CQE Center for Quantitative Economics

Sup-ADF-style bubble-detection methods under test

Verena Monschang † und Bernd Wilfling †

78/2019

† Department of Economics, University of Münster, Germany

wissen.leben
WWU Münster

Sup-ADF-style bubble-detection methods under test

Verena Monschang, Bernd Wilfling*

Westfälische Wilhelms-Universität Münster, Department of Economics (CQE), Am Stadtgraben 9, 48143 Münster, Germany

(Date of this version: February 8, 2019)

Abstract

In this paper we analyze the performance of supremum augmented Dickey-Fuller (SADF), generalized SADF (GSADF), and backward SADF (BSADF) tests, as introduced by Phillips et al. (International Economic Review 56:1043-1078, 2015) for detecting and date-stamping financial bubbles. In Monte Carlo simulations, we show that the SADF and GSADF tests may reveal substantial size distortions under typical financial-market characteristics (like the empirically well-documented leverage effect). We consider the rational bubble specification suggested by Rotermann and Wilfling (Applied Economics Letters 25:1091-1096, 2018) that is able to generate realistic stock-price dynamics (in terms of level trajectories and volatility paths). Simulating stock-price trajectories that contain these parametric bubbles, we demonstrate that the SADF and GSADF tests can have extremely low power under a wide range of bubble-parameter constellations. In an empirical analysis, we use NASDAQ data covering a time-span of 45 years and find that the outcomes of the bubble date-stamping procedure (based on the BSADF test) are sensitive to the data-frequency chosen by the econometrician.

Keywords: Stock markets, present-value model, rational bubble, explosiveness, SADF and GSADF tests, bubble detection, date-stamping.

JEL classification: C15, C32, C58, G15.

*Corresponding author. Tel.: +49 251 83 25040; fax: +49 251 83 25042.
E-mail addresses: verena.monschang@wiwi.uni-muenster.de (V. Monschang), bernd.wilfling@wiwi.uni-muenster.de (B. Wilfling).

1 Introduction

Recently, Phillips et al. (2011) and Phillips et al. (2014, 2015a, 2015b) have established a sound theoretical foundation of recursive right-tailed unit-root testing for explosiveness in time-series data. As a result, the most prominent of their testing procedures—the sup augmented Dickey-Fuller (SADF) test and its generalized variant (the GSADF test)—have been applied in a plethora of empirical articles, in which data explosiveness is interpreted as indicating an asset-price bubble. Along this line of argument, the studies aim at detecting speculative bubbles in alternative types of financial markets, for example in stock markets (Homm and Breitung 2012), commodity markets (Long et al. 2016), as well as in housing (Pan 2017) and currency markets (Bettendorf and Chen 2013; Hu and Oxley 2017). In the meantime, the popularity of the SADF and GSADF tests has been enhanced further by the fact that both testing procedures have become standard routines in econometric software packages like EViews (Caspi 2017).

In simulation experiments, Phillips, Shi, and Yu (2015a, PSY hereafter) demonstrate that SADF and GSADF tests have high discriminatory power when artificial stock-price data are generated under periodically collapsing Evans bubbles (Evans 1991). While the rational Evans bubble has become a benchmark specification in the theoretical and empirical literature, Rotermann and Wilfling (2014, 2018) elaborate two theoretical properties of the Evans bubble that appear irreconcilable with real-world stock-price dynamics. (i) The Evans bubble always collapses completely within one trading unit, implying that stock-price volatility also collapses abruptly within one period. (ii) After a crash, the Evans bubble necessarily reverts to the same expected value, a phenomenon for which there is no theoretical justification. By contrast, Rotermann and Wilfling (2018) propose an alternative rational bubble specification—in the form of a lognormal-mixture process—which is able to generate more realistic, stochastically deflating bubble trajectories.

In this article, we extend and modify the PSY power analysis of the SADF and GSADF tests in various ways. Our investigation has three major findings. (i) In their article, PSY specify a unit-root model with standard (symmetric) GARCH(1,1) errors as the data-generating process and find that this type of conditional heteroscedasticity does not impact substantially on the size of both tests. However, a well-established finding is that specific volatility asymmetries are crucial to modeling equity markets. The most prominent example is the leverage effect, stating that negative shocks often have a relatively larger impact on volatility than positive shocks. To capture such asymmetry, we study the effects of a threshold GARCH (TGARCH) volatility structure on the empirical size of the SADF and GSADF tests and find that both reveal considerable size distortions under TGARCH errors.

(ii) We generate artificial stock-price data under the abovementioned rational Rotermann-Wilfling bubble specification. Our simulations show that both tests often possess low empirical power under this realistic bubble specification. And (iii), we address the PSY date-stamping strategy designed for real-time bubble monitoring. We apply the procedure to NASDAQ data and compare the data-stamping results when using monthly versus daily observations. Our results indicate that the dating strategy is sensitive to the practitioner's choice of data frequency.

The remainder of the paper is organized as follows. Section 2 briefly reviews the essentials of present-value stock-price model and the rational Rotermann-Wilfling bubble specification. Section 3 recapitulates the SADF and GSDAF tests, which we use in Section 4 in our Monte Carlo simulations and the analysis of the date-stamping procedure. Section 5 concludes.

2 Present-value model, rational bubbles, and explosiveness

PSY motivate their SADF and GSADF testing procedures on the basis of the well-known present-value stock-price model with constant expected returns (Campbell et al. 1997). Within this framework, the date-t stock-price P_t is given by the Euler equation

$$P_t = \frac{1}{1+r}\left[\mathbb{E}_t(P_{t+1}) + \mathbb{E}_t(D_{t+1})\right], \tag{1}$$

where $\mathbb{E}_t(\cdot)$ denotes the conditional expectation operator and D_{t+1} the dividend payment between t and $t+1$. $r > 0$ is the constant discount factor, often referred to as the *required rate of return*, which is just sufficient to compensate investors for the riskiness of the stock.[1]

The first-order expectational difference Eq. (1) can be solved routinely by repeatedly substituting future prices forward. The entire class of solutions to Eq. (1) is given by

$$P_t = P_t^f + B_t = \sum_{i=1}^{\infty}\left(\frac{1}{1+r}\right)^i \cdot \mathbb{E}_t(D_{t+i}) + B_t, \tag{2}$$

where B_t is any stochastic process satisfying the (discounted) martingale property

$$\mathbb{E}_t(B_{t+1}) = (1+r)\cdot B_t. \tag{3}$$

The quantities $P_t^f = P_t - B_t$ and B_t in Eqs. (2) and (3) are called 'fundamental stock-price' and 'rational bubble', respectively.

