$) against explosive alternatives $\rho_t = \rho^* > 1$ for all $t = 1, \ldots, T$, Bhargava (1986) proposed the locally most

[2]In that case, Brownian motions are replaced by detrended Brownian motions in the limiting distribution of the test statistics (see below).

powerful invariant test statistic

$$B_0^* = \frac{\sum_{t=1}^{T}(y_t - y_{t-1})^2}{\sum_{t=1}^{T}(y_t - y_0)^2}. \tag{15}$$

Since Bhargava's (1986) alternative does not incorporate a structural break, we employ a modified version of the inverted test statistic:

$$B_\tau = \frac{1}{T-[\tau T]}\left(\frac{\sum_{t=[\tau T]+1}^{T}(y_t - y_{t-1})^2}{\sum_{t=[\tau T]+1}^{T}(y_t - y_{[\tau T]})^2}\right)^{-1} = \frac{1}{s_\tau^2\,(T-[\tau T])^2}\sum_{t=[\tau T]+1}^{T}(y_t - y_{[\tau T]})^2, \tag{16}$$

where $s_\tau^2 = (T-[\tau T])^{-1}\sum_{t=[\tau T]+1}^{T}(y_t - y_{t-1})^2$. To test for a change from I(1) to an explosive process in the interval $\tau \in [0, 1-\tau_0]$, where $\tau_0 \in (0, 0.5)$, we consider the statistic

$$\sup B(\tau_0) = \sup_{\tau\in[0,1-\tau_0]} B_\tau. \tag{17}$$

The test rejects the null hypothesis for large values of $\sup B(\tau_0)$. Note that the original Bhargava (1986) test rejects if B_0^* is small. Since all tests presented below reject for large values, our test statistic is inversely related to the original Bhargava statistic.

The statistic (17) may be motivated as follows. Assume that we want to forecast the value y_{T^*+h} at period $T^* = [\tau T]$. Since it is assumed that the series is a random walk up to period T^*, the forecast results as $\widehat{y}_{T^*+h|T^*} = y_{T^*}$. The B_τ-statistic is based on the sum of squared forecast errors for $y_{T^*+1}, \ldots, y_T$. If the second part of the sample is generated by an explosive process, then the random-walk forecast becomes very poor as h gets large. Therefore, this test statistic is supposed to have good power against explosive alternatives. The supremum of the statistics B_τ is used to cope with the fact that the breakpoint is unknown.

The asymptotic distribution of this test statistic under null hypothesis was not derived in the literature but easily follows from the continuous mapping theorem as

$$\sup B(\tau_0) \Rightarrow \sup_{\tau\in[0,1-\tau_0]}\left\{(1-\tau)^{-2}\int_\tau^1 (W(r)-W(\tau))^2\,\mathrm{d}r\right\},$$

where $\Rightarrow$ denotes weak convergence and W denotes standard Brownian motion on the interval $[0, 1]$.

2.2 The Busetti–Taylor Statistic

Busetti and Taylor (2004) proposed a statistic for testing the hypothesis that a time series is stationary against the alternative that it switches from a stationary to an I(1) process at an unknown breakpoint. Here, we propose a modified version of the statistic to test the null hypothesis (13) against the alternative (14):

$$\text{supBT}(\tau_0) = \sup_{\tau \in [0, 1-\tau_0]} \text{BT}_\tau, \qquad \text{where } \text{BT}_\tau = \frac{1}{s_0^2 (T - [\tau T])^2} \sum_{t=[\tau T]+1}^{T} (y_T - y_{t-1})^2. \tag{18}$$

The supBT test rejects for large values of $\text{supBT}(\tau_0)$. Note that BT_τ employs the variance estimator s_0^2 based on the entire sample, while the inverted Bhargava statistic in Equation (16) employs s_τ^2, which uses only the observations starting at $[\tau T]$. Another way to illustrate the difference between the two test statistics is to note that the BT statistic is based on the sum of squared forecast errors of forecasting the final value y_T from the periods $y_{T^*} + 1, \ldots, y_{T-1}$ by using the null hypothesis that y_t is generated by a random walk. Therefore, the BT statistic fixes the target to be forecasted, whereas the Bhargava statistic uses multiple forecast horizons of a fixed forecast interval. The following result for the limiting distribution of supBT can easily be derived:

$$\sup_{\tau \in [0, 1-\tau_0]} \text{BT}_\tau \Rightarrow \underset{\tau \in [0, 1-\tau_0]}{\text{BT}} \left\{ (1-\tau)^{-2} \int_\tau^1 W(1-r)^2 \mathrm{d}r \right\}.$$

Remark: In their work, Busetti and Taylor (2004) considered the process $y_t = \beta_0 + \mu_t + \epsilon_t$, where β_0 is a constant, $\epsilon_t \sim$ iid $N(0, \sigma^2)$, and μ_t is a process that is I(0) under the null hypothesis and switches from I(0) to I(1) under the alternative. They proposed the statistic $\varphi(\tau) = \hat{\sigma}^{-2}(T - [\tau T])^{-2} \sum_{t=[\tau T]+1}^{T} \left(\sum_{j=t}^{T} \hat{\epsilon}_j \right)^2$ where $\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^{T} \hat{\epsilon}_t^2$ and $\hat{\epsilon}_t$ are the residuals from OLS-regression of y_t on an intercept. To obtain stationary residuals under the null hypothesis (13), we use one-step-ahead forecast errors $y_t - y_{t-1}$ instead of OLS-residuals $\hat{\epsilon}_t$, which leads to Equation (18).

2.3 The Kim Statistic

Another statistic for testing the I(0) null hypothesis against a change from I(0) to I(1) was proposed by Kim (2000) and Kim, Belaire-Franch, and Amador (2002). To transfer the statistic to the bubble testing framework, we apply modifications similar to those in the remark above, which yields the following statistic:

$$\text{sup}K(\tau_0) = \sup_{\tau \in [\tau_0, 1-\tau_0]} K_\tau \qquad \text{with} \quad K_\tau = \frac{(T - [\tau T])^{-2} \sum_{t=[\tau T]+1}^{T} (y_t - y_{[\tau T]})^2}{[\tau T]^{-2} \sum_{t=1}^{[\tau T]} (y_t - y_0)^2}. \tag{19}$$

The test rejects for large values of $\text{sup}K(\tau_0)$. The statistic K_τ is computed over the symmetric interval $[\tau_0, 1 - \tau_0]$. It can be interpreted as the scaled ratio of the sum of squared forecast errors. The prediction is made under the assumption that the time series follows a random walk. y_0 is used to forecast $y_1, \ldots, y_{[\tau T]}$ (denominator) and

$y_{[\tau T]}$ is the forecast of $y_{[\tau T]+1}, \ldots, y_T$. The limiting distribution is obtained as

$$\sup_{\tau \in [\tau_0, 1-\tau_0]} K_\tau \quad \Rightarrow \quad \sup_{\tau \in [\tau_0, 1-\tau_0]} \left\{ \left(\frac{\tau}{1-\tau} \right)^2 \frac{\int_\tau^1 (W(r) - W(\tau))^2 \mathrm{d}r}{\int_0^\tau W(r)^2 \, \mathrm{d}r} \right\}.$$

2.4 The Phillips/Wu/Yu Statistic

To test for speculative bubbles, Phillips, Wu, and Yu (2011) suggest to use a sequence of DF tests. Let $\hat{\rho}_\tau$ denote the OLS estimator of ρ and $\hat{\sigma}_{\rho,\tau}$ the usual estimator for the standard deviation of $\hat{\rho}_\tau$ using the subsample $\{y_1, \ldots, y_{[\tau T]}\}$.[3] The forward recursive DF test is given by

$$\text{supDF}(\tau_0) = \sup_{\tau_0 \leqslant \tau \leqslant 1} \text{DF}_\tau \qquad \text{with} \quad \text{DF}_\tau = \frac{\hat{\rho}_\tau - 1}{\hat{\sigma}_{\rho,\tau}}. \tag{20}$$

Usually, the standard DF test is employed to test H_0 against the alternative $\rho_t = \rho < 1 \quad (t = 1, \ldots, T)$, and the test rejects if DF_1 is small. For the alternative considered here [see Equation (14)], we use upper tail critical values and reject when $\text{supDF}(\tau_0)$ is large. Note that the DF statistic is computed for the asymmetric interval $[\tau_0, 1]$. Following Phillips, Wu, and Yu (2011), we will set $\tau_0 = 0.1$ in the simulation experiments. The limiting distribution derived by Phillips, Wu, and Yu (2011) is

$$\sup_{\tau_0 \leqslant \tau \leqslant 1} \text{DF}_\tau \Rightarrow \quad \sup_{\tau_0 \leqslant \tau \leqslant 1} \frac{\int_0^\tau W(r) \mathrm{d}W(r)}{\sqrt{\int_0^\tau W(r)^2 \mathrm{d}r}}.$$

The test procedure does not take into account that both under the null hypothesis (13) and under the alternative (14), y_t is a random walk for $t = 1, \ldots, [\tau^* T]$. In this sense, the supDF test does not exploit all information.

