

Monschang, Verena; Wilfling, Bernd

Article — Published Version

Sup-ADF-style bubble-detection methods under test

Empirical Economics

Provided in Cooperation with:

Springer Nature

Suggested Citation: Monschang, Verena; Wilfling, Bernd (2020) : Sup-ADF-style bubble-detection methods under test, Empirical Economics, ISSN 1435-8921, Springer, Berlin, Heidelberg, Vol. 61, Iss. 1, pp. 145-172,
<https://doi.org/10.1007/s00181-020-01859-7>

This Version is available at:

<https://hdl.handle.net/10419/288327>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



<https://creativecommons.org/licenses/by/4.0/>



Sup-ADF-style bubble-detection methods under test

Verena Monschang¹ · Bernd Wilfling¹

Received: 8 February 2019 / Accepted: 14 March 2020 / Published online: 29 March 2020
© The Author(s) 2020

Abstract

In this paper, we analyze the capacity of supremum augmented Dickey–Fuller (SADF), generalized SADF (GSADF), and of several heteroscedasticity-adjusted sup-ADF-style tests for detecting and date-stamping financial bubbles. Our Monte Carlo simulations find that the majority of the sup-ADF-style tests exhibit substantial size distortions, when the data-generating process is subject to leverage effects. Moreover, the sup-ADF-style tests often have low empirical power in identifying a (flexible and empirically relevant) rational stock-price bubble, recently proposed in the literature. In a simulation study, we compare the effectiveness of two real-time bubble date-stamping procedures (Procedures 1 and 2), both based on variants of the backward SADF (BSADF) test. While Procedure 1 (predominantly) provides better estimates of the bubbles' origination and termination dates than Procedure 2, the first procedure frequently stamps non-existing bubbles. In an empirical application, we use NASDAQ data covering a time-span of 45 years and find that the bubble date-stamping outcomes of both procedures 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

1 Introduction

In a series of influential articles, Phillips, Wu, and Yu (2011; PWY hereafter) and Phillips et al. (2014, 2015a, b) have established a sound theoretical foundation of right-

✉ Bernd Wilfling
bernd.wilfling@wiwi.uni-muenster.de

Verena Monschang
verena.monschang@wiwi.uni-muenster.de

¹ Department of Economics (CQE), Westfälische Wilhelms-Universität Münster, Am Stadtgraben 9, 48143 Münster, Germany

tailed unit-root testing for explosiveness in time-series data. In the wake of this work, the most prominent testing procedures—the sup augmented Dickey–Fuller (SADF) test and its generalized version (the GSADF test)—have been applied in a plethora of empirical studies, 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 2019; Hu and Oxley 2018a) 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 such econometric software packages as EViews and R (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 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 the subsequent sections, we will make extensive use of this bubble specification.

In this article, we investigate the capacity of various sup-ADF-style testing procedures for detecting and date-stamping financial bubbles. Besides the original SADF and GSADF tests introduced by PWY and PSY, we consider the heteroscedasticity-adjusted variants (bootstrapped and sign-based SADF and GSADF tests), as established in Harvey, Leybourne, Sollis, and Taylor (2016; HLST) and Harvey, Leybourne, and Zu (2020; HLZ). In various ways, we extend and modify the size, power, and date-stamping analyses from the above-mentioned articles. Our investigation has the following major findings. (i) As a prominent equity-market volatility asymmetry, we consider the leverage effect, according to which negative shocks often have a relatively larger impact on volatility than positive shocks. To capture such heteroscedasticity, we study the effects of a threshold GARCH (TGARCH) volatility structure on the empirical size of the sup-ADF-style tests and find that several of them reveal considerable size distortions. (ii) We generate artificial stock-price data under the above-mentioned rational Rotermann-Wilfling bubble specification. Our simulations show that the sup-ADF-style tests often have low empirical power under this realistic bubble model.

(iii) In a simulation study, we contrast two real-time bubble date-stamping strategies—both based on variants of the backward SADF (BSADF) test—namely (1) the procedure established in PSY, and (2) a methodology using the sign-based test statistic of HLZ. For our simulated bubble settings, we find that the first strategy (predominantly) outperforms the second by yielding more accurate estimates of the bubbles’

origination and termination dates. However, the first procedure shows a pronounced tendency to date-stamp non-existing bubbles. (iv) In an empirical application, we apply the two bubble date-stamping strategies to NASDAQ data and compare the date-stamping results when using monthly versus daily observations. Our findings reveal that the dating strategies are sensitive to the practitioner's choice of data frequency. We explain this phenomenon by the structural changes in the data-generating process that come along with the frequency shift.

The remainder of the paper is organized as follows. Section 2 briefly reviews the essentials of the present-value stock-price model and the rational Rotermann–Wilfling bubble specification. Section 3 recapitulates the sup-ADF-style tests and analyzes their properties via Monte Carlo simulation. Section 4 investigates the date-stamping procedures, and Sect. 5 concludes.

2 Present-value model, rational bubbles, and explosiveness

PWY and 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

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

where $\mathbb{E}_t(\cdot)$ denotes the conditional expectation operator and D_{t+1} the dividend payment between t and $t+1$. The constant $r > 0$ is the 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

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

where B_t is any stochastic process satisfying the submartingale property

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

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 submartingale 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 $t \rightarrow \infty$. The most frequently applied rational, parametric specification satisfying all these properties is the Evans (1991) bubble. However, this

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.

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 Sect. 3, we 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 $\frac{1-\alpha}{\psi(1-\pi)} < 1$. The process $\{u_t\}_{t=1}^\infty$ is assumed to be i.i.d. with lognormally distributed $u_t \sim \mathbb{LN}(\frac{-\iota^2}{2}, \iota^2)$.² The constraint $\alpha \in (0, 1)$ ensures that the bubble never collapses to zero and can thus reflate. 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 $\frac{1-\alpha}{\psi(1-\pi)} - 1 < 0$.

Depending on the specific parameter constellation, the bubble specification (4) displays periodically recurring behavior with (stochastically evolving) deflation periods that can range between a ‘small/moderate correction’ and a ‘big crash’ within one or arbitrarily many periods. As an example, Fig. 1 displays four 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 Fig. 1.

For our stock-price simulations in Sect. 3, 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

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

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 .

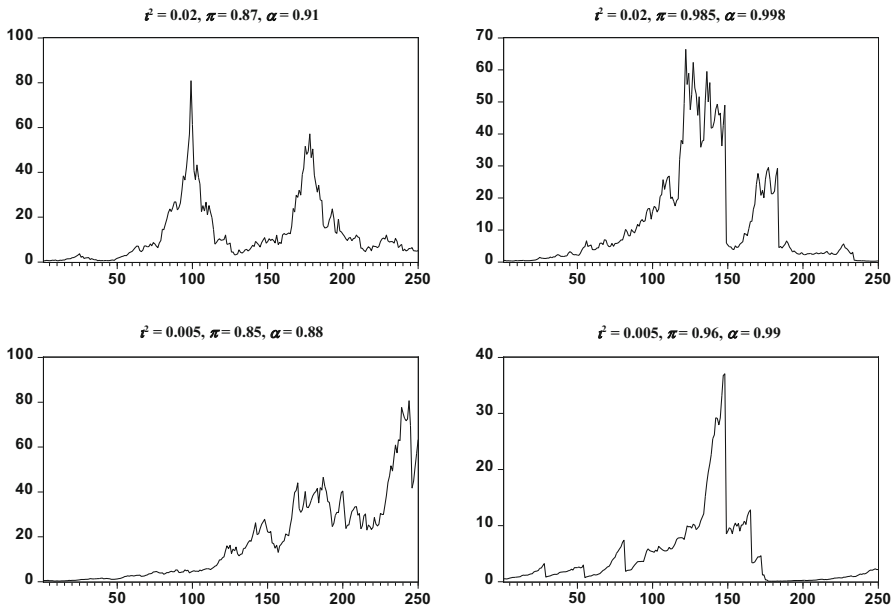


Fig. 1 Bubble trajectories of length $T = 250$ simulated according to Eq. (4) with common parameters $B_0 = 0.5$, $\psi = 0.9840$

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 neither constitutes a necessary nor a sufficient condition for deducing the existence of a bubble.

³ Cochrane (2011) presents overwhelming empirical evidence of time-varying discount rates. PWY indicate how time-varying discount factors may affect bubble detection at various spots in their quantitative considerations. Pertaining to the present-value framework, the authors argue that ‘[...] making the discount rate stationary and time-varying does not change the implication of submartingale (explosive) behavior [...], but complicates the analysis of the rational bubble solution.’ (PWY 2011, p. 204).