Besides the constituting martingale property (3), any rational (stock-price) bubble should satisfy two additional theoretical properties, as pointed out by Diba and Grossman (1988a, b). (i) Rational bubbles cannot start from zero, and (ii) negative bubbles are ruled out as $t \to \infty$. The most frequently applied rational, parametric specification

[1] PSY enrich the model with the process $\{U_t\}$, representing unobservable fundamentals. Since the authors assume $\{U_t\}$ to be either $I(0)$ or an $I(1)$ process, we ignore $\{U_t\}$ without loss of generality.

satifying all these properties is the Evans (1991) bubble. However, the Evans bubble reveals a major empirical shortcoming in that it always bursts entirely from one trading unit to the next. These abrupt bursts not only entail unrealistic stock-price trajectories, but also incompatible volatility dynamics (Rotermann and Wilfling 2014). In Section 4, we therefore consider the bubble specification suggested by Rotermann and Wilfling (2018). This bubble model—a mixture of two lognormal processes—(i) generates realistic trajectories and stock-price volatility-paths, and (ii) satisfies the rationality condition (3) plus the two Diba-Grossman conditions mentioned above.

Specifically, the rational bubble has the form

$$B_t = \begin{cases} \frac{\alpha}{\psi\pi} B_{t-1} u_t & \text{, with probability } \pi \\ \frac{1-\alpha}{\psi(1-\pi)} B_{t-1} u_t & \text{, with probability } 1-\pi \end{cases}, \tag{4}$$

where $\psi \equiv (1+r)^{-1}, \alpha \in (0,1)$ and $\pi \in (0,1]$ such that $\frac{\alpha}{\pi} > 1$ and $\frac{1-\alpha}{\psi(1-\pi)} < 1$. $\{u_t\}_{t=1}^{\infty}$ is assumed to be an i.i.d. lognormal process with $u_t \sim \text{LN}(\frac{-\iota^2}{2}, \iota^2)$.[2] The constraint $\alpha \in (0,1)$ ensures that the bubble never collapses to zero and can thus re-inflate. The constraints $\frac{\alpha}{\pi} > 1$ and $\frac{1-\alpha}{1-\pi} < \psi$ allow us to interpret the two states of the bubble as follows. In State 1 (occurring with probability π) the bubble grows at the (mean) growth rate $\frac{\alpha}{\psi\pi} - 1 = \frac{\alpha}{\pi} - 1 + \frac{\alpha}{\pi} \cdot r > r$, i.e. at a faster rate than the required rate of return. In State 2 (occurring with probability $1-\pi$,) the bubble deflates at the negative (mean) growth rate $\frac{1-\alpha}{\psi(1-\pi)} - 1 < 0$.

Figure 1 about here

Depending on the specific parameter constellation, the bubble specification (4) displays a periodically recurring behavior with (stochastically evolving) deflation periods that can range between a "small/moderate correction" and a "big crash" within one

[2]The lognormal distribution is parameterized in terms of the single parameter ι, so that $u_t > 0$ and $\mathbb{E}_{t-1}(u_t) = 1$ for all t.

or arbitrarily many periods. As an illustrative example, Figure 1 displays 4 simulated bubble trajectories, with each trajectory (of length $T = 250$) having the common values $B_0 = 0.5, \psi = 0.9840$, and the trajectory-specific parameters ι^2, π, α as shown in Figure 1.

For our stock-price simulations in Section 4, we need to specify the fundamental stock-price process $\{P_t^f\}$. To this end, we adopt the frequently encountered assumption that dividends follow a random walk with drift,

$$D_t = \mu + D_{t-1} + e_t, \tag{5}$$

where $\{e_t\}_{t=1}^{\infty}$ is an i.i.d. Gaussian white-noise process with mean 0 and variance σ_e^2 (see, *inter alia*, Homm and Breitung 2012). Taking conditional expectations and adding them as in Eq. (2), we obtain the fundamental stock price as

$$P_t^f = \frac{1+r}{r^2} \cdot \mu + \frac{1}{r} \cdot D_t, \tag{6}$$

which, after inserting Eq. (5) into (6) and rearranging the terms, yields

$$P_t^f = \frac{\mu}{r} + P_{t-1}^f + \frac{e_t}{r} = \mu' + P_{t-1}^f + e_t', \tag{7}$$

showing that the fundamental stock price P_t^f also follows a random walk with drift.

At this stage, some comments are in order on the interrelation between the concepts 'explosiveness in asset prices' and 'existence of a bubble'. In our framework—consisting of Eqs. (1) – (7)—the fundamental stock price P_t^f constitutes a (nonexplosive) $I(1)$ process. Therefore, in view of Eqs. (2) and (3), if we find empirical evidence of explosive behavior in the stock-price process (2), we can attribute this explosiveness to the rational bubble. This profound conclusion, however, hinges crucially on the specific assumptions of our framework and is far from being generally valid. To illustrate, let us consider two alternative model setups. (i) A situation, in which the fundamental stock

price (for whatever economic reason) follows an explosive process. (ii) An extended rational valuation framework with stochastic discount factors (instead of our constant required rate of return r).[3] It is straightforward to verify that, under both scenarios, explosiveness in stock prices (or in the price-dividend ratio) neither constitutes a necessary nor a sufficient condition to deduce the existence of a bubble.

3 Sup-ADF-style tests and bubble date-stamping

The SADF and GSADF tests for explosiveness (applied to the time-series $\{y_t\}_{t=0}^{T}$) rest on well-defined sequences of t-statistics (ADF-statistics) of the parameter θ, estimated from the empirical specification

$$y_t = c + \theta y_{t-1} + \sum_{i=1}^{k} \lambda_i \Delta y_{t-i} + \varepsilon_t, \tag{8}$$

where k is the transient lag-order, Δ denotes the first-difference operator, and $\varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$. The objective is to test the unit-root null hypothesis $H_0 : \theta = 1$ versus the right-tailed alternative of explosiveness, $H_1 : \theta > 1$. For characterizing the respective sequences of ADF-statistics, which are needed to formally represent the ultimate SADF and GSADF test statistics, we consider subsamples over the time domain $\{0, 1, \ldots, T\}$ as fractions of the original sample. For this purpose, let the fractions (i) r_0, (ii) r_1, and (iii) r_2 respectively denote (i) the (fractional) width of the smallest subsample (used to initialize the computation of the test statistic), (ii) the (fractional) starting point of a subsample, and (iii) the (fractional) endpoint of a subsample.