2.5 A Chow-Type Unit Root Statistic for a Structural Break

The information that y_t is a random walk for $t = 1, \ldots, [\tau^* T]$ under both H_0 and H_1 can be incorporated in the test procedure by using a Chow test for a structural break in the autoregressive parameter. Under the assumption that $\rho_t = 1$ for $t = 1, \ldots, [\tau T]$ and $\rho_t - 1 = \delta > 0$ for $t = [\tau T] + 1, \ldots, T$, the model can be written as

$$\Delta y_t = \delta \left(y_{t-1} \mathbb{1}_{\{t > [\tau T]\}} \right) + \varepsilon_t, \tag{21}$$

where $\mathbb{1}_{\{\cdot\}}$ is an indicator function that equals 1 when the statement in braces is true and equals 0 otherwise. Correspondingly, the null hypothesis of interest is

[3] In their paper, Phillips, Wu, and Yu (2011) apply augmented DF tests and use a constant in their regression.

$H_0 : \delta = 0$, which is tested against the alternative $H_1 : \delta > 0$. It is easy to see that the regression t-statistic for this null hypothesis is

$$\mathrm{DFC}_\tau = \frac{\sum_{t=[\tau T]+1}^{T} \Delta y_t y_{t-1}}{\widetilde{\sigma}_\tau \sqrt{\sum_{t=[\tau T]+1}^{T} y_{t-1}^2}}, \tag{22}$$

where

$$\widetilde{\sigma}_\tau^2 = \frac{1}{T-2} \sum_{t=2}^{T} \left(\Delta y_t - \widehat{\delta}_\tau y_{t-1} \mathbb{1}_{\{t>[\tau T]\}} \right)^2,$$

and $\widehat{\delta}_\tau$ denotes the OLS estimator of δ in Equation (21). The Chow-type DF statistic to test for a change from I(1) to explosive in the interval $\tau \in [0, 1 - \tau_0]$ can be written as

$$\mathrm{supDFC}(\tau_0) = \sup_{\tau \in [0,1-\tau_0]} \mathrm{DFC}_\tau. \tag{23}$$

The test rejects for large values of $\mathrm{supDFC}(\tau_0)$. The test, in fact, corresponds to a one-sided version of the "supWald" test of Andrews (1993), where the supremum is taken over a sequence of Wald statistics. Straightforward derivation yields:

$$\mathrm{supDFC}(\tau_0) \Rightarrow \sup_{\tau \in [0,1-\tau_0]} \frac{\int_\tau^1 W(r) \mathrm{d}W(r)}{\sqrt{\int_\tau^1 W(r)^2 \mathrm{d}r}}.$$

Note that the limiting distribution is analogous to the one found in Section 2.4. In finite samples, the null distribution for both the supDFC and the supDF statistics are affected by the initial value of the time series if the series is not demeaned or detrended. To overcome this problem, we suggest to compute the test statistics by using the transformed series $\{\tilde{y}_t\}_{t=1}^T$ with $\tilde{y}_t = y_t - y_0$.

2.6 Further Test Procedures and Infeasible Point Optimal Tests

The test procedures presented so far fall into two categories: recursive DF t-statistics and tests based on scaled sums of forecast errors. Recursive DF t-tests have originally been proposed to test against stationary alternatives (cf. Banerjee, Lumsdaine, and Stock 1992 or Leybourne et al. 2003). In that case, lower tail critical values are appropriate. In order to test the I(1) hypothesis against explosive alternatives, Phillips, Wu, and Yu (2011) proposed the use of forward recursive DF t-statistics and upper tail critical values. In (e), we suggested DF t-statistics, which are essentially backward recursive. In the literature on tests for a change in persistence, several variants of Kim's (2000) and Busetti and Taylor's (2004) tests are available (cf. Taylor and Leybourne 2004; Taylor 2005). We have also adapted these tests to the bubble scenario, using the same logic as for the supBT and supK test. Monte Carlo simulations have shown, however, that the resulting procedures perform worse than the supDFC and the supBT tests in terms of power. To save space, these results are not reported here.

In Equation (14), we have assumed that the break fraction τ^* is unknown. If, instead, τ^* is known, point optimal tests can be constructed by using the Neyman–Pearson lemma. This allows to gauge the performance of the tests in Sections 2.1–2.5 relative to the power envelope. Under the additional assumption that the error terms in Equation (12) follow a normal distribution with known variance σ^2 and for fixed τ^* and ρ^* in Equation (14), the most powerful level-α test of H_0 against H_1 rejects, if

$$\mathrm{PO}(\tau^*,\rho^*) = \frac{1}{\sigma^2}\sum_{t=\tau^*+1}^{T}(y_t - y_{t-1})^2 - (y_t - \rho^* y_{t-1})^2 \; > \; k_\alpha(\tau^*,\rho^*), \tag{24}$$

where $k_\alpha(\tau^*,\rho^*)$ denotes the critical value with respect to a significance level of α. For a known break date, this test is optimal against the alternative $\rho = \rho^*$. Replacing ρ^* with the suitable local alternative $\rho^* = 1 + b/T$ with $b > 0$ and rearranging terms, the asymptotic distribution under H_0 is readily derived as

$$\mathrm{PO}(\tau^*, 1 + b/T) \; \Rightarrow \; 2b\int_{\tau^*}^{1} W(r)\mathrm{d}W(r) - b^2\int_{\tau^*}^{1} W(r)^2\mathrm{d}r.$$

By determining the rejection probabilities of the point optimal tests by means of Monte Carlo simulations, we are able to compute the power envelope for our testing problem.

2.7 Estimation of the Break Date

Assume that the time series under consideration, $\{y_t\}_{t=0}^{T}$, is described by Equations (12) and (14), where τ^* is unknown. An obvious way to estimate τ^* is to use the value of $\tau \in [0, 1 - \tau_0]$ that maximizes the statistic DFC_τ (22),

$$\hat{\tau}_{\mathrm{DFC}} = \underset{\tau\in[0,1-\tau_0]}{\operatorname{argmax}} \, \mathrm{DFC}_\tau. \tag{25}$$

The idea of this estimator is related to that in Leybourne et al. (2003). These authors consider the case of a change from I(0) to I(1) and propose a consistent estimator for the unknown break point. Note that this estimator also maximizes the likelihood function with respect to the break date. Bai and Perron (1998) have shown that the maximum likelihood estimator for the break date is consistent.[4]

[4]In a working paper version of this article, we also consider the break point estimator proposed by Phillips, Wu, and Yu (2011) and adapt the estimator of Busetti and Taylor (2004). As suggested by an associate editor, we do not present these estimators in this paper since $\hat{\tau}_{\mathrm{DFC}}$ has a clearer theoretic justifiaction and turns out to be more accurate in Monte Carlo simulations.

3 REAL-TIME MONITORING

The test statistics considered in Section 2 are designed to detect speculative bubbles within a fixed historical dataset. As argued by Chu, Stinchcombe, and White (1996), such test may be highly misleading when applied to an increasing sample. This is due to the fact that structural break tests are constructed as "one-shot" test procedures, that is, the (asymptotic) size of the test is controlled provided that the sample is fixed and the test procedure is applied only once to the same dataset (cf. Chu, Stinchcombe, and White 1996; Zeileis et al. 2005). To illustrate the problem involved, assume that an investor is interested to find out whether the stock price is subject to a speculative bubble. Applying the tests proposed in Section 2 to a sample of the last 100 trading days (say), he or she is not able to reject the null hypothesis of no speculative bubbles. If the stock price continues to increase in the subsequent days, the investor is interested to find out whether the evidence for a speculative bubble has strengthened. However, repeating the tests for structural breaks when new observations become available eventually leads to a severe over-rejection of the null hypothesis due to multiple application of statistical tests.

Another practical problem is that the tests assume a single structural break from a random-walk regime to an explosive process. The results of our Monte Carlo simulations in Section 4.3 show that the tests generally lack power if the bubble bursts within the sample, that is, if there is an additional structural break back to a random-walk process. The monitoring procedure suggested in this section is able to sidestep the problems due to multiple breaks.

Assume that, when the monitoring starts, a training sample of n observations is available, and that the null hypothesis of no structural break holds for the training sample. Then, in each period $n+1, n+2, \ldots$, a new observation arrives. As we will argue below, it is important to fix in advance the maximal length of the monitoring interval $n+1, n+2, \ldots, N = kn$ as the critical value depends on N. Following Chu, Stinchcombe, and White (1996), we consider two different statistics (detectors):

$$\text{CUSUM: } S_n^t = \frac{1}{\hat{\sigma}_t} \sum_{j=n+1}^{t} y_j - y_{j-1} = \frac{1}{\hat{\sigma}_t}(y_t - y_n) \quad (t > n), \tag{26}$$

$$\text{FLUC: } Z_t = (\hat{\rho}_t - 1)/\hat{\sigma}_{\rho_t} = \text{DF}_{t/n} \quad (t > n), \tag{27}$$

where $\hat{\rho}_t$ denotes the OLS estimate of the autoregressive coefficient, $\hat{\sigma}_{\rho_t}$ denotes the corresponding standard deviation, and $\hat{\sigma}_t^2$ is some consistent estimator of the residual variance based on the sample $\{y_0, \ldots, y_t\}$.