3 Sup-ADF-style tests

3.1 SADF and GSADF tests

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

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

where k is the transient lag-order, Δ denotes the first-difference operator, and $\varepsilon_t \stackrel{\text{i.i.d.}}{\sim} (0, \sigma^2)$. 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 $\{0, 1, \dots, T\}$ 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, PWY define the SADF test statistic as the sup-ADF-statistic from repeated estimation of the empirical regression (8) on a forward expanding sample sequence. Specifically, the authors consider as given, the minimal sample window width r_0 , set the subsample starting point $r_1 = 0$, and let the subsample 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

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

The GSADF test—suggested by PSY 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

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

PSY 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,

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

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

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

and

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

where $W(\cdot)$ denotes the standard Wiener process.

3.2 Heteroscedasticity adjustments

Recently, HLST and HLZ have addressed the issue of bubble detection under heteroscedasticity. We briefly review their testing procedures, which both refer to non-stochastic unconditional heteroscedastic patterns.

3.2.1 Wild-bootstrap SADF and GSADF tests

HLST suggest using a wild-bootstrap re-sampling scheme in order to obtain a size-controlled SADF testing procedure, in the presence of non-stationary (unconditional) volatility. Their wild-bootstrap algorithm consists of the following five steps.

Step 1. Generate a sequence $\{w_t\}_{t=2}^T$ of independent $\mathbb{N}(0, 1)$ random variables and construct the following series of bootstrap innovations $\{\epsilon_t^*\}_{t=1}^T$:

$$\epsilon_t^* = \begin{cases} 0, & t = 1 \\ w_t \Delta y_t, & t = 2, \dots, T \end{cases}$$

Step 2. Construct the bootstrap sample as the partial sums

$$y_t^* \equiv \sum_{s=1}^t \epsilon_s^*, \quad t = 1, \dots, T.$$

Step 3. For a subsample running from $r_1 = 0$ to r_2 , consider the t -value for $\hat{\theta}^*$ in the fitted OLS regression,

$$y_t^* = \hat{c}^* + \hat{\theta}^* y_{t-1}^* + \hat{\epsilon}_t^*,$$

and denote this bootstrap ADF-statistic by $\left[\text{ADF}_{r_1=0}^{r_2} \right]^*$. Compute the bootstrap SADF test statistic (denoted by SADF^*) in the usual way:

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

Step 4. Repeat Steps 1 to 3 B times to generate a sample of bootstrapped SADF statistics, $\left\{ \text{SADF}_b^* \right\}_{b=1}^B$. Use this sample to approximate the cumulative distribution function of SADF^* , denoted by $G_T^*(\cdot)$.

Step 5. To conduct the wild-bootstrap SADF test at the significance level α , compute the SADF test statistic from Eq. (9), using the original sample $\{y_t\}_{t=1}^T$, and reject the unit-root null hypothesis, if the SADF statistic exceeds the $1 - \alpha$ quantile $G_T^{*-1}(1 - \alpha)$.

We denote the wild-bootstrap SADF testing procedure by $\text{SADF}^{\text{bootstrap}}$, in order to distinguish it from the original SADF test described in Sect. 3.1. In our subsequent analysis, we also apply the bootstrap algorithm to the original GSADF test and denote this variant by $\text{GSADF}^{\text{bootstrap}}$. We emphasize that—due to the re-sampling scheme—the critical values of the $\text{SADF}^{\text{bootstrap}}$ and $\text{GSADF}^{\text{bootstrap}}$ tests are always adapted to the specific trajectory $\{y_t\}_{t=1}^T$, which is to be tested for explosiveness. Therefore, the critical values of the bootstrapped tests, obtained from $\{y_t\}_{t=1}^T$, are not universally applicable to other trajectories.

3.2.2 Sign-based SADF and GSADF tests

As a further heteroscedasticity adjustment, HLZ propose the sign-based GSADF test under deterministically time-varying (unconditional) volatility. The idea is to compute the GSADF statistic from Eq. (10) not from the directly observed series $\{y_t\}_{t=1}^T$, but rather from the transformed series

$$C_t = \sum_{s=2}^t \text{sign}(\Delta y_s), \tag{14}$$

where $\text{sign}(x) \in \{-1, 1\}$ for $x \leq 0$ and $x > 0$, respectively, is the sign function. Per definition, $\{C_t\}_{t=2}^T$ is the series of cumulated signs of the first differences.

For a subsample running from r_1 to r_2 , we consider the t -value for $\hat{\vartheta}$ in the fitted OLS regression,

$$C_t = \hat{\vartheta} \cdot C_{t-1} + \hat{\epsilon}_t, \tag{15}$$

which we denote by $[\text{ADF}_{r_1}^{r_2}]^{\text{sign-based}}$. (We note that Eq. (15) is estimated without intercept.) Then, HLZ define their sign-based GSADF test statistic as

$$\text{GSADF}^{\text{sign-based}}(r_0) \equiv \sup_{\substack{r_2 \in [r_0, 1] \\ r_1 \in [0, r_2 - r_0]}} \left\{ [\text{ADF}_{r_1}^{r_2}]^{\text{sign-based}} \right\}. \quad (16)$$

In our analysis below, we also consider sign-based SADF tests (denoted by $\text{SADF}^{\text{sign-based}}$) by setting $r_1 = 0$ in Eq. (16).

3.3 Monte Carlo study

We first approximate asymptotic critical values of the SADF, $\text{SADF}^{\text{sign-based}}$, GSADF and $\text{GSADF}^{\text{sign-based}}$ tests.⁴ In a second step, we analyze the empirical size and power of the tests.

3.3.1 Critical values and size distortions

Given Eqs. (9), (10), (12), (13), and (16), the SADF, $\text{SADF}^{\text{sign-based}}$, GSADF, and $\text{GSADF}^{\text{sign-based}}$ test statistics, as well as their asymptotic distributions depend on the minimal window size r_0 . In line with PSY, we adopt the rule $r_0 = 0.01 + 1.8/\sqrt{T}$, and—whenever considering finite samples—focus on the specific sample sizes $T \in \{100, 200, 400, 800, 1600\}$.

We approximate the asymptotic critical values of the four test 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, $\text{SADF}^{\text{sign-based}}$ and GSADF, $\text{GSADF}^{\text{sign-based}}$ statistics, respectively.

The upper block of Table 1 reports our asymptotic critical values of the SADF and GSADF tests. The values 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 decrease in minimal window size r_0 . (ii) The GSADF 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 prefer our critical values from Table 1 to those provided by PSY. The lower block of Table 1 displays our asymptotic critical values of the $\text{SADF}^{\text{sign-based}}$ and $\text{GSADF}^{\text{sign-based}}$ tests, which exhibit the same qualitative features as those from the upper block.

⁴ We do not provide critical values of the bootstrap variants, since these are trajectory-specific (see Sect. 3.2.1).

Table 1 Asymptotic critical values of SADF, SADF^{sign-based}, GSADF, GSADF^{sign-based} 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
	SADF ^{sign-based}			GSADF ^{sign-based}		
	90%	95%	99%	90%	95%	99%
$r_0 = 0.190$	2.2923	2.6306	3.2613	2.7665	3.0584	3.6267
$r_0 = 0.137$	2.3696	2.6970	3.3206	2.9413	3.2310	3.7556
$r_0 = 0.100$	2.4304	2.7502	3.3635	3.0963	3.3675	3.8710
$r_0 = 0.074$	2.4844	2.7987	3.4024	3.2372	3.5175	4.0018
$r_0 = 0.055$	2.5261	2.8402	3.4331	3.3771	3.6236	4.1449

Large-sample critical values are approximated by simulating the data-generating process from Eq. (11) with $\theta = 1$ and $a = \eta = 1$, and using Eq. (14). The sample size is $T = 5000$. The numbers of replications are 100,000 for the SADF, SADF^{sign-based}, and 12,500 for the GSADF, GSADF^{sign-based} tests, respectively

Table 2 Sizes of SADF, SADF^{sign-based}, GSADF, GSADF^{sign-based} 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	0.0390	0.0444	0.0385	0.0431	0.0457
GSADF	0.0519	0.0566	0.0496	0.0484	0.0499
SADF ^{sign-based}	0.0472	0.0489	0.0467	0.0466	0.0455
GSADF ^{sign-based}	0.0911	0.0905	0.0782	0.0637	0.0601

Sizes are obtained by simulating data (under the null hypothesis) from the specification in Eq. (11) with parameters $a = \eta = \theta = 1$, and Eq. (14). Using 10,000 replications, the sizes are computed via the (asymptotic) 95% critical values from Table 1