Using this notation, Phillips et al. (2011) define the SADF test statistic as the sup ADF-statistic from repeated estimation of the empirical regression (8) on a forward expanding sample sequence. Concretely, the authors consider as given, the minimal sample window width r_0, set the subsample starting point $r_1 = 0$, and let the subsample

[3] Cochrane (2011) presents overwhelming empirical evidence of time-varying discount rates.

endpoint r_2 range between r_0 and 1. Denoting the ADF-statistic for a subsample running from r_1 to r_2 by $\text{ADF}_{r_1}^{r_2}$, they define the SADF test statistic as

$$\text{SADF}(r_0) \equiv \sup_{r_2 \in [r_0, 1]} \left\{ \text{ADF}_{r_1=0}^{r_2} \right\}. \tag{9}$$

The GSADF test—suggested by Phillips et al. (2015a, b) with the goal of improving the detection capacity under multiple stock-price bubbles—essentially pursues the same idea as the SADF test, but processes more subsamples to estimate the ADF-regression (8). In contrast to the SADF variant, the GSADF test allows the fractional starting point r_1 to range between 0 and $r_2 - r_0$, implying a double recursive subsample structure. The corresponding test statistic is defined as

$$\text{GSADF}(r_0) \equiv \sup_{\substack{r_2 \in [r_0, 1] \\ r_1 \in [0, r_2 - r_0]}} \left\{ \text{ADF}_{r_1}^{r_2} \right\}. \tag{10}$$

Phillips et al. (2014, 2015a) derive the asymptotic null distributions of the SADF and GSADF test statistics on the basis of the prototypical model with weak (local to zero) intercept form,

$$y_t = a \cdot T^{-\eta} + \theta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2), \tag{11}$$

with constants a and $\eta > 1/2$. Under the null hypothesis of a unit root ($\theta = 1$), the limiting distributions of the test statistics are given by

$$\text{SADF}(r_0) \xrightarrow{d} \sup_{r_2 \in [r_0, 1]} \left\{ \frac{\frac{1}{2} r_2 \left[W(r_2)^2 - r_2 \right] - \int_0^{r_2} W(s) ds \; W(r_2)}{{r_2}^{\frac{1}{2}} \left\{ r_2 \int_0^{r_2} W(s)^2 ds - \left[\int_0^{r_2} W(s) ds \right]^2 \right\}^{\frac{1}{2}}} \right\}, \tag{12}$$

and

$$\text{GSADF}(r_0) \xrightarrow{d}$$

$$\sup_{\substack{r_2\in[r_0,1]\\ r_1\in[0,r_2-r_0]}} \left\{ \frac{\frac{1}{2}(r_2-r_1)\left[W(r_2)^2-W(r_1)^2-(r_2-r_1)\right]-\int_{r_1}^{r_2}W(s)ds\,\left[W(r_2)-W(r_1)\right]}{(r_2-r_1)^{\frac{1}{2}}\left[(r_2-r_1)\int_{r_1}^{r_2}W(s)^2ds-\left(\int_{r_1}^{r_2}W(s)ds\right)^2\right]^{\frac{1}{2}}} \right\}, \tag{13}$$

where W denotes a standard Wiener process.

In addition to the SADF and GSADF testing for bubbles, PSY propose a date-stamping procedure for estimating the (fractional) origination and termination dates (denoted by r_e and r_f) of a bubble. The underlying idea rests on a recursive test procedure called the backward SADF (BSADF) test. The BSADF test follows the same principle as the SADF test, but processes the sample in the reverse direction. The test proceeds in two steps. (i) It computes a sequence of ADF statistics using a series of samples, in which each individual sample has the same fixed endpoint r_2, while the starting point ranges between 0 and $r_2 - r_0$. (ii) The BSADF test statistics is then defined as the sup-value of the ADF statistics computed in Step (i):

$$\mathrm{BSADF}_{r_2}(r_0) \equiv \sup_{r_1\in[0,r_2-r_0]} \left\{\mathrm{ADF}_{r_1}^{r_2}\right\}. \tag{14}$$

In order to test for explosiveness in the time-series process at date $t = \lfloor Tr_2 \rfloor$ ($\lfloor\cdot\rfloor$ is the floor function), the $\mathrm{BSADF}_{r_2}(r_0)$ statistic is compared to a critical value obtained from Monte Carlo simulation. Letting r_2 range between r_0 and 1, PSY (i) define the origination date $\lfloor Tr_e \rfloor$ of a bubble as that point in time with the first chronological observation, at which the BSADF statistic exceeds the critical value, and (ii) suggest estimating the origination date $\lfloor Tr_e \rfloor$ via

$$\hat{r}_e = \inf_{r_2\in[r_0,1]} \left\{r_2 : \mathrm{BSADF}_{r_2}(r_0) > \mathrm{scv}_{r_2}^{\beta_T}\right\}, \tag{15}$$

where β_T denotes the significance level and $\mathrm{scv}_{r_2}^{\beta_T}$ is the $100(1-\beta_T)\%$ critical value of

the SADF test statistic based on $\lfloor Tr_2 \rfloor$ observations.[4]

In a similar vein, the termination date $\lfloor Tr_f \rfloor$ of the bubble is defined as the point in time with the first chronological observation, at which the BSADF statistic falls below the critical value. Additionally, assuming a minimal time lag of $\delta \log(T)$ observations to exist between the origination and the termination date of the bubble, PSY propose estimating the termination date $\lfloor Tr_f \rfloor$ via

$$\hat{r}_f = \inf_{r_2 \in [\hat{r}_e + \delta \log(T)/T, 1]} \{r_2 : \text{BSADF}_{r_2}(r_0) < \text{scv}_{r_2}^{\beta_T}\}. \tag{16}$$

We end this section by noting that (i) the parameter δ in Eq. (16) controls for the minimal duration of the bubble, and (ii) the GSADF and the BSADF test statistics are interrelated by

$$\text{GSADF}(r_0) = \sup_{r_2 \in [r_0, 1]} \{\text{BSADF}_{r_2}(r_0)\}.$$

4 Monte Carlo study and empirical analysis

In this section, we first approximate asymptotic critical values of the SADF and GSADF tests, and analyze the empirical size and power of both tests. Second, we report the results of a date-stamping analysis with daily and monthly NASDAQ observations.

4.1 Simulation results

Given Eqs. (9), (10), (12) and (13), the SADF and GSDF statistics, as well as their asymptotic distributions depend on the minimal window size r_0. To be in line with the PSY article, we adopt the rule $r_0 = 0.01 + 1.8/\sqrt{T}$, and—whenever considering finite samples—focus on the specific sample sizes $T \in \{100, 200, 400, 800, 1600\}$.

In a first step, we approximate the asymptotic critical values of the SADF and

[4] PSY point out that the significance level β_T may depend on the sample size T and shrink to zero as $T \to \infty$.