Note that Chu, Stinchcombe, and White (1996) suggest a fluctuation test statistic based on the coefficients $\hat{\rho}_t$. Since the coefficient ρ_t is equal to unity under the null hypothesis, the FLUC is essentially similar to the fluctuation statistic advocated by Chu, Stinchcombe, and White (1996). Also note that both the FLUC detector and the recursive DF test from Section 2 make use of standard DF t-statistics. However, the two procedures apply to different scenarios. While the latter is intended to analyze a given dataset with a fixed last observation, the former applies to a dataset that increases with the duration of the monitoring. Moreover, the recursive DF *test* is only concerned with whether or not a bubble has emerged within a given dataset, while for the FLUC and CUSUM *detectors*, it also plays a role how quickly a structural change is detected. Instead of using a constant critical value, one might prefer to use a critical boundary that increases during the monitoring phase. Given that the bubble starts at the beginning of the monitoring, an increasing instead of constant critical boundary should improve chances to detect the bubble quickly.

Under the null hypothesis, the functional central limit theorem implies as $n \to \infty$

$$\frac{1}{\sqrt{n}} S_n^{[\lambda n]} \Rightarrow W(\lambda) - W(1) \quad (1 \leqslant \lambda \leqslant k)$$

$$Z_{[\lambda n]} \Rightarrow \int_0^{\lambda} W(r)\mathrm{d}W(r) \Big/ \sqrt{\int_0^{\lambda} W(r)^2 \mathrm{d}r} \quad (1 \leqslant \lambda \leqslant k),$$

where $W(r)$ is a Brownian motion defined on the interval $r \in [0, k]$. Our CUSUM monitoring is based on the fact that for any $k > 1$ (see Chu, Stinchcombe, and White 1996)

$$\lim_{n\to\infty} P\left(|S_n^t| > c_t\sqrt{t} \text{ for some } t \in \{n+1, n+2, \ldots, kn\}\right) \leqslant \exp(-b_\alpha/2), \tag{28}$$

where

$$c_t = \sqrt{b_\alpha + \log(t/n)} \quad (t > n). \tag{29}$$

Since our null hypothesis is one sided (i.e., we reject the null hypothesis for large positive values of S_n^t) and S_n^t is distributed symmetrically around zero, a one-sided decision rule is adopted as follows. The null hypothesis is rejected if S_t exceeds the threshold c_t the first time, that is,

$$\text{reject } H_0 \text{ if } S_n^t > c_t\sqrt{t} \text{ for some } t > n. \tag{30}$$

For a significance level $\alpha = 0.05$, for instance, the one-sided critical value b_α used to compute c_t in Equation (29) is 4.6.

Such a test sequence has the advantage that if the evidence for a bubble process is sufficiently large, the monitoring procedure eventually stops before the bubble collapses. Accordingly, such a monitoring procedure sidesteps the problem of multiple breaks due to a possible burst of the bubble.

For the second monitoring using the statistic $\mathrm{DF}_{t/n}$, we apply the following rule:

$$\text{reject } H_0 \text{ if } \mathrm{DF}_{t/n} > \kappa_t \text{ for some } t = n+1, \ldots, N = kn, \tag{31}$$

where $\kappa_t = \sqrt{b_{k,\alpha} + \log(t/n)}$.[5] Since the limiting distributions of the CUSUM and the FLUC detectors differ and no theoretical result similar to that in Equation (28) is available for the FLUC detector, we determine the critical value $b_{k,\alpha}$ by means of simulation (see Section 4.4). This ensures that, under the null hypothesis, the probability of the event $\{\mathrm{DF}_{t/n} > \kappa_t,$ for some $t = n+1, \ldots, kn\}$ does not exceed α. It turns out that for the usual significance levels, $b_{k,\alpha}$ is monotonically increasing in k, the length of the monitoring period (including the training sample) relative to the training sample. Thus, the maximal size of monitoring period has to be fixed before starting the monitoring.

To account for a linear time trend in the data generating process (12), the time series can be detrended before computing the detectors. However, it is well known that the DF t-statistic possesses a sizable negative mean and, therefore, the critical values for the detrended FLUC monitoring may be negative. However, our boundary function κ_t is restricted to positive values. To overcome this problem, the FLUC detector is computed by using the standardized DF t-statistics $\frac{Z_t - m_{\mathrm{DF}}}{\sigma_{\mathrm{DF}}}$. The asymptotic first moment m_{DF} and standard deviation σ_{DF} of the DF t-statistic are taken from Nabeya (1999), where $m_{\mathrm{DF}} = -2.1814$ and $\sigma_{\mathrm{DF}} = 0.7499$. Regarding the CUSUM procedure, instead of using the OLS-detrended series to compute S_t in Equation (26), one can replace the forecast error $y_j - y_{j-1}$ with $w_j = \sqrt{\frac{j-1}{j}}(y_j - y_{j-1} - \hat{\mu}_{j-1})$, where $\hat{\mu}_{j-1} = (j-1)^{-1}\sum_{l=1}^{j-1}\Delta y_l$. Note that w_j is the recursive CUSUM residual in the regression of Δy_t on a constant. As is well known, the same asymptotic results for S_n^t hold when this replacement is made. This means that one can proceed as in the case without drift and use the same boundary function $c_t = \sqrt{b_\alpha + \log(t/n)}$ with the same values for b_α.

4 MONTE CARLO ANALYSIS

We start our Monte Carlo analysis within our basic framework in Equations (12)–(14). In Section 4.1, we report critical values, present the results for the empirical power of the tests, and evaluate the break point estimator $\hat{\tau}_{\mathrm{DFC}}$. Furthermore, we consider price processes that contain explicitly modeled bubbles. We investigate the power of the tests to detect randomly starting bubbles in Section 4.2. Periodically collapsing bubbles are considered in Section 4.3, where we apply both tests and monitoring procedures. In Section 4.4, we present further results for the monitoring procedures.

4.1 The Basic Framework: Tests and Break Point Estimation

We use Monte Carlo simulation to calculate critical values for the test statistics supDF, supB, supBT, supK, supDFC, and for the point optimal statistics. Here and

[5] It is possible to employ different functional forms of the boundary function κ_t. Our choice is motivated by facilitating comparisons between the performance of CUSUM and FLUC monitoring.

in the remainder of the paper, we set $\tau_0 = 0.1$ for all test statistics and estimators. The data are generated according to Equation (12) with $\rho_t = 1$ (for all t), an initial value $y_0 = 0$, and Gaussian white noise. To approximate asymptotic critical values, we use a sample size of $T = 5000$. The number of replications is 10,000. We apply the test statistics to the original and to the detrended series, that is, to the residuals from the regression of y_t on a constant and a linear time trend. The results are reported in Table 1. To save space, we leave out the critical values for the point optimal tests. These tests are of little practical use, when the break date is unknown.

To evaluate the empirical power of the tests, we generate data according to Equations (12) and (14) with Gaussian white noise. Two thousand replications are performed for the sample sizes $T = 100$, $T = 200$, and $T = 400$. We consider a range of different break points τ^* and growth rates ρ^*. The power of the tests is evaluated at a nominal size of 5%, that is, a test rejects the null hypothesis when the corresponding statistic is larger than the respective asymptotic 0.95 quantile in Table 1. The results are reported in Tables 2 and 3. The row labeled "actual size" shows that the size of the tests is close to the nominal size, that is, the asymptotic critical values also apply to finite samples. Only the supB test is somewhat undersized. The actual size of the point optimal tests depends, apart from the sample size, on τ^* and on the value ρ^* of the autoregressive parameter under H_1. The actual size of the point optimal tests ranges between 4.3% and 5.7%.

With regard to testing power, the supDFC test and the supBT test exhibit the best performance among those tests that do not use knowledge of the true break date or of ρ^*. Moreover, these two tests come close to the power envelope computed from the infeasible point optimal tests. The difference between the power

Table 1 Large sample upper tail critical calues for test statistics

	Test statistics				
Quantiles	supDF	supDFC	supK	supBT	supB
(a) Critical values without detrending					
0.90	2.4152	1.5762	31.4531	1.9317	3.2796
0.95	2.7273	1.9327	43.7172	2.4748	3.9253
0.99	3.3457	2.6285	79.5410	3.8878	5.3746
(b) Critical values with detrending					
0.90	0.5921	0.9436	28.400	1.7374	2.7614
0.95	0.8726	1.3379	38.072	2.2736	3.3472
0.99	1.4176	2.0741	64.863	3.6088	4.6162

The critical values are estimated by simulation of Equations (12)–(13) using Gaussian white noise, a sample size of $T = 5000$, and 10,000 replications.