In order to check the validity of our asymptotic critical values, when applied to finite samples, we simulate sizes for the four 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., we use the 95% critical values from Table 1). For the SADF, SADF^{sign-based}, and the GSADF tests, we do not find substantial size distortions. By contrast, the GSADF^{sign-based} test appears to be slightly over-

Table 3 Finite-sample critical values of SADF^{sign-based} and GSADF^{sign-based} tests

	SADF ^{sign-based}			GSADF ^{sign-based}		
	90%	95%	99%	90%	95%	99%
$T = 100, r_0 = 0.190$	2.2621	2.6467	3.4033	3.0236	3.4634	4.5808
$T = 200, r_0 = 0.137$	2.3425	2.6935	3.3623	3.1485	3.5083	4.3896
$T = 400, r_0 = 0.100$	2.4027	2.7383	3.3719	3.2211	3.5397	4.2054
$T = 800, r_0 = 0.074$	2.4631	2.7796	3.4014	3.3287	3.6257	4.2266
$T = 1600, r_0 = 0.055$	2.5080	2.8130	3.4118	3.4167	3.6794	4.2255

Finite-sample critical values are approximated by simulating the data-generating process from Eq. (11) with parameters $a = \eta = \theta = 1$, and Eq. (14). The numbers of replications are 100,000 and 12,500 for the SADF^{sign-based} and GSADF^{sign-based} tests, respectively

sized, when applying asymptotic critical values in finite samples (in particular, for $T \in \{100, 200, 400\}$).

The latter finding is in line with HLZ, who demonstrate that the convergence of finite-sample critical values to their asymptotic counterparts is rather slow for the GSADF^{sign-based} test. To be on the safe side, we compute finite-sample critical values for both sign-based tests (SADF^{sign-based}, GSADF^{sign-based}), using the sample sizes $T \in \{100, 200, 400, 800, 1600\}$. Table 3 displays the finite-sample critical values, which we use in our analysis below, to ensure correctly sized sign-based tests.

The process specification in Eq. (11) assumes homoscedastic errors and ignores conditional heteroscedasticity, a well-documented phenomenon in all types of financial markets. To assess the impact on the SADF and GSADF testing procedures, PSY address the sizes of both tests under unit-root processes with GARCH errors. Using standard GARCH(1, 1) errors as 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 the unit-root DGP from Eq. (11) with $a = \eta = \theta = 1$ under (threshold) TGARCH(1, 1) errors,

$$\varepsilon_t = s_t \sqrt{h_t}, \quad s_t \stackrel{i.i.d.}{\sim} \mathbb{N}(0, 1), \tag{17}$$

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

where $\mathbb{I}(\cdot)$ 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 (Zakoian 1994).

The first two rows of Table 4 display the simulated sizes of the SADF and GSADF 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)

Table 4 Sizes of six sup-ADF-style 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
SADF ^{sign-based}	0.0435	0.0428	0.0417	0.0464	0.0505
GSADF ^{sign-based}	0.0443	0.0429	0.0439	0.0478	0.0480
SADF ^{bootstrap}	0.0412	0.0277	0.0181	0.0107	0.0032
GSADF ^{bootstrap}	0.0279	0.0231	0.0292	0.0316	0.0440

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, (i) the sizes of the SADF, GSADF tests are computed with the (asymptotic) 95% critical values from Table 1, (ii) the sizes of the SADF^{sign-based}, GSADF^{sign-based} tests are computed with the finite-sample critical values from Table 3. (iii) For the SADF^{bootstrap}, GSADF^{bootstrap} tests, we set $B = 499$ in Step 4 of the HLST wild-bootstrap algorithm. For the size of the GSADF^{bootstrap} test, we only use 2,000 replications in the case of $T = 1,600$ (due to computational burden)

tranquil period between January 1988 and December 1994.⁵ Apparently, the SADF and GSADF tests exhibit substantial size distortions under volatility asymmetry in the form of TGARCH(1, 1) errors.

Rows 3–6 of Table 4 display the sizes of the four heteroscedasticity-adjusted tests under TGARCH(1, 1) errors. We first recall that the four tests—on the grounds of their construction in the HLST and HLZ articles—adjust for unconditional (but not for conditional) heteroscedasticity patterns in the data-generating processes. However, Rows 3, 4 of Table 4 indicate that the SADF^{sign-based} and GSADF^{sign-based} tests operate well in controlling the actual sizes under TGARCH errors. By contrast, the SADF^{bootstrap} and GSADF^{bootstrap} tests (Rows 5, 6 in Table 4) exhibit substantial undersizing. These tests—for which, in line with HLST, we use $B = 499$ bootstrap replications—reject the null hypothesis of a unit-root too rarely (as opposed to the SADF and GSADF tests from Rows 1–2, which reject the null too often).

3.3.2 Empirical power

We now address the empirical power properties of the SADF, SADF^{sign-based}, GSADF and GSADF^{sign-based} tests. We distinguish between homoscedastic and TGARCH(1, 1) stock-price trajectories.

Homoscedastic trajectories

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

⁵ We fitted TGARCH(1, 1) volatility specifications to several subsamples of the NASDAQ data. In line with the leverage-effect literature, we find significant ϕ -estimates (at conventional levels) for the majority of the subsamples. We describe the data set in Sect. 4.

Table 5 Empirical power of SADF, SADF^{sign-based}, SADF^{bootstrap}, GSADF, GSADF^{sign-based}, GSADF^{bootstrap} tests under a rational bubble

	$T = 100$ $r_0 = 0.190$	$T = 200$ $r_0 = 0.137$	$T = 400$ $r_0 = 0.100$	$T = 800$ $r_0 = 0.074$	$T = 1,600$ $r_0 = 0.055$
(a) Homoscedastic trajectories					
SADF	0.0373	0.0857	0.5579	0.9758	1.0000
GSADF	0.0448	0.1075	0.5806	0.9782	1.0000
SADF ^{sign-based}	0.0727	0.0982	0.3025	0.8195	0.9986
GSADF ^{sign-based}	0.0683	0.0971	0.2925	0.7523	0.9936
SADF ^{bootstrap}	0.0308	0.0579	0.2834	0.6348	0.8298
GSADF ^{bootstrap}	0.0130	0.0399	0.2780	0.6305	0.8203
(b) TGARCH(1, 1) trajectories					
SADF ^{adjusted}	0.0109	0.0115	0.1538	0.8714	0.9997
GSADF ^{adjusted}	0.0114	0.0120	0.1321	0.9607	0.9993
SADF ^{sign-based}	0.1323	0.1435	0.2055	0.6557	0.9931
GSADF ^{sign-based}	0.1051	0.1038	0.1637	0.6007	0.9825
SADF ^{bootstrap}	0.0107	0.0108	0.0745	0.5129	0.8182
GSADF ^{bootstrap}	0.0086	0.0075	0.0790	0.5120	0.8114

The stock-price series are generated as $P_t = P_t^f + B_t$, with P_t^f from Eqs. (5) and (6), where Eq. (5) includes either homoscedastic errors with $\sigma_e^2 = 0.4476$ [Block (a)], or TGARCH(1, 1) errors [Block (b)] with TGARCH parameters as in Table 4. B_t is generated as in Eq. (4). The remaining parameters are set to $\mu = 0$, $D_0 = 1.6942$, $B_0 = 10.1925$, $\psi = 0.9840$, $\iota^2 = 0.0061$, $\pi = 0.9595$ and $\alpha = 0.9675$. For the power calculation, we use 10,000 replications, the 95% critical values from Table 1 for the SADF and GSADF tests, and the finite-sample critical values from Table 3 for the SADF^{sign-based}, GSADF^{sign-based} tests. For the SADF^{adjusted}, GSADF^{adjusted} tests in Block (b), we simulated critical values (not reported). For the SADF^{bootstrap}, GSADF^{bootstrap} tests, we set $B = 499$ in Step 4 of the HLST wild-bootstrap algorithm

rational bubble B_t following the lognormal 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. Specifically, 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 $\alpha = 0.9675$.