GSADF statistics via Monte Carlo simulation. In contrast to PSY, we do not make use of the asymptotic null distributions from Eqs. (12) and (13), the simulation of which requires an approximation of the Wiener process. Instead, we simulate the critical values by restricting the data-generating process to the prototypical specification in Eq. (11) with parameters $\theta = 1$ and $a = \eta = 1$ and use the sample size $T = 5000$. (The parameter values for a and η are taken from PSY.) With these settings, we approximate the asymptotic critical values by simulating 100,000 and 12,500 replications of the SADF and GSADF statistics, respectively.

Table 1 about here

Table 1 reports our asymptotic critical values, which share two features with their analogs from the PSY simulations via the asymptotic null distributions from Eqs. (12) and (13). (i) The critical values of both test statistics increase with a decreasing minimal window size r_0. (ii) The GSDAF critical values always exceed their SADF counterparts. In most cases, our asymptotic critical values are slightly larger than those reported by PSY, yielding more conservative rejections of the null hypothesis. In our analysis below, we rely on our critical values from Table 1.

Table 2 about here

In order to check the validity of our asymptotic critical values, when applied to finite samples, we simulate sizes for both tests. For this purpose, we generate data from the prototypical process in Eq. (11) under the null hypothesis (with parameters $a = \eta = \theta = 1$) for the finite sample sizes $T \in \{100, 200, 400, 800, 1600\}$. On the basis of 10,000 replications, we compute simulated sizes as the fractions of replications, for which the tests erroneously reject the null of a unit root in favor of the alternative, thus indicating explosiveness. We report the results in Table 2 for a nominal size of

5% (i.e. using the 95% critical values from Table 1) and do not find substantial size distortions.

The process specification in Eq. (11) ignores (conditional) heteroscedasticity, a well-documented phenomenon in all types of financial markets. To assess the impact of heteroscedasticity on the SADF and GSADF procedures, PSY address the sizes of both tests under unit-root processes with GARCH errors. Under standard GARCH$(1,1)$ errors—as originally defined in Bollerslev (1986)—they do not find critical size distortions.

However, the standard GARCH$(1,1)$ specification does not account for volatility asymmetries, such as the highly relevant leverage effect, according to which negative stock-market shocks tend to exert a larger impact on volatility than positive shocks (e.g. Black 1976; Christie 1982; Schwert 1989). Therefore, we modify the PSY size analysis and consider (as the data-generating process) the unit-root specification from Eq. (11) with $a = \eta = \theta = 1$ under (threshold) TGARCH$(1,1)$ errors,

$$\varepsilon_t = s_t\sqrt{h_t}, \qquad s_t \overset{\text{i.i.d.}}{\sim} N(0,1), \tag{17}$$

$$h_t = \omega + \gamma \cdot \varepsilon_{t-1}^2 + \beta \cdot h_{t-1} + \phi \cdot \varepsilon_{t-1}^2 \cdot \mathrm{I}(\varepsilon_{t-1} < 0), \tag{18}$$

where $\mathrm{I}(\cdot)$ denotes the indicator function, which takes on the value 1 if the market is shocked by bad news ($\varepsilon_{t-1} < 0$), and is 0 in the case of good news (Zakoïan 1994).

Table 3 about here

Table 3 displays the simulated sizes of both tests under TGARCH$(1,1)$ errors for the nominal size of 5% and the sample sizes $T \in \{100, 200, 400, 800, 1600\}$ on the basis of 10,000 replications. We set the TGARCH$(1,1)$ parameters from Eq. (18) to $\omega = 0.4387, \gamma = 0, \beta = 0.9319$, $\phi = 0.1306$, so as to coincide with the maximum likelihood estimates obtained from monthly observations of the NASDAQ price-dividend ratio

covering the (relatively) tranquil period between January 1988 and December 1994.[5] Apparently, both tests exhibit substantial size distortions under volatility asymmetry in the form of TGARCH$(1,1)$ errors.

Next, we address the empirical power properties of the SADF and GSADF tests.[6] In preparation, we simulate 10,000 stock-price series from the present-value equation $P_t = P_t^f + B_t$, with the fundamental stock price P_t^f evolving according to Eqs. (5) and (6), and the rational bubble B_t following the log-normal mixture from Eq. (4). The involved parameters are set equal to the estimates obtained in Rotermann and Wilfling (2018), who fit the bubble specification (4) to monthly NASDAQ observations between January 1990 and October 2013, applying a particle-filter technique. Explicitly, we use the parameter values $\mu = 0, \sigma_e^2 = 0.4476, D_0 = 1.6942, B_0 = 10.1925, \psi = 0.9840, \iota^2 = 0.0061, \pi = 0.9595$ and $\alpha = 0.9675$.

Table 4 about here

Table 4 reports the results of our power analysis for the sample sizes $T \in \{100, 200, 400, 800, 1600\}$. Again, we use the 95% critical values from Table 1, and 10,000 Monte Carlo replications. Our analysis has 4 major findings. (i) The power of the GSADF test exceeds the power of the SADF test—except for $T = 1600$, where both tests have power equal to 1. (ii) The power of both tests increases with increasing sample size. (iii) The power of both tests is extremely low for $T = 100, 200$. It improves only slightly for $T = 400$. (iv) Both tests perform acceptably for $T = 800$. For $T = 1600$, both tests identify the bubble in each of the 10,000 simulated stock-price series.

Table 5 about here

[5] We describe the data set in Section 4.2.

[6] We note that adjusting both tests for the size distortions under TGARCH errors would entail a decrease in the empirical power.

Table 5 reports the power of both tests, when some model parameters are singly (or, as π and α, jointly) varied (see Table 5), while all other parameters are held constant at their levels presented in Table 4 (*ceteris-paribus* analysis). Explicitly, we let (i) the dividend drift μ range between 0 and 0.003, (ii) the discount factor ψ range between 0.975 and 0.990, and (iii) the probability π range between 0.35 and 0.95. In the case of the π-variation, we simultaneously adjust the parameter α, so that the mean bubble growth factor from Eq. (4), $\alpha/(\psi\pi)$, remains constant at 1.05.