Table 2 Empirical power

Break point		Test statistics					
		PO(τ^*, ρ^*)	supDF	supDFC	supK	supBT	supB
(a) Power for $T = 100$							
	Actual size		0.057	0.050	0.049	0.060	0.023
$\tau^* = 0.7$	$\rho^* = 1.02$	0.347	0.166	0.312	0.085	0.282	0.128
	$\rho^* = 1.03$	0.526	0.293	0.483	0.137	0.466	0.218
	$\rho^* = 1.04$	0.679	0.429	0.634	0.211	0.615	0.342
	$\rho^* = 1.05$	0.780	0.559	0.750	0.316	0.741	0.459
$\tau^* = 0.8$	$\rho^* = 1.02$	0.273	0.107	0.247	0.067	0.214	0.072
	$\rho^* = 1.03$	0.414	0.181	0.379	0.088	0.348	0.118
	$\rho^* = 1.04$	0.545	0.264	0.508	0.117	0.468	0.168
	$\rho^* = 1.05$	0.662	0.369	0.605	0.171	0.589	0.240
$\tau^* = 0.9$	$\rho^* = 1.02$	0.177	0.069	0.169	0.054	0.139	0.031
	$\rho^* = 1.03$	0.276	0.086	0.238	0.061	0.207	0.034
	$\rho^* = 1.04$	0.364	0.112	0.322	0.068	0.288	0.035
	$\rho^* = 1.05$	0.460	0.150	0.397	0.077	0.372	0.034
(b) Power for $T = 200$							
	Actual size		0.059	0.054	0.039	0.055	0.031
$\tau^* = 0.7$	$\rho^* = 1.02$	0.694	0.439	0.633	0.216	0.615	0.451
	$\rho^* = 1.03$	0.870	0.673	0.810	0.455	0.802	0.676
	$\rho^* = 1.04$	0.944	0.811	0.905	0.764	0.902	0.813
	$\rho^* = 1.05$	0.973	0.901	0.944	0.894	0.946	0.886
$\tau^* = 0.8$	$\rho^* = 1.02$	0.572	0.271	0.504	0.135	0.467	0.282
	$\rho^* = 1.03$	0.779	0.472	0.698	0.231	0.686	0.478
	$\rho^* = 1.04$	0.876	0.644	0.810	0.430	0.802	0.640
	$\rho^* = 1.05$	0.931	0.761	0.876	0.690	0.873	0.736
$\tau^* = 0.9$	$\rho^* = 1.02$	0.381	0.114	0.328	0.070	0.283	0.088
	$\rho^* = 1.03$	0.560	0.185	0.481	0.095	0.440	0.156
	$\rho^* = 1.04$	0.698	0.303	0.606	0.135	0.584	0.230
	$\rho^* = 1.05$	0.795	0.404	0.705	0.181	0.692	0.303

The empirical power is computed at a nominal size of 5%. The simulations are conducted with 2000 replications of Equations (12) and (14). The true (simulated) breakpoint τ^* is measured relative to the sample size. ρ^* is the autoregressive parameter under the explosive regime. In the row labeled actual size, we report rejection frequencies when the data generating process obeys H_0.

of the supDFC test and the power envelope is never larger than 10% and in many cases only about 5% or smaller.[6] The supBT test performs comparably well. Taking

[6]For a fixed value of ρ among 1.02, 1.03, 1.04, 1.05, and known variance σ^2, we also considered statistics of the form $\sup_{\tau \in [0,0.9]} \mathrm{PO}(\tau, \rho)$ [cf. Equation (24)]. None of the resulting feasible tests dominates the supDFC test in terms of empirical power. That is, only for a few parameter constellations (T, τ^*, ρ^*)

Table 3 Empirical power cont

Break point		Test statistics					
		PO(τ^*, ρ^*)	supDF	supDFC	supK	supBT	supB
(a) Power for $T = 400$							
	Actual size		0.053	0.045	0.040	0.056	0.039
$\tau^* = 0.7$	$\rho^* = 1.02$	0.938	0.824	0.907	0.759	0.904	0.844
	$\rho^* = 1.03$	0.991	0.942	0.978	0.951	0.975	0.949
	$\rho^* = 1.04$	0.999	0.984	0.992	0.986	0.992	0.984
	$\rho^* = 1.05$	1.000	0.995	0.998	0.994	0.998	0.992
$\tau^* = 0.8$	$\rho^* = 1.02$	0.855	0.655	0.817	0.444	0.815	0.694
	$\rho^* = 1.03$	0.954	0.851	0.930	0.824	0.932	0.870
	$\rho^* = 1.04$	0.986	0.934	0.974	0.938	0.973	0.943
	$\rho^* = 1.05$	0.997	0.970	0.990	0.972	0.991	0.969
$\tau^* = 0.9$	$\rho^* = 1.02$	0.687	0.289	0.593	0.115	0.579	0.337
	$\rho^* = 1.03$	0.844	0.503	0.772	0.270	0.771	0.537
	$\rho^* = 1.04$	0.917	0.670	0.862	0.554	0.864	0.673
	$\rho^* = 1.05$	0.960	0.776	0.915	0.728	0.919	0.743

The empirical power is computed at a nominal size of 5%. The simulations are conducted with 2000 replications of Equations (12) and (14). The true (simulated) breakpoint τ^* is measured relative to the sample size. ρ^* is the autoregressive parameter under the explosive regime. In the row labeled actual size, we report rejection frequencies when the data generating process obeys H_0.

into account that the original version of Busetti and Taylor (2004) was constructed to test for a change from I(0) to I(1) and not from I(1) to explosive, the favorable performance is quite remarkable. The power of the supB test is comparable with that of the supDF test if $T \geqslant 200$. For $T = 100$, supB performs worse than supDF, which is probably due to the fact that supB is undersized.[7] The supK test lacks power if the sample size is small. Note that the supDFC test and the supBT test perform better than the supDF test in all parameter constellations. The advantage over the supDF test tends to increase as the break fraction τ^* increases. For instance, when $T = 400$, $\rho^* = 1.03$, and $\tau^* = 0.7$, the power of the three tests is approximately equal, while for $T = 400$, $\rho^* = 1.03$, and $\tau^* = 0.9$, the supDF test rejects the null hypothesis in only 50% of the cases, whereas the supDFC test and the supBT test can reject the null in more than 75% of the cases.

Results regarding the break point estimator $\hat{\tau}_{\text{DFC}}$ in Equation (25) are reported in Table 4. Equations (12) and (14) are used to generate the data. The experiments were conducted for different break points τ^* and for sample sizes $T = 200$ and

for the data generating process can supDFC be sizeably outperformed, and there are parameter choices where supDFC performs better (not shown).

[7] Further simulations have shown that if finite sample critical values are used, supB performs very similar to supDF for all parameter constellations.

Table 4 Break date estimates of $\hat{\tau}_{DFC}$ when $\rho^* = 1.05$

Break point	$\tau^* = 0.4$	$\tau^* = 0.5$	$\tau^* = 0.6$	$\tau^* = 0.7$	$\tau^* = 0.8$	$\tau^* = 0.9$
$T = 200$	0.4616	0.5582	0.6493	0.7388	0.8184	0.8682
	(0.1037)	(0.0980)	(0.0862)	(0.0747)	(0.0612)	(0.0887)
$T = 400$	0.4232	0.5207	0.6208	0.7207	0.8143	0.8890
	(0.0456)	(0.0436)	(0.0441)	(0.0427)	(0.0329)	(0.0512)

The table reports means and standard deviations (in parentheses) of the break point estimator $\hat{\tau}_{DFC}$. These values as well as the true breakpoint τ^* are expressed as fraction of the sample size T. Data are generated according to Equations (12) and (14) for the sample size $T = 200$ and $T = 400$. The autoregressive parameter under the explosive regime is $\rho^* = 1.05$. In each case, 2000 replications are used.

$T = 400$. The autoregressive parameter ρ^* was set to 1.05. For each parameter constellation, 2000 replications were performed. The table shows the empirical mean and standard deviation (in parentheses) of the estimates. The estimates are reasonably close to the true breakpoint. Additional (unreported) simulations also show that $\hat{\tau}_{DFC}$ is more reliable than the estimator proposed by Phillips, Wu, and Yu (2011) or an adapted versions of Busetti and Taylor's (2004) estimator.

4.2 Randomly Starting Bubbles

We now investigate how well the tests can detect randomly starting bubbles. We use the model presented in Section 1. That is, we simulate the price process as $P_t = P_t^f + B_t$, where B_t is generated as in Equation (7) and P_t^f follows Equations (10) and (11). The parameters are also specified as in Section 1. Since the fundamental price follows a random walk with drift, we detrend the time series before applying the tests. The results are given in Table 5. The null and the alternative hypotheses are as in Equations (12)–(14). The nominal size is 5% and the sample sizes are $T = 100$ or $T = 200$. For each parameter constellation, 2000 replications of the price process $P_t = P_t^f + B_t$ are generated. Cases in which a bubble remains at its initial value for more than 90% of its realizations are excluded from the experiment. Note that for the first row of Table 5, no bubble component enters the price series. This suggests that all tests have correct size.

When the sample size equals $T = 100$, the supDFC test and the supBT test have the highest rejection frequencies. In all cases, they reject the null hypothesis at least 20% more often than the supDF test or the supB test. As in Section 4, the supK test performs worst. When the sample size is $T = 200$ and B_0 is 0.05,[8] the differences between the tests are less pronounced. Still, the supK test shows the weakest performance among the tests, while the supDFC test and the supBT test have the highest empirical power.

[8]When B_0 is set to 2, all tests, except the supK test, have rejection frequencies close to 1 for $\pi \geqslant 0.05$.

Table 5 Rejection frequencies in the case of randomly starting bubbles

		Test statistics				
Initial value	π	supDF	supDFC	supK	supBT	supB
Rejection frequencies when the sample size is $T = 100$						
$B_0 = 2$	No bubble	0.036	0.049	0.034	0.060	0.020
	0.02	0.419	0.636	0.054	0.647	0.413
	0.05	0.382	0.741	0.051	0.732	0.388
	0.10	0.343	0.825	0.061	0.829	0.365
	0.25	0.288	0.907	0.067	0.907	0.365
	0.50	0.277	0.930	0.072	0.924	0.359
	1.00	0.272	0.931	0.072	0.929	0.367
Rejection frequencies when the sample size is $T = 200$						
$B_0 = 0.05$	No bubble	0.037	0.047	0.038	0.060	0.031
	0.01	0.558	0.645	0.104	0.657	0.566
	0.02	0.672	0.770	0.117	0.775	0.678
	0.05	0.863	0.935	0.158	0.938	0.854
	0.10	0.968	0.990	0.169	0.990	0.957

The tests are applied to series generated as $P_t = P_t^f + B_t$, with P_t^f as in Equation (11) and B_t as in Equation (7). The number of replications is 2000. The "no-bubble" hypothesis is rejected when the test statistic exceeds the 95% quantile from Table 1. B_0 is the initial value of the bubble. π is the probability that the bubble enters the phase of explosive growth.