Block (a) of Table 5 ('Homoscedastic trajectories') reports the results of our power analysis for the sample sizes $T \in \{100, 200, 400, 800, 1600\}$. We use the 95% critical values from Table 1 for the SADF, GSADF tests, the 95% finite-sample critical values from Table 3 for the SADF^{sign-based}, GSADF^{sign-based} tests, $B = 499$ bootstrap replications for the SADF^{bootstrap}, GSADF^{bootstrap} tests, and 10,000 Monte Carlo replications. Our analysis yields six 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) By contrast, the power of the SADF^{sign-based} test always (slightly) exceeds the power of the GSADF^{sign-based} test. (iii) The SADF^{sign-based} and GSADF^{sign-based} tests have higher power than

the original SADF, GSADF tests for $T = 100$, and in the SADF-case also for $T = 200$. In all other cases, the sign-based tests exhibit lower power. (iv) The power of all six tests increases with increase in sample size. (v) The power of the first four tests (SADF, GSADF, SADF^{sign-based}, GSADF^{sign-based}) is extremely low for $T = 100, 200$ and improves only moderately for $T = 400$. The tests perform satisfactorily for $T = 800$. For $T = 1600$, the four tests identify the bubble in (almost) each of the 10,000 simulated stock-price series. (vi) The power values of the SADF^{bootstrap}, GSADF^{bootstrap} tests unambiguously fall below the power values of all other tests in the five scenarios. In anticipation of our subsequent analysis, we emphasize that this clear-cut result also holds under TGARCH(1, 1) trajectories, as shown in Block (b) of Table 5. Again, in each scenario, the SADF^{bootstrap}, GSADF^{bootstrap} tests exhibit the (by far) lowest power values among all 6 tests analyzed. We note that the computation of the 20 SADF^{bootstrap}, GSADF^{bootstrap} power values in Table 5 is extremely time-consuming.⁶ Thus, owing to (1) sizing problems, (2) low power values, and (3) computational burdens, we exclude the SADF^{bootstrap}, GSADF^{bootstrap} tests from our further analysis.

The Blocks (a) of Table 6 ('Homoscedastic trajectories') report the power of the remaining four sup-ADF-style tests, when some model parameters are singly (or, as π and α , jointly) varied, while all other parameters are held constant at their base-levels given in the Note of Table 6 (*ceteris-paribus* analysis). Explicitly, we let (i) the dividend drift μ range between 0 and 0.003 (Settings 1–4), (ii) the discount factor ψ range between 0.975 and 0.990 (Settings 5–8), and (iii) the probability π range between 0.35 and 0.95 (Settings 9–12). 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 1.06. We conduct our analysis with 10,000 Monte Carlo replications for each parameter setting, and the sample size $T = 400$.⁷

The Blocks (a) in Table 6 ('Homoscedastic trajectories') yield the following four findings. (i) The SADF and GSADF tests have higher power than their sign-based counterparts in 10 of 12 settings (the two exceptions are the Settings 8 and 12). (ii) A variation in the dividend drift μ (*ceteris paribus*) does not substantially affect the power of either test. This result is not surprising for two reasons. First, the SADF and GSADF tests are based on Eq. (8), which captures the effects of the dividend drift. Second, the computation of the sign-based test statistics rests on the series $\{C_t\}$ from Eq. (14), which—by construction—is independent of μ . (iii) The power of the four tests decreases dramatically with an increase in discount factor ψ . For instance, for $\psi = 0.99$ (Setting 8), the tests only detect 9.73%, 12.35%, 11.00%, and 9.67% of the simulated bubbles, respectively. 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.

⁶ The computation needed 13 days, using parallelized MATLAB algorithms in a computer environment with more than 100 kernels. Technical details are available upon request.

⁷ We emphasize that our 12 parameter settings in Table 6 are all empirically relevant. Each set of parameters is able to produce bubble trajectories that are qualitatively similar to those in Fig. 1. The chosen sample size $T = 400$ covers time spans of (i) more than 33 years under monthly observations, and (ii) less than 2 years in the case of daily data. We choose $T = 400$ in order to be comparable with the settings in PSY.

Table 6 Power of SADF, GSADF, SADF^{sign-based}, GSADF^{sign-based} tests under varying bubble parameters

	$\mu = 0.000$ (Setting 1)	$\mu = 0.001$ (Setting 2)	$\mu = 0.002$ (Setting 3)	$\mu = 0.003$ (Setting 4)
(a) Homoscedastic trajectories				
SADF	0.5579	0.5603	0.5543	0.5527
GSADF	0.5806	0.5814	0.5774	0.5751
SADF ^{sign-based}	0.3025	0.3021	0.2935	0.2983
GSADF ^{sign-based}	0.2925	0.2883	0.2846	0.2905
(b) TGARCH(1, 1) trajectories				
SADF ^{adjusted}	0.1538	0.1551	0.1552	0.1600
GSADF ^{adjusted}	0.1321	0.1347	0.1343	0.1372
SADF ^{sign-based}	0.2055	0.1932	0.2066	0.2083
GSADF ^{sign-based}	0.1637	0.1582	0.1698	0.1685
	$\psi = 0.975$ (Setting 5)	$\psi = 0.98$ (Setting 6)	$\psi = 0.985$ (Setting 7)	$\psi = 0.990$ (Setting 8)
(a) Homoscedastic trajectories				
SADF	0.9891	0.8646	0.4531	0.0973
GSADF	0.9890	0.8755	0.4852	0.1235
SADF ^{sign-based}	0.8749	0.5724	0.2412	0.1100
GSADF ^{sign-based}	0.8338	0.5448	0.2343	0.0967
(b) TGARCH(1, 1) trajectories				
SADF ^{adjusted}	0.8790	0.4986	0.1029	0.0166
GSADF ^{adjusted}	0.8540	0.4566	0.0903	0.0171
SADF ^{sign-based}	0.6795	0.3613	0.1837	0.1612
GSADF ^{sign-based}	0.6584	0.3377	0.1435	0.1103

(iv) Increasing π -probabilities (Settings 9–12) entail substantial decreases in the power of the SADF and GSADF tests, while the impact on the power of the SADF^{sign-based}, GSADF^{sign-based} tests appears ambiguous. A plausible explanation of the SADF and GSADF power reduction could be as follows. Recall that π represents the likelihood of ongoing bubble growth at the constant rate of 6% (which we achieve by an appropriate adjustment of α). For $\pi = 0.95$, $\alpha = 0.9975$ (Setting 12), both tests only detect 8.04% and 9.72% of the bubbles. *Prima facie*, this finding seems counter-intuitive, since we would expect the tests to exhibit higher power when the probability of bubble inflation (bubble growth) increases. Our explanation of this phenomenon rests on the fact that the joint variation of the parameters π and α keeps the bubble inflation rate, given by $\alpha/(\psi\pi) - 1$, stable at the 6% 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 of the SADF, GSADF tests is capable of coping with these opposing effects on bubble dynamics.

Table 6 continued

	$\pi = 0.35$ $\alpha = 0.3675$ (Setting 9)	$\pi = 0.55$ $\alpha = 0.5775$ (Setting 10)	$\pi = 0.75$ $\alpha = 0.7875$ (Setting 11)	$\pi = 0.95$ $\alpha = 0.9975$ (Setting 12)
(a) Homoscedastic trajectories				
SADF	0.5626	0.5108	0.4250	0.0804
GSADF	0.5813	0.5355	0.4555	0.0972
SADF ^{sign-based}	0.2065	0.1844	0.2970	0.1416
GSADF ^{sign-based}	0.1914	0.1723	0.2843	0.1149
(b) TGARCH(1, 1) trajectories				
SADF ^{adjusted}	0.1492	0.1530	0.1363	0.0345
GSADF ^{adjusted}	0.1277	0.1339	0.1231	0.0315
SADF ^{sign-based}	0.1859	0.1779	0.2051	0.1694
GSADF ^{sign-based}	0.1460	0.1412	0.1622	0.1163

The stock-price series are generated as $P_t = P_t^f + B_t$, with P_t^f as in Eqs. (5) and (6), where Eq. (5) includes either homoscedastic errors with $\sigma_\epsilon^2 = 0.4476$ [Block (a)], or TGARCH(1, 1) errors [Block (b)] with TGARCH parameters as in Table 4. B_t is generated as in Eq. (4). The basic set of parameters is $\mu = 0, D_0 = 1.6942, B_0 = 10.1925, \psi = 0.9840, \iota^2 = 0.0061, \pi = 0.9595, \alpha = 0.9675$. 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 $T = 400, r_0 = 0.1$, use 10,000 replications for each parameter setting, the 95% critical values from Table 1 for the SADF, GSADF tests, and the finite-sample critical values from Table 3 for the SADF^{sign-based}, GSADF^{sign-based} tests. For the SADF^{adjusted}, GSADF^{adjusted} tests under TGARCH trajectories, we simulated critical values (not reported)