Table 5 displays the results of our analysis, conducted with 10,000 Monte-Carlo replications (for each parameter setting), and the sample size $T = 400$ (which covers a time span of more than 33 years under monthly observations). We obtain the following 3 findings. (i) A variation in the dividend drift μ (*ceteris paribus*) does not substantially affect the power of either test. This result is not surprising, since both tests are based on Eq. (8), capturing the effects of the dividend drift. (ii) The power of both tests decreases dramatically with an increasing discount factor ψ. For instance, for $\psi = 0.99$, both tests only detect 9.73% and 12.35% of the simulated bubbles. An explanation may be that—with an increasing discount factor ψ—the bubble's positive (mean) growth factor from Eq. (4), $\alpha/(\psi\pi)$, decreases, rendering the detection of explosiveness more difficult. (iii) The power of both tests decreases substantially with an increasing probability π. Recall that π represents the likelihood of ongoing bubble growth at the constant rate of 5% (what we achieve by an appropriate adjustment of α). For $\pi = 0.95, \alpha = 0.0972$, both tests only detect 8.04 % and 9.72 % of the bubbles. *Prima facie*, this finding appears counter-intuitive, since we would expect the tests to exhibit higher power when the probability of bubble inflation (bubble growth) increases. One explanation of this phenomenon rests on the fact that the joint variation of the parameters π and α keeps the inflation rate, given by $\alpha/(\psi\pi) - 1$, stable at the 5 % level. However, at the same time this variation increases the mean bubble deflation rate in Eq. (4), given by

$(1-\alpha)/[\psi(1-\pi)]-1$, in absolute value, namely from -0.0111 to -0.9492. Evidently, neither test is capable of coping with these opposing effects on bubble dynamics.

Figure 2 about here

Figure 2 illustrates the overall results of our power analysis. For this empirical-quantile graph, we multiplied the 24 power values from Table 5 (12 SADF, 12 GSADF power values) by 10,000, so that the values now represent the number of bubbles correctly detected by both tests in each parameter setting.[7] Two findings are striking. (i) The bubble-detection capacity of the GSADF test only slightly outperforms that of the SADF test. (ii) In 10 out of 12 parameter settings, both tests detect less than 6,000 (out of 10,000) bubbles per setting. In other words, for these 10 (out of 12) settings, more than 40% of the existing bubbles remain undetected.[8] Only in 2 of 12 settings, do both tests identify more than 8,000 (out of 10,000) bubbles.

4.2 Date-stamping analysis

In this section, we apply the PSY date-stamping methodology [Eqs. (14)-(16)] to the NASDAQ stock market. We use *Thomson Reuters Datastream*, which provides daily and monthly observations of (i) the NASDAQ composite price index, (ii) the NASDAQ composite dividend yields, and (iii) the US Consumer Price Index (CPI) for all urban consumers. The data cover the time-span 2 January 1973 – 29 December 2017 ($T = 540$ monthly, $T = 11739$ daily observations).

We use the NASDAQ price index and the dividend yields to compute our dividend time series, and the CPI to deflate the nominal series. We obtain daily observations

[7] Recall that we use 10,000 replications in each parameter setting, where each replication contains a bubble.

[8] In 4 settings, more than 5,000 (out of 10,000), and in 2 settings even more than 8,000 (out of 10,000) bubbles remain undetected.

on the CPI by the linear interpolation

$$\mathrm{CPI}_{M,D} = \mathrm{CPI}_{M,15} + \left(\frac{D-15}{30}\right)\left(\mathrm{CPI}_{M+1,15} - \mathrm{CPI}_{M,15}\right),$$

where $\mathrm{CPI}_{M,D}$ is the consumer price index for day D in month M (which is assumed to have 30 days). We use the NASDAQ price index and the dividend time series to obtain the price-dividend ratio. In order to estimate the termination date, according to Eq. (16), the date-stamping procedure requires an assumption regarding the minimal (fractional) bubble duration $\delta \log(T)/T$. We impose a minimal duration of 6 months (180 days), implying the (approximate) values $\delta = 2.2$ (monthly observations) and $\delta = 44.2$ (daily observations).

Figure 3 about here

Our date-stamping analysis starts with the monthly NASDAQ sample (upper panel in Figure 3). As the training period, we use the first 47 (out of 540) observations. As described in Section 3, we investigate explosive behaviour in the price-dividend ratio via the intersections of the BSADF test statistics with its corresponding 95% critical values (obtained from 2,000 Monte Carlo replications). The upper panel in Figure 3 indicates 4 bubbly periods (grey shaded areas) lasting, respectively, from (i) May 1983 – June 1984 (14 months), (ii) December 1985 – November 1987 (24 months), (iii) July 1995 – April 2001 (68 months), and (iv) October 2008 and May 2009 (8 months).

The final 3 of these bubbly periods can obviously be ascribed to (i) the bull market prior to Black Monday in October 1987, (ii) the dotcom bubble with its crash starting at the beginning of 2000, and (iii) the short-term stock-market recovery after the Lehman Brothers insolvency in September 2008. Especially the latest bubble during the subprime mortgage crisis raises crucial doubts about the correct bubble date-stamping. The stamped bubble starts during the NASDAQ downturn after the Lehman collapse,

and ends in the middle of the recovery period.

In the lower panel of Figure 3, we analyze the date-stamping performance for the daily NASDAQ observations, where our training period consists of the first 312 (out of 11739) observations. Obviously, the sequence of the BSADF statistics appears far more volatile under daily observations. The date-stamping procedure identifies a number of short-lived periods of explosiveness that do not show up under monthly data. In the lower panel of Figure 3, we mark the centers of these periods by vertical, dashed lines. However, since we impose a minimal bubble duration of 180 days ($\delta = 44.2$), the procedure only identifies the first 3 bubbly periods (i.e. one less than under monthly data). The period during the subprime mortgage crisis, identified as bubbly under monthly observations, now remains unstamped.

Apart from that, 2 out of the 3 stamped bubbles under daily observations (the first and third) differ substantially in their origination and termination dates from their analogs under monthly data. The first bubble under daily data (lower panel) already starts on 23 December 1983, i.e. 5 months earlier than its analog in the upper panel. Since both termination dates basically coincide (June 1984 and 14 May 1984), this is tantamount to a 5-month-longer-lasting bubble under daily observations. Similarly, the dotcom bubble basically has the same origination date under both data-frequencies, but ends 5 months later under monthly observations (April 2001 vs [23] November 2000).