It is interesting to note that for $T = 100$ and $B_0 = 2$, the power of some tests does not increase if the probability of the bubble starting at the next period increases. This seems odd since the expected time for the start of the bubble, $1/\pi$, decreases with increasing probability π. However, the parameter π also has another effect on the bubble process. As can be seen in Equation (7), there is a jump in the bubble process at the time when it starts to grow. The jump size is however inversely related to π. These two effects may offset each other, yielding an ambiguous effect on the power. This phenomenon can be observed for the supDF test, the supK test, and the supB test.

Summing up, the simulation experiments with randomly starting bubbles are very much in line with the results of Section 4.1, that is, the supDFC and the supBT tests clearly outperform all other tests.

4.3 Periodically Collapsing Bubbles

We now analyze how well the tests are suited to detect Evans' (1991) periodically collapsing bubbles. The price process $P_t = P_t^f + 20B_t$ is generated as described in

Section 1. Again, all tests are applied to detrended data. As in the previous section, rejections at the 5% significance level are considered as evidence for a speculative bubble.

Table 6 reports rejection frequencies of both test and monitoring procedures. In the Monte Carlo experiment, 2000 replications of the price process P_t have been generated for a sample size of $T = 100$ and different values for the probability π that the bubble continues. First, we discuss the results for the test procedures, whereas results for the monitoring scheme are considered in the next section. Obviously, the power of the tests decreases if the probability π gets smaller. The more striking finding is, however, the superiority of the supDF test relative to all the other tests. For instance, if $\pi = 0.85$, the supDF test rejects in roughly 60% of the cases, while the supK test rejects in 32% of the cases and the other tests have rejection frequencies less than 10%.

There is an intuitive explanation for the superiority of the supDF test. In contrast to the other tests, the supDF test is based on subsamples of the form $\{P_1, \ldots, P_{[\tau T]}\}$. The DF t-statistic for this subsample tends to be large, if it contains a long period with a steady growth of the bubble up to the end of the subsample. An obvious case is provided in Figure 3. It shows a realization of the price process P_t with $\pi = 0.85$ and a corresponding plot of the sequence of DF_τ statistics.

In contrast to the supDF test, the idea behind the supK test is to compare the first part of the sample $\{P_1, \ldots, P_{[\tau T]}\}$ to the second part $\{P_{[\tau T]+1}, \ldots, P_T\}$. However, if the bubble collapses several times during the whole sample, the subsamples mix up periods with and without a bubble. Therefore, the test statistics that are based on a comparison of the pre- and postbreak subsamples have difficulties to indicate a bubble component. Thus, among the alternative tests, the supDF test

Table 6 Rejection frequencies in the case of periodically collapsing bubbles

	Test statistics					Detectors	
π	supDF	supDFC	supK	supBT	supB	FLUC	CUSUM
No bubble	0.043	0.049	0.033	0.062	0.020	0.045	0.045
0.999	0.803	0.881	0.045	0.934	0.117	0.947	0.942
0.990	0.824	0.589	0.192	0.642	0.119	0.788	0.763
0.950	0.715	0.164	0.371	0.223	0.076	0.618	0.535
0.850	0.593	0.057	0.324	0.072	0.026	0.535	0.379
0.750	0.515	0.040	0.261	0.040	0.018	0.467	0.269
0.500	0.393	0.020	0.235	0.022	0.010	0.342	0.031
0.250	0.248	0.020	0.190	0.023	0.010	0.205	0.015

The tests and monitoring procedures are applied to series generated as $P_t = P_t^f + 20B_t$, with P_t^f as in Equation (11) and B_t as in Equation (8). The sample size equals $T = 100$. π is the probability that the bubble does not collapse in the next period. For each case, 2000 replications are performed. The nominal size of the test and monitoring procedure is 5%.

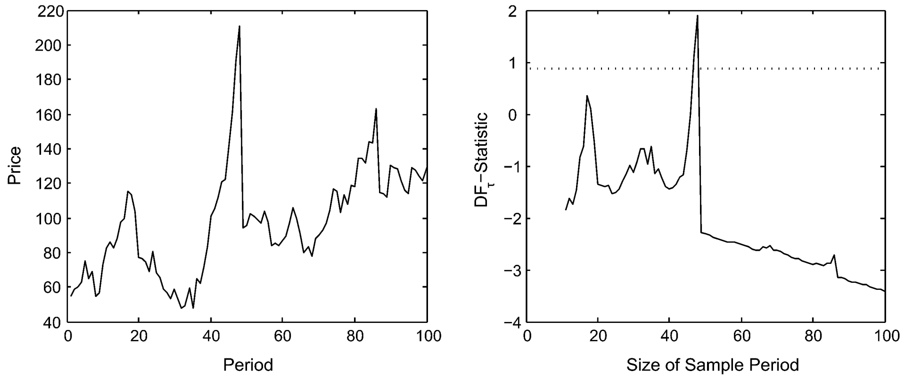

Figure 3 Price containing periodically collapsing bubble (left) and DF$_\tau$-statistic (right).

has the highest power when applied to periodically collapsing bubbles. However, in applications, one typically has a clear indication for the end of the explosive regime, just from visual inspection of the time series. If the subsequent observations are excluded from the sample, the supDFC test and the supBT test are most powerful among the tests considered here.

4.4 Monitoring

The critical values $b_{k,\alpha}$ in the expression for the boundary function $\kappa_t = \sqrt{b_{k,\alpha} + \log(t/n)}$ from Section 3 are reported in Tables 7 and 8 for different values of the significance level α and different sizes of the monitoring sample k (measured relative to the training sample). They have been obtained through application of the monitoring procedures to data simulated according to Equation (12) under the random-walk null hypothesis with Gaussian white noise innovations. We tabulate critical values for training sample sizes of $n = 20$, $n = 50$, and $n = 100$. It is important to note that $b_{k,\alpha}$ is monotonic increasing in k and the maximal length of the monitoring period as measured by k has to be fixed before starting the monitoring. Table 8 reports values $b_{k,\alpha}$ for CUSUM monitoring in finite samples. Using the asymptotic 5% critical value, $b_\alpha = 4.6$ can lead to an empirical size below 1%, when the CUSUM monitoring procedure includes estimation of a drift term.

The results for the FLUC and CUSUM monitoring (with detrending) when applied to periodically collapsing bubbles are reported in the last two columns of Table 6. The first $n = 20$ observations are used as training sample and the monitoring ends after 100 observations. The tests are conducted at a nominal size of 5%, that is, we use the finite sample critical values $b_{5,0.05}$ with $n = 20$. The residual variance for the CUSUM monitoring is estimated as $\hat{\sigma}_t^2 = (t-1)^{-1} \sum_{j=1}^{t} (\Delta y_j - \hat{\mu}_t)^2$, where $\hat{\mu}_t$ is the mean of $\{\Delta y_1, \ldots, \Delta y_t\}$. Similar to the supDF test, the monitoring procedures are more robust against periodically collapsing bubbles. However,

Table 7 Critical values for FLUC monitoring

n	k / α	2	3	4	5	6	8	10
i) FLUC monitoring without detrending								
100	0.10	3.05	3.60	3.93	4.15	4.31	4.48	4.57
	0.05	4.50	5.14	5.55	5.69	5.89	6.05	6.26
	0.01	7.76	8.59	9.06	9.48	9.62	9.79	9.99
50	0.10	2.80	3.33	3.62	3.80	3.96	4.14	4.27
	0.05	4.19	4.80	5.11	5.34	5.50	5.72	5.81
	0.01	7.30	8.11	8.43	8.82	8.86	9.25	9.49
20	0.10	2.49	3.12	3.44	3.65	3.78	3.99	4.12
	0.05	3.88	4.56	4.86	5.06	5.19	5.38	5.52
	0.01	7.00	7.84	8.26	8.49	8.66	9.12	9.19
ii) FLUC monitoring with detrending								
100	0.10	6.16	7.55	8.15	8.53	8.98	9.30	9.51
	0.05	8.12	9.82	10.45	10.82	11.20	11.55	11.80
	0.01	12.89	14.22	14.94	15.39	16.22	16.14	16.39
50	0.10	5.47	6.74	7.15	7.69	8.07	8.54	8.80
	0.05	7.29	8.68	9.30	9.62	10.11	10.46	10.87
	0.01	11.54	12.46	13.57	13.81	14.08	14.71	15.44
20	0.10	4.18	5.16	5.71	6.17	6.38	6.78	7.24
	0.05	5.50	6.52	7.24	7.96	8.04	8.45	9.13
	0.01	8.45	9.55	10.82	11.41	11.93	12.35	13.08

The table shows critical values for FLUC monitoring for different significance levels α, different lengths n of the training sample, and for different lengths k of the monitoring period (measured in multiples of n).

except for one case, the monitoring procedures have less power than the supDF test. Moreover, the FLUC monitoring performs better than the CUSUM monitoring. There are two conceptual differences between the supDF test and the FLUC monitoring. First, the critical value of the FLUC monitoring increases in time. Second—and more important—the different detrending methods give rise to the different performance of these two related approaches. Before applying the supDF test, the time series is detrended once using the entire sample. In contrast, the monitoring approach uses a sequential detrending scheme. Additional simulations suggest that the empirical power of the FLUC monitoring is virtually identical to that of the supDF test if the FLUC monitoring uses the same detrending method (i.e., the residuals of a full sample regression) as the supDF test.