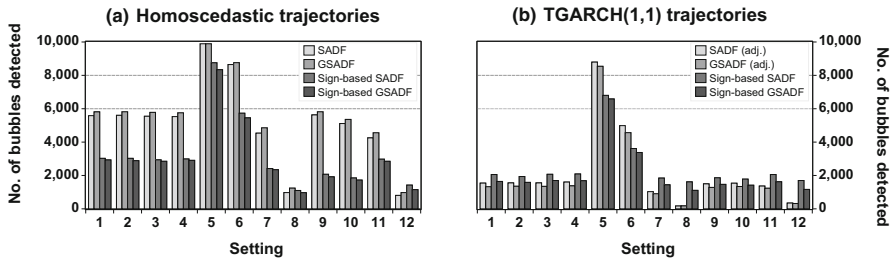


Fig. 2 Numbers of bubbles detected by SADF, GSADF, SADF^{sign-based}, GSADF^{sign-based} tests under (a) homoscedastic trajectories, (b) TGARCH(1,1) trajectories

Panel (a) of Fig. 2 (‘Homoscedastic trajectories’) illustrates the overall results of our power analysis. For this bar graph, we multiplied all power values in Table 6 by 10,000, so that the values now represent the number of bubbles correctly detected by the tests in each parameter setting.⁸ Four findings are striking. (i) The bubble-detection capacity of the GSADF test only slightly outperforms that of the SADF test. For the sign-based tests, the reverse is true, with the SADF^{sign-based} tests slightly outperforming the GSADF^{sign-based} tests. (ii) In 10 of 12 parameter settings, all four tests detect less than 6,000 (of 10,000) bubbles per setting, that is, more than 40%

⁸ Recall that we use 10,000 replications in each setting, where each replication contains a bubble.

of the existing bubbles remain undetected. (iii) Pertaining to the sign-based tests, the bubble-detection capacity is even poorer. The $SADF^{\text{sign-based}}$, $GSADF^{\text{sign-based}}$ tests detect less than 6,000 bubbles in 11 of 12 settings (and in 10 of 12 settings, even less than 3,100 bubbles). (iv) Only in 2 of 12 settings, the SADF and GSADF tests identify more than 8,000 bubbles, while the sign-based tests detect more than 8,000 bubbles only in 1 of 12 settings.

TGARCH(1,1) trajectories

Finally, we include TGARCH-heteroscedasticity in the stock-price trajectories. We consider identical parameter settings as under homoscedastic trajectories in Block (a) of Tables 5 and 6, but specify TGARCH(1,1) errors (instead of Gaussian white-noise) in the fundamental stock-price P_t^f from Eqs. (5) and (6). We choose the same TGARCH parameters as in Sect. 3.3.1 ($\omega = 0.4387$, $\gamma = 0$, $\beta = 0.9319$, $\phi = 0.1306$).

Block (b) of Table 5 reports the power of the SADF, GSADF tests and their sign-based counterparts for the sample sizes $T \in \{100, 200, 400, 800, 1600\}$ under TGARCH trajectories. Since the conventional SADF and GSADF tests exhibit substantial size distortions under TGARCH heteroscedasticity (see Sect. 3.3.1), we follow the lines of HLZ (p. 11) and size-correct both tests infeasibly. To this end, we simulate finite-sample critical values using the data-generating process from Eq. (11) with $a = \eta = \theta = 1$ and TGARCH(1, 1) errors. We do not report the critical values here (they are available upon request), but denote the infeasibly size-corrected tests by $SADF^{\text{adjusted}}$ and $GSADF^{\text{adjusted}}$. Comparing the power values from Block (b) in Table 5 with those from Block (a), we find qualitatively similar features under TGARCH as under homoscedastic trajectories. However, for the sample sizes $T \in \{400, 800, 1600\}$ the tests under TGARCH have lower power than their counterparts under homoscedasticity.

The latter finding is further strengthened by Table 6, in which Block (b) displays the power values under TGARCH(1, 1) stock-price trajectories for $T = 400$, $r_0 = 0.1$, when some of the model parameters are varied. Panel (b) of Fig. 2 ('TGARCH(1,1) trajectories') shows the numbers of bubbles detected by the respective tests. Visual inspection of both panels in Fig. 2 reveals that the bubble-detection rates are often substantially lower under TGARCH than under homoscedastic stock-price trajectories. However, Panel (b) shows that the $SADF^{\text{sign-based}}$ and $GSADF^{\text{sign-based}}$ tests outperform their (size-adjusted) counterparts in 10 of 12 parameter settings (exceptions are the settings 5 and 6).

4 Bubble date-stamping

The aim of this section is threefold. In Sect. 4.1, we review bubble date-stamping procedures, which are based on the sup-ADF-style tests from Sect. 3. In Sect. 4.2, we investigate the performance of these date-stamping procedures in a simulation study, using a completely specified data-generating process with known TGARCH(1, 1)-

heteroscedasticity. In Sect. 4.3, we apply the procedures to NASDAQ data, for which the DGP—and in particular the heteroscedasticity pattern—is completely unknown.

4.1 Date-stamping procedures

In addition to mere bubble detection, PSY propose a date-stamping procedure for estimating the (fractional) origination and termination dates (denoted by r_e and r_f) of bubbles in real time.⁹ 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 statistic is then defined as the sup-value of the ADF statistics computed in Step (i):

$$\text{BSADF}_{r_2}(r_0) \equiv \sup_{r_1 \in [0, r_2 - r_0]} \{ \text{ADF}_{r_1}^{r_2} \}. \quad (19)$$

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 $\text{BSADF}_{r_2}(r_0)$ 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

$$\hat{r}_e = \inf_{r_2 \in [r_0, 1]} \{ r_2 : \text{BSADF}_{r_2}(r_0) > \text{scv}_{r_2}^{\beta_T} \}, \quad (20)$$

where β_T denotes the significance level and $\text{scv}_{r_2}^{\beta_T}$ is the $100(1 - \beta_T)\%$ critical value of the SADF test statistic based on $\lfloor Tr_2 \rfloor$ observations.¹⁰

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

$$\hat{r}_f = \inf_{r_2 \in [\hat{r}_e + \delta \log(T)/T, 1]} \{ r_2 : \text{BSADF}_{r_2}(r_0) < \text{scv}_{r_2}^{\beta_T} \}. \quad (21)$$

We note that the parameter δ in Eq. (21) controls for the minimal duration of the bubble.

Equations (19)–(21) illustrate that PSY date-stamping ultimately rests on Dickey–Fuller test statistics, which may lead to unreliable conclusions under heteroscedastic

⁹ PWY establish a predecessor date-stamping methodology, similar to the PSY procedure. However, the PWY procedure may be inconsistent in the presence of multiple bubbles (see PSY, pp. 1044–1045).

¹⁰ PSY point out that the significance level β_T may depend on the sample size T and shrink to zero as $T \rightarrow \infty$.

data. In view of this, we additionally consider a modified PSY dating strategy that uses the HLZ sign-based test statistic from Sect. 3.2.2, and which we denote by $PSY^{\text{sign-based}}$. To this end, we first define the sign-based BSADF test statistic

$$[BSADF_{r_2}(r_0)]^{\text{sign-based}} \equiv \sup_{r_1 \in [0, r_2 - r_0]} \left\{ [ADF_{r_1}^{r_2}]^{\text{sign-based}} \right\}. \tag{22}$$

We then estimate the origination date $[Tr_e]$ of a bubble via

$$\hat{r}_e^{\text{sign-based}} = \inf_{r_2 \in [r_0, 1]} \left\{ r_2 : [BSADF_{r_2}(r_0)]^{\text{sign-based}} > \text{sbcv}_{r_2}^{\beta_T} \right\}, \tag{23}$$

where β_T is the significance level and $\text{sbcv}_{r_2}^{\beta_T}$ the $100(1 - \beta_T)\%$ critical value of the $SADF^{\text{sign-based}}$ test statistic based on $[Tr_2]$ observations. Similarly, we estimate the associated termination date $[Tr_f]$ via

$$\hat{r}_f^{\text{sign-based}} = \inf_{r_2 \in [\hat{r}_e^{\text{sign-based}} + \delta \log(T)/T, 1]} \left\{ r_2 : [BSADF_{r_2}(r_0)]^{\text{sign-based}} < \text{sbcv}_{r_2}^{\beta_T} \right\}. \tag{24}$$

We note that HLZ discuss the $PSY^{\text{sign-based}}$ dating strategy from Eqs. (22)–(24) and find that the procedure may be unable to estimate the termination date consistently. To overcome this drawback, they propose an alternative procedure that provides consistent estimates of the bubble’s origination and termination dates. However, this refined methodology (i) is only able to date-stamp one bubble, and (ii) constitutes an *ex-post* rather than a real-time strategy. We therefore do not consider it in our analysis.