We end by noting that the econometrician's imposition of a minimal bubble duration (6 months / 180 days in our analysis) at the outset of the date-stamping procedure turns out to be crucial. In the lower panel of Figure 3, we observe 4 short-lived periods of explosiveness, that are not recognized as bubbles, due to our arbitrary setting of $\delta = 44.2$ for daily observations. The practitioner will frequently be confronted with the judgmental and subjective question of whether a (rather) short-lived sequence of

BSADF statistics exceeding the critical values either (i) is to be interpreted as a bubbly period, or (ii) is to be viewed as a statistical artefact.[9]

5 Conclusion

This paper investigates the performance of SADF and GSADF tests, as suggested by Phillips et al. (2015a, b), in detecting stock-market bubbles. In a first step, we address (i) the empirical size of both tests under typical financial-market volatility asymmetries (like the leverage effect), and (ii) the empirical power of the tests in detecting a class of rational bubbles, as proposed by Rotermann-Wilfling (2018). Our Monte Carlo simulations find that both tests exhibit substantial size distortions when the data-generating process is subject to (threshold) TGARCH errors, which are often used to capture leverage effects.[10] Moreover, the SADF and GSADF tests often have low empirical power in identifying the (flexible) Rotermann-Wilfling bubble. As shown in Figure 2, in 10 of our 12 scenarios (=83%), more than 40% of the existing bubbles remain undetected by the two tests. In a second step, we investigate the bubble date-stamping performance (based on the BSADF test) using monthly and daily NASDAQ data over a period of 45 years. We find that the date-stamping results may be sensitive to the data-frequency in two respects. (i) Some periods are stamped as bubbly under monthly data, but remain unstamped under daily observations. (ii) A bubbly period that is stamped under both data-frequencies, may have different origination and termination dates under the two frequencies (implying substantially differing durations).

We emphasize that it is by no means our intention to discredit any of the sup ADF-style testing procedures discussed in the previous sections. The probabilistic

[9]Fulop and Yu (2017) suggest estimating the termination date within a Bayesian framework by setting $\delta = 0$ in Eq. (16).

[10]Fulop and Yu (2017) propose a 2-regime Markov-switching bubble framework, which appears to exhibit some robustness in bubble detection, when the leverage effect is present in the data-generating process.

background of the methodologies, designed to test for (statistical) explosiveness, is mathematically rigorous and, from a purely econometric perspective, worth investigating further. However, when it comes to financial bubble detection, our advice is to apply the routines with caution. At the end of Section 2, we ascertain that in many realistic financial-market settings, (statistical) explosiveness neither constitutes a necessary nor a sufficient condition to deduce the existence of a bubble. Therefore, we cannot expect statistical tests for explosiveness to constitute a generally valid tool for bubble detection.

A further aspect concerns the theoretical setting (the present-value model), in which bubbles are often discussed. Within this framework, rational bubbles cannot start from zero (Diba and Grossman 1988a, b), implying that, at the present date t, either (i) a rational bubble exists (i.e. $B_t > 0$, although B_t may take on very small positive values), or (ii) there is currently no (rational) bubble (i.e. $B_t = 0$), but then there will never be one. Viewed from this angle, an appealing alternative to detecting bubbles via statistical tests for explosiveness could be to assume a parametric bubble specification (such as the one we use in this paper), and to estimate the parameters via appropriate techniques.

References

Bettendorf T, Chen W (2013) Are there bubbles in the Sterling-dollar exchange rate? New evidence from sequential ADF test. Economics Letters 120:350-353

Black F (1976) Studies on stock price volatility changes. Proceedings of the 1976 Meetings of the Business and Economics Statistics Section, pp. 177-181, American Statistical Association, Washington, DC

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31:307-328

Campbell JY, Lo AW, MacKinlay AC (1997) The Econometrics of Financial Markets. Princeton University Press

Caspi I (2017) Rtadf: Testing for bubbles with Eviews. Journal of Statistical Software 81:1-16

Christie AA (1982) The stochastic behavior of common stock variances: Value, leverage and interest rate effects. Journal of Financial Economics 10:407-432

Cochrane JH (2011) Presidential address: Discount rates. Journal of Finance 66:1047-1108

Diba B, Grossman H (1988a) Explosive rational bubbles in stock prices? American Economic Review 78:520-530

Diba B, Grossman H (1988b). The theory of rational bubbles in stock prices. Economic Journal 98:746-754

Evans G (1991) Pitfalls in testing for explosive bubbles in asset prices. American Economic Review 81:922-30

Fulop A, Yu J (2017) Bayesian analysis of bubbles in asset prices. Econometrics 4, Article 47:1-23

Homm U, Breitung J (2012) Testing for speculative bubbles in stock markets: A comparison of alternative methods. Journal of Financial Econometrics 10: 198-231

Hu Y, Oxley L (2017) Are there bubbles in exchange rates? Some new evidence from G10 and emerging market economies. Economic Modelling 64:419-442

Long W, Li D, Li Q (2016) Testing explosive behavior in the gold market. Empirical Economics 51:1151-1164

Pan WF (2017) Detecting bubbles in China's regional housing markets. Empirical Economics. https://doi.org/10.1007/s00181-017-1394-3

Phillips PCB, Shi S, Yu J (2014) Specification sensitivity in right-tailed unit root testing for explosive behaviour. Oxford Bulletin of Economics and Statistics 76:315-333

Phillips PCB, Shi S, Yu J (2015a) Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. International Economic Review 56:1043-1078

Phillips PCB, Shi S, Yu J (2015b) Testing for multiple bubbles: Limit theory of real-time detectors. International Economic Review 56:1043-1078

Phillips PCB, Wu Y, Yu J (2011) Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values? International Economic Review 52:201-226

Rotermann B, Wilfling B (2014) Periodically collapsing Evans bubbles and stock-price volatility. Economics Letters 123:383-386

Rotermann B, Wilfling B (2018) A new stock-price bubble with stochastically deflating trajectories. Applied Economics Letters 25:1091-1096

Schwert GW (1989) Why does stock market volatility change over time? Journal of Finance 44:1115-1153

Zakoïan JM (1994) Threshold heteroscedastic models. Journal of Economic Dynamics and Control 18:931-955

Tables and Figures

Table 1
Asymptotic critical values of SADF and GSADF tests

	SADF			GSADF		
	90%	95%	99%	90%	95%	99%
$r_0 = 0.190$	1.1136	1.4107	1.9754	1.7329	1.9920	2.4838
$r_0 = 0.137$	1.1848	1.4738	2.0339	1.8734	2.1017	2.5771
$r_0 = 0.100$	1.2222	1.5195	2.0936	1.9945	2.2046	2.6551
$r_0 = 0.074$	1.2887	1.5613	2.1081	2.0993	2.3121	2.7254
$r_0 = 0.055$	1.3272	1.5963	2.1330	2.2009	2.3995	2.8002

Note: Large-sample critical values are approximated by simulating the data-generating process from Eq. (11) with $\theta = 1$ and $a = \eta = 1$. The chosen sample size is $T = 5000$. The numbers of replications are 100,000 and 12,500 for the SADF and GSADF tests, respectively.