We also evaluate the performance of the monitoring procedures in terms of time needed to detect a bubble, a property that is irrelevant for the test procedures. The results are presented in Table 9. The data were generated as in Equations (12)

Table 8 Critical values for CUSUM monitoring

n	k / α	2	3	4	5	6	8	10
i) CUSUM monitoring without drift estimation								
100	0.10	0.92	1.39	1.62	1.80	1.92	2.08	2.19
	0.05	1.51	2.14	2.47	2.73	2.88	3.20	3.36
	0.01	2.86	3.92	4.57	4.94	5.30	5.72	6.02
50	0.10	0.87	1.31	1.55	1.70	1.82	1.97	2.06
	0.05	1.43	2.03	2.34	2.62	2.80	3.03	3.12
	0.01	2.85	3.87	4.25	4.77	5.03	5.47	5.94
20	0.10	0.81	1.18	1.43	1.61	1.75	1.91	2.02
	0.05	1.27	1.96	2.35	2.53	2.72	3.00	3.13
	0.01	2.61	3.91	4.49	4.97	5.16	5.52	5.74
ii) CUSUM monitoring with drift estimation								
100	0.10	0.92	1.38	1.59	1.78	1.90	2.06	2.16
	0.05	1.45	2.11	2.43	2.66	2.80	3.06	3.18
	0.01	2.76	3.78	4.49	4.84	5.12	5.68	5.96
50	0.10	0.91	1.32	1.56	1.69	1.80	1.96	2.08
	0.05	1.43	2.03	2.37	2.57	2.75	2.99	3.14
	0.01	2.71	3.90	4.39	4.74	4.97	5.43	5.62
20	0.10	0.83	1.20	1.43	1.60	1.72	1.86	2.00
	0.05	1.34	1.90	2.24	2.47	2.67	2.88	3.02
	0.01	2.65	3.76	4.42	4.77	5.01	5.53	5.65

The table shows critical values for CUSUM monitoring for different significance levels α, different lengths n of the training sample, and for different lengths k of the monitoring period (measured in multiples of n).

and (14) with slope coefficient $\rho^* = 1.03$ in the explosive regime. The total sample size was set to $N = 200$ and $N = 400$. The size of the training sample was set to $n = 40$ and $n = 80$, respectively. We used 5% critical values $b_{5,0.05}$ with training sample size $n = 50$ and $n = 100$, respectively. The residual variance for the CUSUM monitoring is estimated as $\hat{\sigma}_t^2 = t^{-1}\sum_{j=1}^{t}(\Delta y_j)^2$. The breakpoint τ^* is measured relative to the total sample size. For each case, 5000 replications were performed.

The actual size is given in the first row of each panel in Table 9. The columns labeled *delay* report the mean delay (measured relative to the sample size) for detecting the bubble. Since a rejection before the bubble starts is spurious, only rejections after the start of the bubble enter the mean delay and standard deviations. Table 9 shows that the empirical power of the CUSUM monitoring is comparable to that of the FLUC monitoring. The average time needed to detect the bubble is also very similar for both monitoring procedures. The mean and the standard

Table 9 Performance of monitoring procedures

	FLUC			CUSUM		
Breakpoint	Size/power	Delay	σ(delay)	Size/power	Delay	σ(delay)
a) Size of training sample $n = 40$, size of total sample $kn = 200$						
Actual size	0.0570	—	—	0.0574	—	—
$\tau^* = 0.2$	0.9766	0.2787	0.1779	0.9744	0.2816	0.1600
$\tau^* = 0.4$	0.9330	0.2454	0.1379	0.9310	0.2415	0.1369
$\tau^* = 0.6$	0.8148	0.2019	0.0969	0.8090	0.2042	0.0982
$\tau^* = 0.8$	0.4696	0.1243	0.0471	0.4588	0.1255	0.0492
b) Size of training sample $n = 80$, size of total sample $kn = 400$						
Actual size	0.0512	—	—	0.0488	—	—
$\tau^* = 0.2$	0.9996	0.1368	0.0977	0.9996	0.1448	0.0906
$\tau^* = 0.4$	0.9974	0.1318	0.0900	0.9972	0.1352	0.0908
$\tau^* = 0.6$	0.9854	0.1250	0.0782	0.9844	0.1319	0.0803
$\tau^* = 0.8$	0.8362	0.0977	0.0474	0.8126	0.1032	0.0484

The data are generated as a random-walk switching to an explosive process at the relative break point τ^* [cf. Equations (12) and (14)]. The autoregressive parameter is set to $\rho^* = 1.03$ during the explosive phase. The number of replications is 5000. The empirical power is computed at the 5% significance level. In the column delay, we report the average time needed to detect the regime switch relative to the sample size kn. The column σ(delay) contains standard deviations of the detection delay relative to the sample size.

deviation for both procedures decrease as the time of the break date τ^* increases. This is probably due to the fact that the maximum delay is bounded by $1 - \tau^*$.

5 APPLICATIONS

In Section 5.1, we start with a detailed discussion of the Nasdaq Composite Index and the so-called dot.com bubble. A range of other financial time series that are often supposed to be affected by speculative bubbles are considered in Section 5.2.

5.1 The Nasdaq Composite Index and the Dot.Com Bubble

A price movement for which the term *bubble* is widely used is the development of the Nasdaq composite price index in the late 1990s. A mere look at the plot of this time series appears to justify the name "dot.com bubble" (see Figure 4). Phillips, Wu, and Yu (2011) found decisive empirical evidence for a speculative bubble during this period (see Shiller 1981 for a detailed discussion of this speculative bubble). In this section, we reproduce their findings and also apply the alternative tests presented in Section 2.

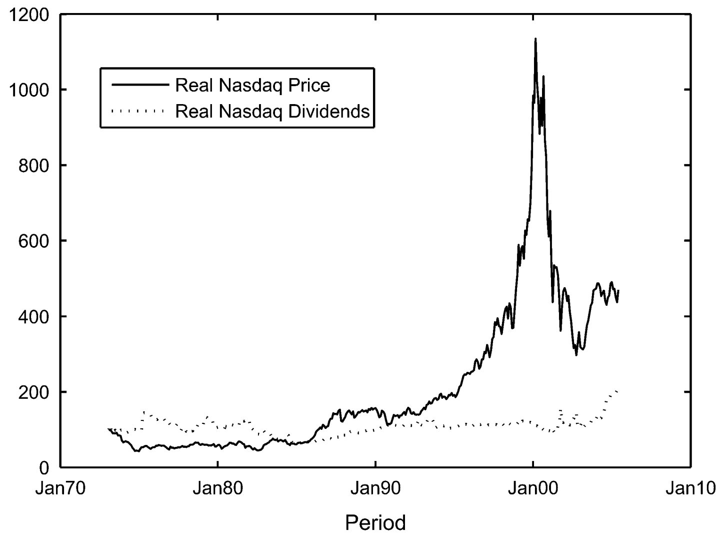

Figure 4 Real Nasdaq price and dividends.

We use the same data as in Phillips, Wu, and Yu (2011). The times series of the Nasdaq composite price index and the Nasdaq composite dividend yield are taken from *Datastream International*. Phillips, Wu, and Yu (2011) considered monthly Nasdaq data from February 1973 to June 2005. Figure 4 suggests that if there was a bubble, it has definitely crashed in March 2000, after the Nasdaq reached its all time high. If the observations after the peak are included, the chance that the bubble tests detect the bubble will be very low (see Section 4.3). Therefore, we apply the tests to the restricted sample period from February 1973 to March 2000, which yields $T = 326$ observations. The Nasdaq composite price index is multiplied by the Nasdaq composite dividend yield to compute the time series of the total Nasdaq dividends. By use of the Consumer Price Index from the Federal Reserve Bank of St. Louis, nominal data are transformed to real data. Figure 4 shows a plot of the time series for the real Nasdaq price index and the real Nasdaq dividends, where the time series have been normalized to 100 at the beginning of the sampling period. While the price index shows an accelerated increase during the late 1990s followed by a sharp drop, real index dividends are more or less constant.

For the logarithm of these two time series, we conduct tests of the random-walk hypothesis against the alternative of a switch from I(1) to explosive. If the time series of the Nasdaq composite price index turns out to have changed from I(1) to explosive, while the index dividend series remained constant I(1), this would suggest the presence of bubbles.