A third obvious date-stamping strategy may consist of embedding the bootstrap ADF-statistic from Sect. 3.2.1 into Eqs. (19)–(21) ($PSY^{\text{bootstrap}}$). However, besides the low power values reported in Table 5, the computational burden, associated with the bootstrap tests, renders the $PSY^{\text{bootstrap}}$ date-stamping procedure hard to apply in practice (cf. Footnote 6). We therefore do not present computational results for the $PSY^{\text{bootstrap}}$ procedure.¹¹

4.2 Simulation study

We analyze the impact of conditional heteroscedasticity on the performance of PSY date-stamping from Eqs. (19)–(21) and compare the results with the $PSY^{\text{sign-based}}$ strategy from Eqs. (22)–(24). We consider the data-generating process suggested in Phillips and Shi (2018), but modify it by imposing TGARCH(1, 1) errors. Formally,

¹¹ Recently, Phillips and Shi (2020) propose a new, computationally less demanding (than $PSY^{\text{bootstrap}}$ based on HLST) bootstrap date-stamping procedure, which is implemented in the (psymonitor) R package. This new procedure is, however, practically not amenable to our simulation study in Sect. 4.2. As shown in Figs. 2 and 3 of Phillips and Shi (2020), this date-stamping procedure typically requires subjective judgmental decisions (via eyeballing) on the estimated bubble termination and origination dates, which is impossible to execute in an investigation with 10,000 trajectories. We readdress this issue in Sect. 4.3.

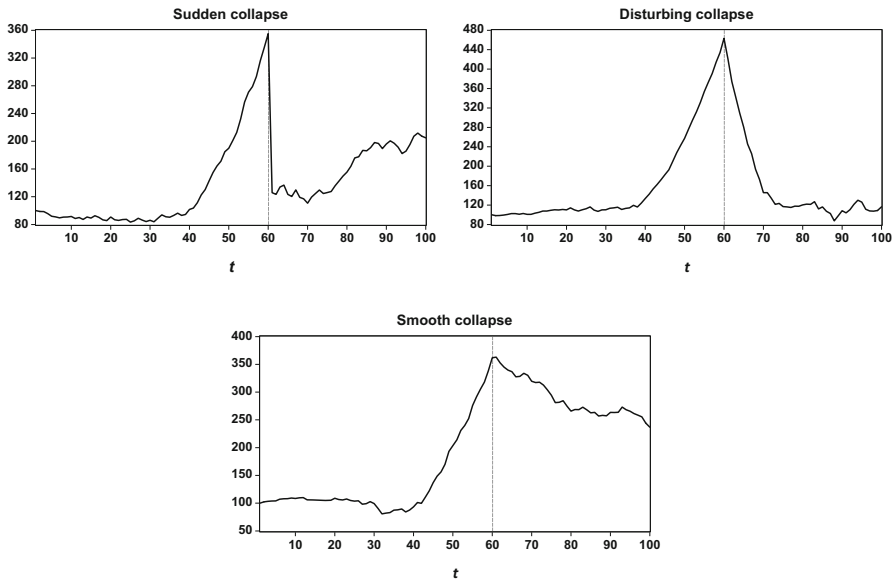


Fig. 3 Bubble trajectories according to Eq. (25) with alternative collapse patterns

the DGP is given by

$$y_t = \begin{cases} a \cdot T^{-\eta} + y_{t-1} + \varepsilon_t, & t \in N_0 \cup N_1 \\ \delta_T y_{t-1} + \varepsilon_t, & t \in B \\ \gamma_T y_{t-1} + \varepsilon_t, & t \in C \end{cases}, \tag{25}$$

where $\varepsilon_t \sim \text{TGARCH}(1, 1)$ as in Eqs. (17) and (18), $\delta_T = 1 + c_1 T^{-\tilde{\alpha}}$, $\gamma_T = 1 - c_2 T^{-\tilde{\beta}}$. Besides the parameters, Phillips and Shi (2018) define the following time-periods. (i) The set $B = [\lfloor Tr_e \rfloor, \lfloor Tr_f \rfloor]$, during which the bubble inflates. This period has duration $d_B = \lfloor T(r_f - r_e) \rfloor$. (ii) The set $C = (\lfloor Tr_f \rfloor, \lfloor Tr_r \rfloor]$, during which the bubble deflates/collapses. Here, $\lfloor Tr_r \rfloor$ is the date of market recovery, and this period has duration $d_C = \lfloor T(r_r - r_f) \rfloor$. (iii) The set $N_0 \cup N_1 = [1, \lfloor Tr_e \rfloor] \cup (\lfloor Tr_r \rfloor, T]$ represents the two ‘normal market periods’ at the beginning and the end of the sample.¹²

In our simulation, we use the TGARCH(1, 1) parameters from Sect. 3 and choose the bubbly trajectory parameters in Eq. (25) equal to those suggested in Phillips and Shi (2018): $y_0 = 100, T = 100, a = \eta = c_1 = c_2 = 1, \tilde{\alpha} = 0.6, r_e = 0.4, r_f = 0.6, d_B = \lfloor 0.20T \rfloor$. We consider three alternative collapse patterns: (i) sudden collapse ($\tilde{\beta} = 0.1, r_r = 0.61, d_C = \lfloor 0.01T \rfloor$), (ii) disturbing collapse ($\tilde{\beta} = 0.5, r_r = 0.7, d_C = \lfloor 0.10T \rfloor$), and (iii) smooth collapse ($\tilde{\beta} = 0.9, r_r = 0.8, d_C = \lfloor 0.20T \rfloor$). Figure 3 displays bubbly trajectories (one for each collapse pat-

¹² The (fractional) recovery date r_r is not estimated in the date-stamping analysis, but used as a mere parameter in the simulation of the bubbly trajectories.

tern), simulated with these parameters. We generate 10,000 bubbly trajectories per collapse pattern, each with the sample size $T = 100$. For PSY date-stamping, we use the 95% critical-value sequence $\{\text{scv}_{r_2}^{0.05}\}_{r_2 \in [19, 100]}$, obtained from 10,000 Monte Carlo replications. For PSY^{sign-based} date-stamping, we use the 95% critical-value sequence $\{\text{sbcv}_{r_2}^{0.05}\}_{r_2 \in [19, 100]}$, also obtained from 10,000 Monte Carlo replications. As the minimal bubbly period duration, we impose 3 points in time.

Table 7 displays our date-stamping results under the alternative collapse patterns. In the upper block, we report three bubble-detection rates ('No bubble,' 'One bubble,' '> 1 bubble'). The lower block displays the mean deviations of the estimated origination (termination) dates from their actual simulation values (along with standard errors). We recall that each of the 30,000 trajectories is simulated with exactly one bubbly period.

Our date-stamping simulation provides the following main findings. (i) Only on rare occasions the PSY date-stamping procedure does not stamp any bubble (Row 'No bubble' in Table 7), while—irrespective of the collapse pattern—this occurs more frequently under the PSY^{sign-based} procedure. (ii) Both procedures routinely stamp more than one bubble (Row '> 1 bubble'). For the PSY procedure, this rate of erroneous stamping increases from 20.33% to 39.18% with increasing $\tilde{\beta}$ values (i.e., with less abrupt collapse patterns). By contrast, this rate decreases from 24.57% to 5.88% with increasing $\tilde{\beta}$ values for the PSY^{sign-based} procedure.

(iii) In 2 of 3 settings ($\tilde{\beta} = 0.5, 0.9$), the PSY^{sign-based} procedure outperforms the PSY procedure in detecting the correct number of exactly one bubble (Row 'One bubble'). (iv) On average, both procedures overestimate the origination date (Row 'Mean ($\hat{r}_e - r_e$)').¹³ This result is stable across the distinct collapse patterns. (v) The PSY procedure exhibits an increasing bias in estimating the termination date, when it comes to less abrupt collapse patterns (Row 'Mean ($\hat{r}_f - r_f$)'). By contrast, the PSY^{sign-based} strategy exhibits a large bias in estimating the termination date in case of a sudden bubble collapse ($\tilde{\beta} = 0.1$).