Table 2
Sizes of SADF and GSADF tests in finite samples when using asymptotic critical values

	$T = 100$ $r_0 = 0.190$	$T = 200$ $r_0 = 0.137$	$T = 400$ $r_0 = 0.100$	$T = 800$ $r_0 = 0.074$	$T = 1,600$ $r_0 = 0.055$
SADF	0.0393	0.0404	0.0410	0.0435	0.0478
GSADF	0.0483	0.0520	0.0489	0.0501	0.0499

Note: Sizes are obtained by simulating data (under the null hypothesis) from the specification in Eq. (11) with parameters $a = \eta = \theta = 1$. Using 10,000 replications, the sizes are computed via the (asymptotic) 95% critical values from Table 1.

Table 3
Sizes of the SADF and GSADF tests in finite samples under TGARCH$(1, 1)$ errors

	$T = 100$ $r_0 = 0.190$	$T = 200$ $r_0 = 0.137$	$T = 400$ $r_0 = 0.100$	$T = 800$ $r_0 = 0.074$	$T = 1,600$ $r_0 = 0.055$
SADF	0.2221	0.1999	0.1488	0.1020	0.0620
GSADF	0.1528	0.1537	0.1613	0.2018	0.2936

Note: Sizes are obtained by simulating data (under the null hypothesis) from the specification in Eq. (11) with parameters $a = \eta = \theta = 1$ under TGARCH errors, as described in Eqs. (17) and (18). The TGARCH parameters are $\omega = 0.4387, \gamma = 0, \beta = 0.9319, \phi = 0.1306$. Using 10,000 replications, the sizes are computed via the (asymptotic) 95% critical values from Table 1.

Table 4
Empirical power of SADF and GSADF tests under a rational bubble

	$T = 100$ $r_0 = 0.190$	$T = 200$ $r_0 = 0.137$	$T = 400$ $r_0 = 0.100$	$T = 800$ $r_0 = 0.074$	$T = 1,600$ $r_0 = 0.055$
SADF	0.0373	0.0857	0.5579	0.9758	1.0000
GSADF	0.0448	0.1075	0.5806	0.9782	1.0000

Note: The stock-price series are generated as $P_t = P_t^f + B_t$ with P_t^f from Eqs. (5) and (6), and B_t as in Eq. (4). The parameters are set to $\mu = 0, \sigma_e^2 = 0.4476, D_0 = 1.6942, B_0 = 10.1925, \psi = 0.9840, \iota^2 = 0.0061, \pi = 0.9595$ and $\alpha = 0.9675$. For the power calculations, we use the 95% critical values from Table 1, and 10,000 replications.

Table 5
Power of SADF and GSADF tests under varying bubble parameters

	$\mu = 0.000$	$\mu = 0.001$	$\mu = 0.002$	$\mu = 0.003$
SADF	0.5579	0.5603	0.5543	0.5527
GSADF	0.5806	0.5814	0.5774	0.5751
	$\psi = 0.975$	$\psi = 0.98$	$\psi = 0.985$	$\psi = 0.990$
SADF	0.9891	0.8646	0.4531	0.0973
GSADF	0.9890	0.8755	0.4852	0.1235
	$\pi = 0.35$ $\alpha = 0.3675$	$\pi = 0.55$ $\alpha = 0.5775$	$\pi = 0.75$ $\alpha = 0.7875$	$\pi = 0.95$ $\alpha = 0.9975$
SADF	0.5626	0.5108	0.4250	0.0804
GSADF	0.5813	0.5355	0.4555	0.0972

Note: The stock-price series are generated as $P_t = P_t^f + B_t$, with P_t^f as in Eqs. (5), (6), and B_t as in Eq. (4). As in Table 4, the basic set of parameters is $\mu = 0, \sigma_e^2 = 0.4476, D_0 = 1.6942, B_0 = 10.1925, \psi = 0.9840, \iota^2 = 0.0061, \pi = 0.9595, \alpha = 0.9675$. The parameters μ, ψ, π, α are then singly (or, as π and α, jointly) varied as indicated. The remaining parameters are held constant at their basic values. For the power calculations, we set $T = 400, r_0 = 0.1$ and use the 95% critical values from Table 1, with 10,000 replications for each parameter setting.

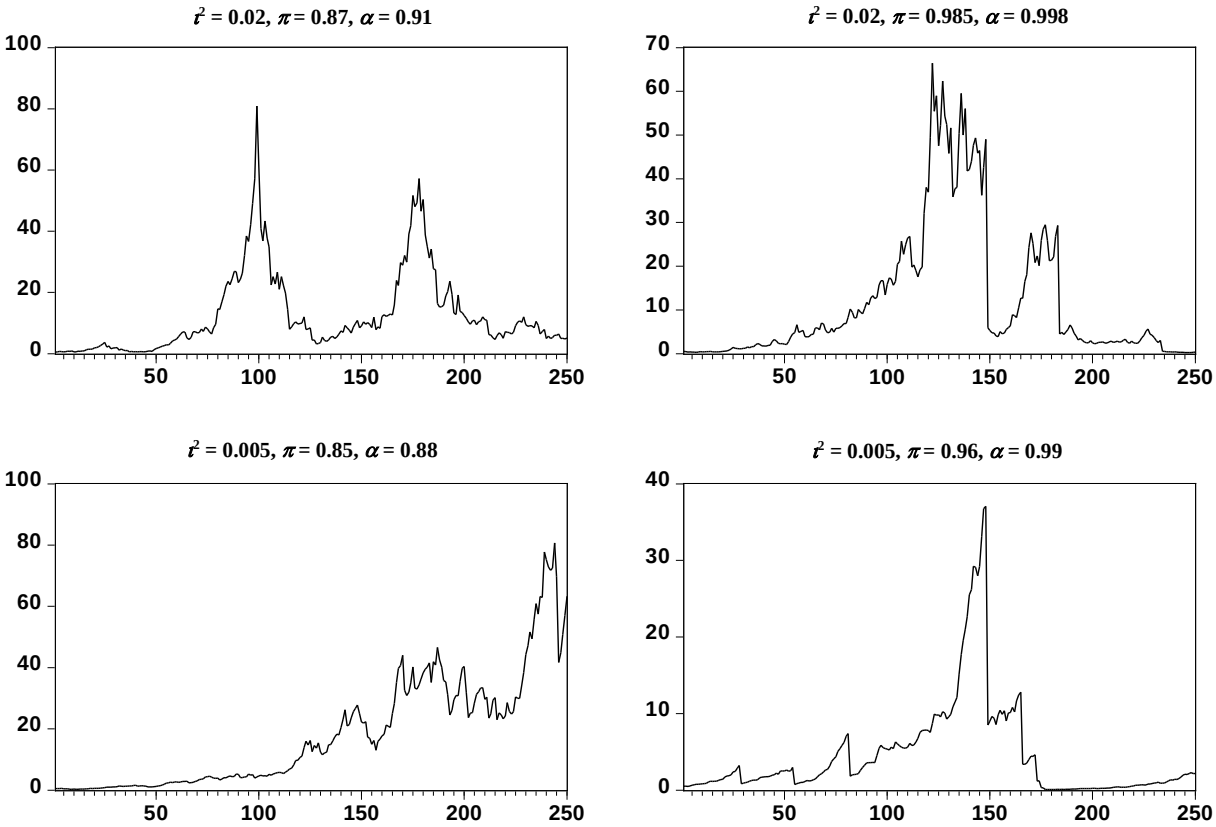

Figure 1. Bubble trajectories of length T = 250 simulated according to Eq. (4) with common parameters B_0 = 0.5, ψ = 0.9840