In the applications, we follow Phillips, Wu, and Yu (2011) more closely and use an augmented Dickey–Fuller (ADF) test that includes an intercept term in the regression, that is, the regression equation is

Table 10 Testing for an explosive root in the Nasdaq index

	Test statistics				
	supADF	supDFC	supK	supBT	supB
Log price index	3.0420	4.5874	12.5203	7.8722	3.0242
Log dividends	−0.8563	−0.2086	5.0756	0.1438	2.0057
Upper tail critical values					
1%	2.094	2.6285	79.5410	3.8878	5.3746
5%	1.468	1.9327	43.7172	2.4748	3.9253
10%	1.184	1.5762	31.4531	1.9317	3.2796

This Table reports the values of the test statistics applied to the log real Nasdaq price index and dividends. The table also shows the corresponding critical values. Those for the supADF statistic are taken from Phillips, Wu, and Yu (2011). The others correspond to our simulation results.

$$y_t = \alpha + \rho y_{t-1} + \sum_{j=1}^{p} \phi_j \Delta y_{t-j} + \epsilon_t, \quad \text{for } t = 1, \ldots, [\tau T], \tag{32}$$

where y_t denotes the logarithm of either the Nasdaq price or dividends and ϵ_t is white noise. The lag order p is determined by a general-to-specific test sequence. Equation (32) leads to the augmented DF t-statistic ADF_τ. The supremum of ADF_τ over $\tau \in [0.1, 1]$ is referred to as supADF. The pertaining critical values are taken from Phillips, Wu, and Yu (2011).

The results are shown in Table 10. For the dividend series, none of the tests rejects the constant I(1) hypothesis at the 5% significance level. Even at the 10% significance level, the null hypothesis is not rejected. The application of the supADF test leads to results that are very similar to those obtained by Phillips, Wu, and Yu (2011).[9] The supDFC test and the supBT test indicate a bubble in the Nasdaq index, while the supB test and the supK test fail to reject the null hypothesis. Most of the results do not change when we consider weekly or daily data, the only exception being the supB test, which becomes significant for higher sampling frequencies (not shown). Thus, the supDF test, the supDFC test, the supBT test, and partly the supB test support the view that the Nasdaq price index has changed from I(1) to explosive, while the dividend series remained I(1) throughout the sample. Under the assumptions of the present value model in Equations (1) and (2), one may conclude that a bubble was indeed present in the Nasdaq index. Finally, the estimator $\hat{\tau}_{\mathrm{DFC}}$ identifies the starting date of the bubble as $\hat{\tau}_{\mathrm{DFC}} = 72.7\%$, corresponding to October 1992.

[9]They report a test value of 2.894 for the index price and a value of −1.018 for the index dividends.

5.2 Further Applications: Major Stock Indices, House Prices, and Commodities

Besides the Nasdaq Composite index, other financial time series have shown phases of massive growth, often combined with subsequent crashes. Frequently, the term bubble has been used to describe such phases. Specifically, we consider the Japanese stock market index, Nikkei225, which, during the 1980s, exhibited a tremendous increase along with Japanese urban land prices (see Figures 5 and 6). We also check whether the dot.com bubble has been paralleled by bubbles in the S&P500 and the FTSE100. Furthermore, we analyze more recent upswings in major Chinese stock market indices, the Hang Seng and the Shanghai stock index. During the recent subprime crisis, housing markets have received much interest from financial analysts and researchers. We consider the US S&P/Case–Shiller home price index (10-City composite) among other housing indices. Finally, we also test for bubbles in two commodity time series, gold and crude oil. Note that house and commodity prices do not directly fit into the theoretical framework of rational bubbles, in which the fundamental price is based on a stream of future dividend payments. Nevertheless, detecting a change from I(1) to explosive clearly points to excessive speculation.

Our results are presented in Table 11. We focus on the supDFC test, which together with supBT performed very well in our Monte Carlo simulations, and on the supDF test. In most cases, the supBT test and the supDFC test lead to similar results, so we skip reporting the results for the former. We apply the tests to the logarithm of the inflation-adjusted time series at different sampling frequencies: monthly, weekly, and daily. To adjust weekly and daily data for inflation, we employ linearly interpolated monthly consumer price indices. Data for house/land

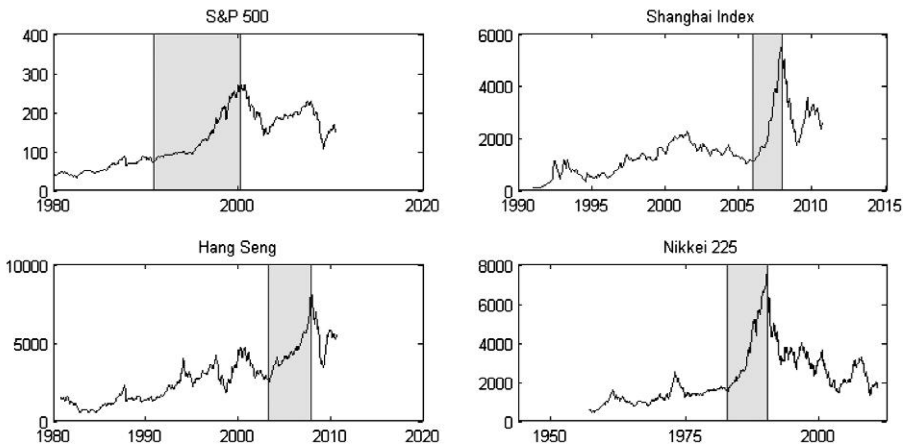

Figure 5 The figure shows several stock price indices with time on the x-axis and index values on the y-axis.

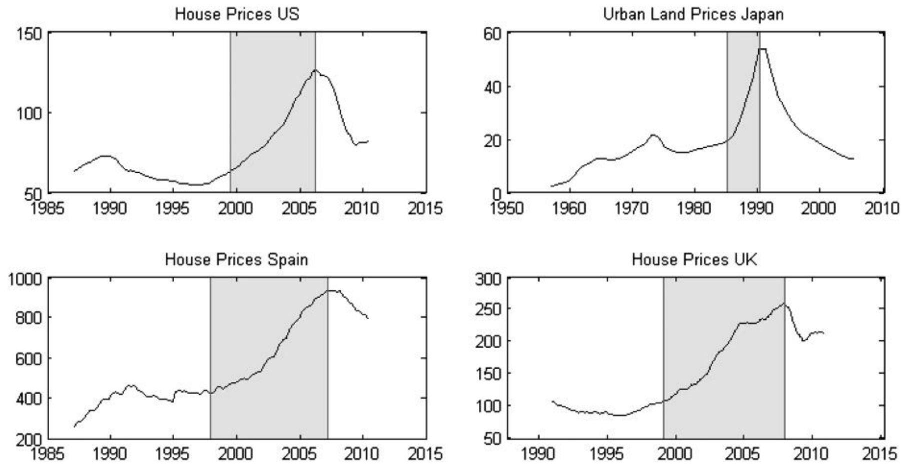

Figure 6 The figure shows several house/land price indices with time on the x-axis and index values on the y-axis.

price indices were only available at monthly, quarterly, or annual frequency. The range of the sampling period for each time series is given in Table 11. As in the foregoing section, the price series end at their maximum. With regard to the real gold price, we focus on the recent run-up in gold prices and consider the time span from January 1985 to November 2010. Most series were obtained from *Datastream International*. Consumer Price Indices for the United States and Hong Kong were downloaded from the websites of the Federal Reserve Bank of St. Louis and the Hong Kong Census and Statistics Department, respectively. The source for the Japanese Urban Land Price Index (six Major Cities average) is the Statistics Bureau of Japan.

As Table 11 shows, there is strong evidence that house prices have run through explosive phases. For the US S&P/Case–Shiller home price index, the UK house price index, and the Spanish house price index, both tests reject the random-walk hypothesis at the 1% level, indicating bubble-like growth during the time preceding 2006/2007. Similar results are obtained for the Japanese land price index, where the explosive phase occurred before 1990. Interestingly, also for the Nikkei225 index, the supDFC and the supDF tests detect explosive behavior prior to January 1990. These results indicate that a land price bubble in Japan was paralleled by a stock market bubble.

With regard to the S&P 500, both the supDFC and supDF tests detect explosive behavior. The supDFC test rejects at the 5% or 1% level, while the supDF test rejects at the 10% level. Thus, there is evidence that in the 90s, not only the Nasdaq but also the S&P 500 was driven by a bubble.

The supDF test detects a bubble in the FTSE 100 series for weekly and daily data. The explosive behavior, however, is not identified during the 90s but several month before Black Monday 1987. The DF_{τ} statistic exceeds the critical values only when τ corresponds to July 1987.

Table 11 Testing for an explosive root

Series name	Monthly data		Weekly data		Daily data	
	supDFC	supDF	supDFC	supDF	supDFC	supDF
Stock market indices						
S&P 500 (January 1980 to March 2000)	***	*	**	*	***	*
FTSE 100 (December 1985 to December 1999)	*			*	*	**
Nikkei 225 (January 1957 to January 1990)	***	**	***	**	***	***
Hang Seng (October 1980 to October 2007)	**		**		**	
Shanghai (January 1991 to November 2007)	**		***		***	
Commodities						
Crude Oil (January 1985 to July 2008)						
Gold (January 1985 to November 2010)	*		*		*	
House/land prices						
United States (January 1987 to March 2006)	***	***	—	—	—	—
Spain (1987Q1–2007Q1)	***	***	—	—	—	—
United Kingdom (January 1991 to October 2007)	***	***	—	—	—	—
Japan (1957–1990)	***	**	—	—	—	—

The table reports significance levels of the supDFC and supDF test applied to the logarithm of the respective time series at monthly, weekly, and daily frequencies. For the house/land prices, data were not available at a weekly or daily frequency. For Spanish house prices, the data frequency are quarterly, and for Japanese urban land prices, data were available only at an annual frequency.
"*(**/***)" signifies significance at the 10% (5%/1%) level.
"—" signifies that the data are not available at the corresponding frequency.