Overall, the PSY procedure outperforms the PSY^{sign-based} procedure in estimating the origination and termination dates of a bubble (in terms of smaller biases) in 5 of 6 settings. However, the PSY procedure often erroneously stamps more than one bubble, especially under the realistic scenarios of disturbing and smooth bubble collapses.¹⁴

4.3 Empirical analysis

In this section, we bubble date-stamp the NASDAQ stock market. We use *Thomson Reuters Datastream*, which provides daily and monthly observations of the NASDAQ composite dividend yields. The data cover the time-span 2 January 1973–29 December 2017 ($T = 540$ monthly, $T = 11739$ daily observations). We obtain the price-dividend ratio as the inverse of the dividend yield. In order to estimate the termination dates,

¹³ In computing the 'Mean deviations' of the estimated origination and termination dates from their true values, we only used those trajectories, for which the procedures stamped exactly one bubble.

¹⁴ We also analyzed the PSY date-stamping procedure using infeasibly size-adjusted critical values. Compared to the original (non-adjusted) PSY strategy, we do not find substantial differences in bubble-date estimation accuracy. However, under the infeasibly size-adjusted PSY procedure the rate of erroneously stamping more than a single bubble is considerably lower, irrespective of the collapse pattern.

Table 7 PSY and PSY^{sign-based} date-stamping results under the bubbly Phillips-Shi (2018) process with TGARCH(1, 1) errors

	Sudden collapse ($\tilde{\beta} = 0.1$)		Disturbing collapse ($\tilde{\beta} = 0.5$)		Smooth collapse ($\tilde{\beta} = 0.9$)	
	PSY	PSY ^{sign-based}	PSY	PSY ^{sign-based}	PSY	PSY ^{sign-based}
Detection rate						
No bubble	0.0025	0.1881	0.0017	0.1003	0.0018	0.1005
One bubble	0.7942	0.5662	0.7017	0.8048	0.6064	0.8407
> 1 bubble	0.2033	0.2457	0.2966	0.0949	0.3918	0.0588
Mean deviation						
Mean ($\hat{r}_e - r_e$)	0.0278	0.0834	0.0286	0.0789	0.0296	0.0798
Standard error	(0.0291)	(0.0806)	(0.0301)	(0.0602)	(0.0313)	(0.0655)
Mean ($\hat{r}_f - r_f$)	0.0093	0.1424	0.0324	0.0487	0.1237	0.0602
Standard error	(0.0125)	(0.1381)	(0.0173)	(0.0559)	(0.0341)	(0.0610)

The trajectories are generated according to Eq. (25) with TGARCH(1,1) errors parameterized as $\omega = 0.4387, \gamma = 0, \beta = 0.9319, \phi = 0.1306$. The other parameters are $y_0 = 100, T = 100, a = \eta = c_1 = c_2 = 1, \alpha = 0.6, r_e = 0.4, r_f = 0.6, d_B = [0.20T]$, and (i) $\beta = 0.1, r_r = 0.61, d_C = [0.01T]$ for the sudden collapse, (ii) $\tilde{\beta} = 0.5, r_r = 0.7, d_C = [0.10T]$ for the disturbing collapse, and (iii) $\tilde{\beta} = 0.9, r_r = 0.8, d_C = [0.20T]$ for the smooth collapse. We use 10, 000 replications for each setting

according to Eqs. (21) and (24), the date-stamping procedures require 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).

In contrast to Sect. 4.2, the data-generating process, underlying the NASDAQ, is completely unknown. In particular, we neither know (i) the true number of bubbles (0, 1, 2, ...), nor (ii) the DGP's true heteroscedasticity pattern. A well-documented stylized empirical fact is that financial time series are typically subject to time-varying (and, most likely, overlapping) unconditional and conditional heteroscedasticity patterns of unknown form (e.g., Schwert 1989; HLST, Reher and Wilfling 2016). For example, for the NASDAQ data we may find—among other forms of heteroscedasticity—significant TGARCH effects over certain sampling periods, but not over others (see Footnote 5). In this context, we recall that the bootstrap (HLST) and sign-based (HLZ) bubble-detection tests from Sect. 3 only adjust for specific patterns of unconditional heteroscedasticity. Under (conditional) TGARCH heteroscedasticity, however, both tests may exhibit substantial power problems.

This latter issue becomes important, when it comes to the econometrician's selection of the appropriate bubble date-stamping procedure (PSY, PSY^{sign-based}, and conceivably, PSY^{bootstrap}). Since the NASDAQ-DGP is subject to heteroscedasticity of unknown form, there is, a priori, no econometric justification for why to prefer one of the three date-stamping procedures to any of the other. In our subsequent date-stamping analysis, we again discard the PSY^{bootstrap} procedure due to computational infeasibility and apply PSY and PSY^{sign-based}. A further methodological issue pertains to the econometrician's choice of the data frequency. We therefore contrast the NASDAQ date-stamping results for monthly with daily observations.

4.3.1 Monthly data

We start our date-stamping analysis with the monthly NASDAQ sample (Fig. 4). As the training period, we use the first 47 (out of 540) observations. As described in Sect. 4.1, we investigate explosive behavior in the price-dividend ratio via the intersections of the BSADF and BSADF^{sign-based} test statistics with their corresponding 95% critical values (obtained from 2,000 Monte Carlo replications). The upper panel in Fig. 4, analyzing the PSY date-stamping procedure, stamps (i) three short-lived periods of explosiveness (shorter than 6 months), indicated by vertical dashed lines, and (ii) four potential bubble periods, indicated by gray shaded areas. These four periods last from (i) June 1983–March 1984 (10 months), (ii) April 1986–November 1987 (20 months), (iii) October 1996–January 2001 (52 months), and (iv) November 2008–April 2009 (6 months).

The final three of these periods can 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 2008.¹⁵ Especially the latest period, stamped during the

¹⁵ In describing the historical events, we use the terminology 'bubbly periods' and 'dotcom bubble' in line with the empirical financial-market literature. The question of whether these episodes in fact represent 'theoretical' stock-market bubbles (such as those described in Sect. 2) is subject to controversial debate.

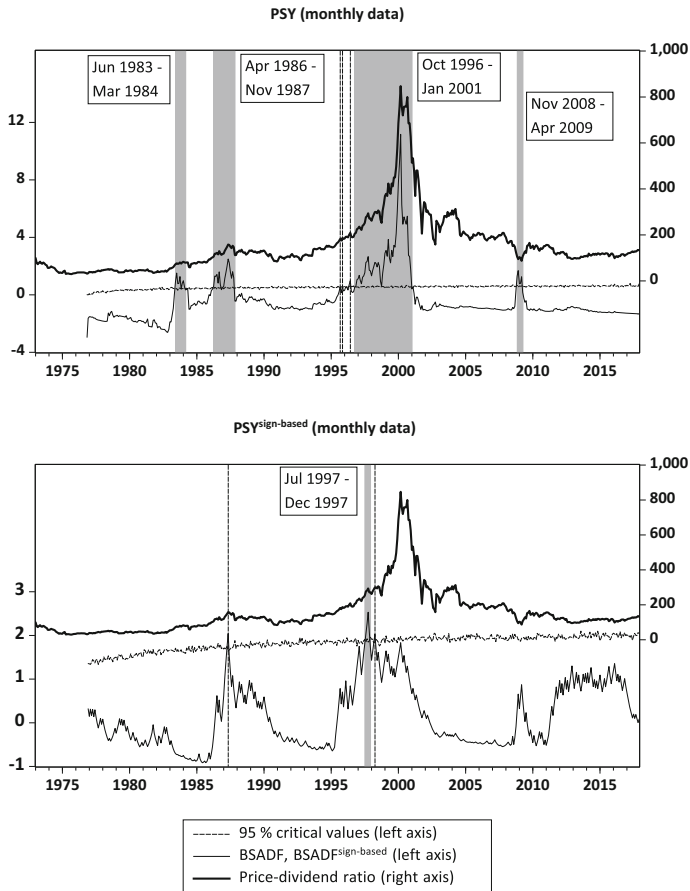


Fig. 4 PSY (upper panel) and $PSY^{\text{sign-based}}$ (lower panel) bubble date-stamping of monthly NASDAQ price-dividend ratios

subprime mortgage crisis, reflects a feature of the PSY procedure that is also reported by other authors. Sometimes the PSY procedure identifies collapse periods rather than bubble periods (see, *inter alia*, Hu and Oxley 2018b).

The lower panel of Fig. 4 displays the monthly date-stamping results for the $PSY^{\text{sign-based}}$ procedure. The procedure only detects one bubbly period (gray shaded area), lasting from July 1997–December 1997. Additionally, the procedure marks two short-lived periods of explosiveness (vertical, dashed lines) around (i) May 1987, and (ii) May 1998, the lengths of which fall below our imposed minimal bubble duration of six months. Obviously, the $PSY^{\text{sign-based}}$ procedure only slightly signals the bull market prior to the Black Monday crash in 1987, and completely fails to stamp (i) the principal part of the dotcom bubble during the years 1999 and 2000, and (ii) the market recovery after the Lehman Brothers insolvency in autumn 2008.