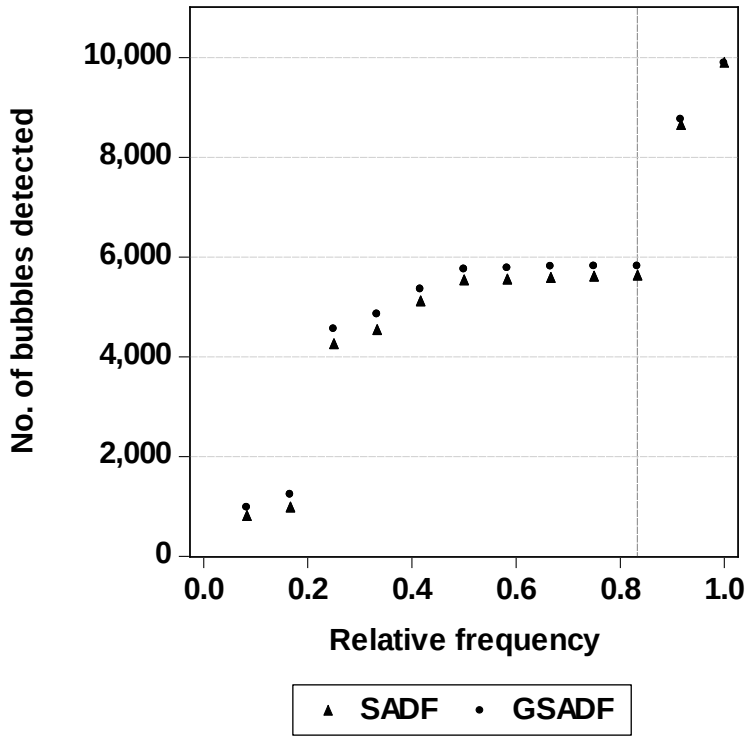

Figure 2. Number of bubbles detected by SADF and GSADF tests (empirical quantiles)

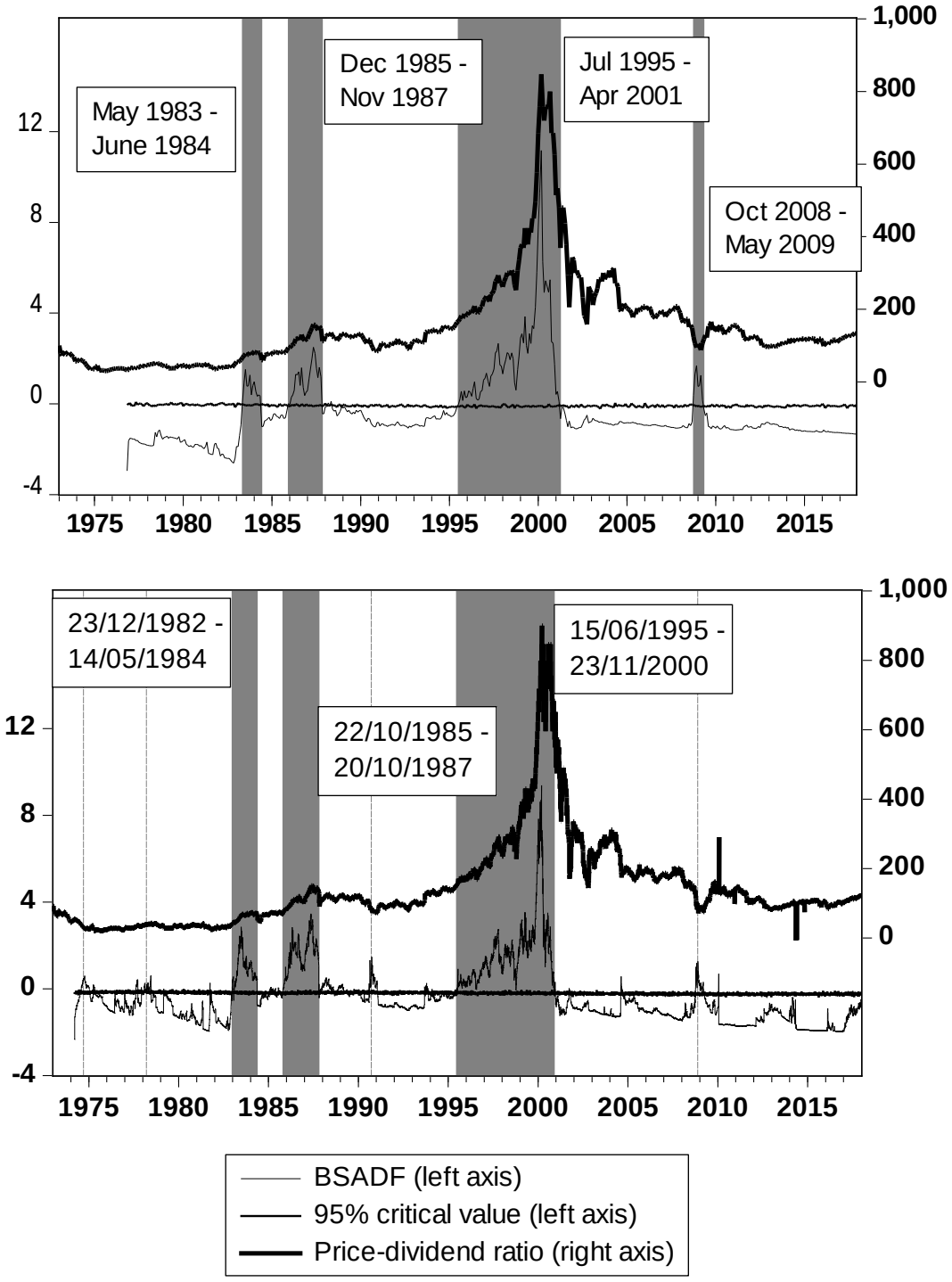

Figure 3. Date-stamping explosive behavior in monthly (upper panel) and daily (lower panel) NASDAQ price-dividend ratios