Turning to Chinese stock market indices, the results in Table 11 show that for the Shanghai Stock Exchange Index, the supDFC test finds clear evidence of explosive growth before November 2007, and for the Hang Seng Index, the supDFC test rejects the no bubble hypothesis at the 5% level (independent of the sampling frequency).

Table 12 Estimates for the date of change form I(1) to explosive

Series name	Data frequency		
	Monthly	Weekly	Daily
Stock market indices			
S&P 500	1990-10-31	1990-10-16	1990-10-11
Nikkei 225	1982-10-03	1982-10-04	1982-10-01
Hang Seng	2003-03-31	2003-04-25	2003-04-25
Shanghai	2005-12-02	2005-12-05	2005-12-05
House/land Prices			
United States	1999-15-06	—	—
Spain	1997Q4	—	—
United Kingdom	1999-01-15	—	—
Japan	1985	—	—

The table reports break date estimates for different financial time series using the estimator $\hat{\tau}_{\mathrm{DFC}}$ from Section 2.7.

Furthermore, there is no evidence for a bubble in the barrel price of Brent Crude oil. For the price of a troy ounce of gold at the London Bullion Market, the evidence is mixed. The supDFC test is significant at the 10% level for monthly and weekly data, while the supDF test is insignificant for all sampling frequencies.[10] Test results are very different, when the real gold price from January 1968 until its all time high in January 1980 is considered. The supDFC and the supDF tests reject the hypothesis of no structural breaks at the 1% level, irrespective of the data frequency.

Finally, the dividend series for the stock market indices were not available in all cases. However, for the S&P 500 and for the Hang Seng, we were not able to reject the constant I(1) null hypothesis for the dividend series (not shown), which gives further support to the view, that there has been a bubble in those stock markets.

For those time series where the supDFC detected a change from I(1) to explosive at the 5% level, we use the related estimator $\hat{\tau}_{\mathrm{DFC}}$ to estimate the date of the change. The results are reported in Table 12. Figures 5 and 6 show the plots of the corresponding series. The shaded regions highlight the phase from the estimated start date of the bubble until its presumed collapse. The estimate for the Spanish house price index and for the Japanese house price index should be interpreted with care since the time series considered are rather short (81 and 34 observations, respectively). From Table 12, one can also see that the break date estimate for a

[10] Interestingly, when we estimate the start of a supposed bubble using $\hat{\tau}_{\mathrm{DFC}}$, the result, July 2007, closely coincides with the beginning of the subprime crisis and the failure of several Bear Stearns hedge funds.

given time series is robust to the choice of the sampling frequency. Also, note from the results in Table 11 that in most cases, changing the observation frequency has only a minor impact on the p values of the tests. This seems to mirror the finding of Shiller and Perron (1985) for unit root tests against stationary alternatives. Their theoretical analysis and Monte Carlo simulations suggest that power depends more on the span of the data rather than on the number of observations.

6 CONCLUSION

In this paper, the ability of several tests to detect rational bubbles has been investigated. The main focus was on rational bubbles in stock markets. The basic idea is that a rational bubble gives rise to an explosive component in stock prices. Therefore, a change from a random walk to an explosive regime is considered to be an indication for the emergence of a speculative bubble. Hence, all tests aim at detecting a switch from a random walk to an explosive regime. The sequential DF test, proposed by Phillips, Wu, and Yu (2011) serves as a reference point. We have also adapted various tests for a change in persistence to accommodate a possible change from I(1) to an explosive regime. In a simple simulation framework, it was shown that a Chow-type DF (supDFC) test and our modified version of Busetti and Taylor's (2004) (supBT) test have higher finite sample power than the test of Phillips, Wu, and Yu (2011), especially when the change from I(1) to explosive occurs late in the sample. This result was confirmed in simulation experiments with randomly starting bubbles. Moreover, the estimator that results from maximizing the Chow-type test statistic ($\hat{\tau}_{DFC}$) yields a reliable and roughly unbiased estimator of the starting date.

Since Evans (1991) clearly demonstrated the potential problems of unit root tests to detect periodically collapsing bubbles, we also study the properties of the tests under multiple structural breaks. It turns out that the Phillips, Wu, and Yu (2011) test is much more robust against multiple breaks than all other tests. We also considered sequential DF t-statistics and CUSUM statistics within a real-time monitoring framework and present critical values for different sizes of the monitoring sample.

The test of Phillips, Wu, and Yu (2011) provides strong evidence for a bubble in the Nasdaq index at the end of the 1990s. Their findings are even strengthened by our sequential Chow-type DF test and the modified version of Busetti and Taylor's (2004) test. The estimation of the starting date of the bubble indicates that the explosive regime emerged at the end of the year 1992. Moreover, our results support the view that bubbles have occurred in several other stock markets. Our sequential Chow-type DF test and the supDF test of Phillips, Wu, and Yu (2011) also indicate explosive behavior in US, UK, and Spanish house price indices prior to the so-called subprime crisis.

Received February 18, 2010; revised July 5, 2011; accepted September 12, 2011.

REFERENCES

Andrews, D. W. K. 1993. Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* 61 (4): 821–856.

Bai, J., and P. Perron. 1998. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* 66 (1): 47–78.

Banerjee, A., R. L. Lumsdaine, and J. H. Stock. 1992. Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence. *Journal of Business & Economic Statistics* 10 (3): 271–287.

Bhargava, A. 1986. On the Theory of Testing for Unit Roots in Observed Time Series. *Review of Economic Studies* 53: 369–384.

Blanchard, O. J., and M. W. Watson. 1982. "Bubbles, Rational Expectations, and Financial Markets." In P. Wachtel (ed.), *Crisis in the Economic and Financial Structure*, 295–315. Lexington: Lexington Books.

Busetti, F., and A. M. R. Taylor. 2004. Tests of Stationarity Against a Change in Persistence. *Journal of Econometrics* 123: 33–66.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay. 1997. *The Econometrics of Financial Markets*. Princeton: Princeton University Press.

Chu, C.-S. J., M. Stinchcombe, and H. White. 1996. Monitoring Structural Change. *Econometrica* 64: 1045–1065.

Cuñado, J., L. A. Gil-Alana, and F. P. de Gracia. 2005. A Test for Rational Bubbles in the NASDAQ Stock Index: A Fractionally Integrated Approach. *Journal of Banking and Finance* 29: 2633–2654.

Diba, B. T., and H. I. Grossman. 1988. Explosive Rational Bubbles in Stock Prices? *American Economic Review* 78(3): 520–530.

Evans, G. W. 1991. Pitfalls in Testing for Explosive Bubbles in Asset Prices. *American Economic Review* 81(4): 922–930.

Frömmel, M., and R. Kruse. 2011. Testing for a Rational Bubble Under Long Memory. *Quantitative Finance* (in press).

Froot, K., and M. Obstfeld. 1991. Intrinsic Bubbles: The Case of Stock Prices. *American Economic Review* 81: 1189–1214.

Galbraith, J. K. 1993. *A Short History of Financial Euphoria*. New York: Viking.

Gürkaynak, R. S. 2008. Econometric Tests of Asset Price Bubbles: Taking Stock. *Journal of Economic Surveys* 22 (1): 166–186.

Kim, J.-Y. 2000. Detection of Change in Persistence of a Linear Time Series. Journal of Econometrics 95: 97–116.

Kim, J.-Y., J. Belaire-Franch, and R. B. Amador. 2002. Corrigendum to 'Detection of Change in Persinstence of a Linear Time Series'. *Journal of Econometrics* 109: 389–392.

LeRoy, S. F., and R. D. Porter. 1981. The Present-Value Relation: Tests Based on Implied Variance Bounds. *Econometrica* 49: 555–574.

Leybourne, S., T.-H. Kim, V. Smith, and P. Newbold. 2003. Tests for a Change in Persistence Against the Null of Difference-Stationarity. *The Econometrics Journal* 6 (2): 291–311.

Nabeya, S. 1999. Asymptotic Moments of Some Unit Root Test Statistics in the Null Case. *Econometric Theory* 15: 139–149.

Phillips, P. C. B., Y. Wu, and J. Yu. 2011. Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values? *International Economic Review* 52 (1): 201–226.

Shiller, R. J. 1981. Do Stock Prices Move to Much to be Justified by Subsequent Changes in Dividends? *American Economic Review* 71: 421–436.

Shiller, R. J., and P. Perron. 1985. Testing the Random Walk Hypothesis: Power Versus Frequency of Observation. *Economics Letters* 18 (4): 381–386.

Taylor, A. M. R. 2005. Fluctuation Tests for a Change in Persistence. *Oxford Bulletin of Economics and Statistics* 67 (2): 207–230.

Taylor, R., and S. Leybourne. 2004. Some New Tests for a Change in Persistence. Economics Bulletin 3 (39): 1–10.

West, K. 1987. A Specification Test for Speculative Bubbles. *Quarterly Journal of Economics* 102: 553–580.

Zeileis, A., F. Leisch, C. Kleiber, and K. Hornik. 2005. Monitoring Structural Change in Dynamic Econometric Models. *Journal of Applied Econometrics* 20: 99–121.