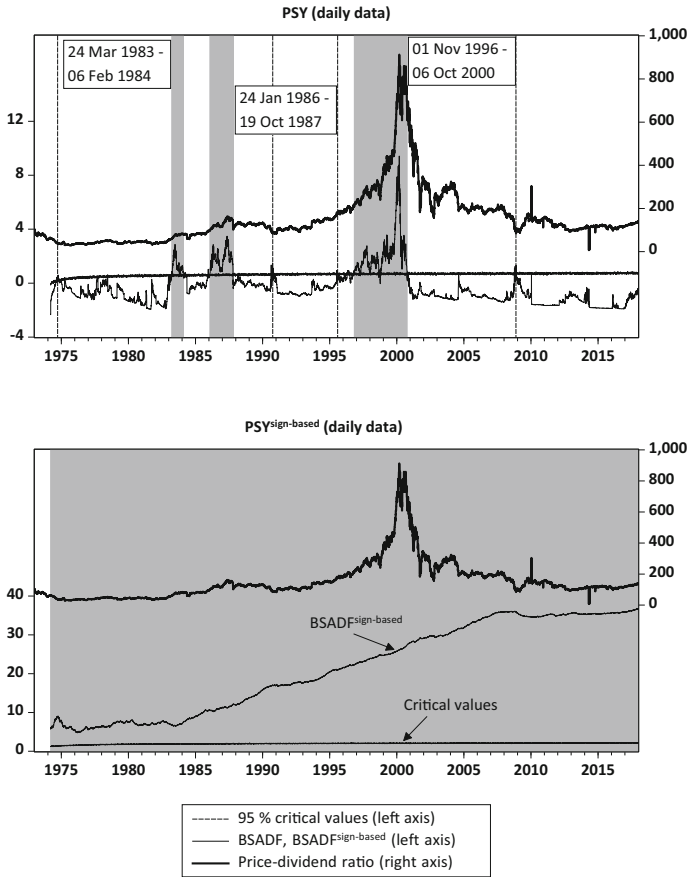


Fig. 5 PSY (upper panel) and $PSY^{sign-based}$ (lower panel) bubble date-stamping of daily NASDAQ price-dividend ratios

4.3.2 Daily data

In Fig. 5, we analyze the performance of the two date-stamping procedures for the daily NASDAQ observations, where our training period consists of the first 312 (out of 11739) observations. The PSY date-stamping procedure in the upper panel again identifies a number of short-lived periods of explosiveness. However, since we impose a minimal bubble duration of 180 days ($\delta = 44.2$), the PSY procedure now only identifies three bubbly periods under the daily data. In contrast to monthly data, the collapse period during the subprime mortgage crises now remains unstamped.

Evidently, the PSY-stamped bubbles under daily observations can differ considerably in the origination and/or termination dates from their analogs under monthly data. Concretely, the first bubble in the upper panel of Fig. 5 already starts on 24 March 1983 under daily data, but not before June 1983 under monthly data in the upper panel of Fig. 4 (i.e., nearly 2.5 months later). Similarly, the second bubble under daily data

starts on 24 January 1986, but—under monthly data—not before April 1986. On the other hand, the dotcom bubble ends in January 2001 under monthly observations, but already on 6 October 2000, when using daily data. Thus, for the dotcom bubble, the different estimates of the termination and origination dates imply a 5-month-longer-lasting bubble under monthly observations.

The daily $\text{PSY}^{\text{sign-based}}$ NASDAQ date-stamping is displayed in the lower panel of Fig. 5. Obviously, the $\text{BSADF}^{\text{sign-based}}$ test statistics exceed the critical values over the entire sample, signaling a permanent NASDAQ bubble over 44 years. This clear-cut result for the $\text{PSY}^{\text{sign-based}}$ procedure again documents an apparent sensitivity of the date-stamping procedures to data frequency shifts. A frequency shift is typically associated with differing econometric properties of the underlying DGP, for example in terms of altering (i) degrees-of-explosiveness (over certain subsamples), and/or (ii) (local) heteroscedasticity patterns. The differing DGP properties can then have the potential to induce divergent date-stamping results.

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 Figs. 4 and 5, we mark a number of short-lived periods of explosiveness not recognized as bubbles, due to our arbitrary setting of $\delta = 2.2$ and $\delta = 44.2$ for monthly/daily observations. The practitioner will frequently be confronted with the judgmental and subjective question of whether a (relatively) 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 artifact.¹⁶

5 Conclusion

This paper investigates the performance of SADF, GSADF, and several heteroscedasticity-adjusted sup-ADF-style tests in detecting financial bubbles. We address (i) the empirical size of the 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 and Wilfling (2018). Our Monte Carlo simulations find that the majority of the tests exhibit substantial size distortions when the data-generating process is subject to leverage effects.¹⁷ Moreover, the sup-ADF-style tests often have low empirical power in identifying the (flexible) Rotermann–Wilfling (2018) bubble. As shown in the Panels (a) and (b) of Fig. 2, in 21 of our 24 scenarios (= 87.5%), more than 40% of the existing bubbles remain undetected by the sup-ADF-style tests. In addition, we investigate the performance of two real-time bubble date-stamping procedures (PSY and $\text{PSY}^{\text{sign-based}}$) in a Monte Carlo simulation. While the PSY procedure outperforms the $\text{PSY}^{\text{sign-based}}$ strategy in terms of bubble-date estimation accuracy, it frequently stamps non-existing bubbles. Finally, we apply both date-stamping procedures to monthly and daily NASDAQ data over a

¹⁶ This phenomenon particularly applies to the bootstrap date-stamping procedure suggested by Phillips and Shi (2020). See Footnote 11 above.

¹⁷ Fulop and Yu (2017) propose a (Bayesian) two-regime Markov-switching bubble framework, which appears to exhibit some robustness in bubble detection in the presence of leverage effects.

period of 44 years. We find that the date-stamping results can be highly sensitive to the chosen data frequency.

We emphasize that it is by no means our intention to discredit any of the sup-ADF-style testing procedures. The probabilistic background of the methodologies, designed to test for (statistical) explosiveness, is mathematically rigorous and constitutes an interesting, rapidly evolving strand of the econometric literature (Harvey et al. 2019; Kurozumi 2020). However, when it comes to financial bubble detection, our advice is to apply the routines with caution.¹⁸ At the end of Sect. 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 never will 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.

Acknowledgements Open Access funding provided by Projekt DEAL. We are grateful to Robert Kunst and two anonymous reviewers for their constructive and extensive comments, which greatly improved the paper. The usual disclaimer applies.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Human and animal rights This article does not contain any studies with animals or human beings performed by any of the authors.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bettendorf T, Chen W (2013) Are there bubbles in the Sterling-dollar exchange rate? New evidence from sequential ADF tests. *Econ Lett* 120:350–353
- Black F (1976) Studies on stock price volatility changes. In: Proceedings of the 1976 meetings of the business and economics statistics section. American Statistical Association, Washington, DC, pp 177–181

¹⁸ See Hu and Oxley (2018a), and Phillips and Shi (2018) for a similar recommendation.

- Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. *J Econom* 31:307–328
- Campbell JY, Lo AW, MacKinlay AC (1997) *The econometrics of financial markets*. Princeton University Press, Princeton
- Caspi I (2017) Rtdaf: testing for bubbles with eviews. *J Stat Softw* 81:1–16
- Christie AA (1982) The stochastic behavior of common stock variances: value, leverage and interest rate effects. *J Financ Econ* 10:407–432
- Cochrane JH (2011) Presidential address: discount rates. *J Finance* 66:1047–1108
- Diba B, Grossman H (1988a) Explosive rational bubbles in stock prices? *Am Econ Rev* 78:520–530
- Diba B, Grossman H (1988b) The theory of rational bubbles in stock prices. *Econ J* 98:746–754
- Evans G (1991) Pitfalls in testing for explosive bubbles in asset prices. *Am Econ Rev* 81:922–30
- Fulop A, Yu J (2017) Bayesian analysis of bubbles in asset prices. *Econometrics* 4:1–23 (Article 47)
- Harvey DI, Leybourne SJ, Sollis R, Taylor AMR (2016) Tests for explosive financial bubbles in the presence of non-stationary volatility. *J Empir Finance* 38:548–574
- Harvey DI, Leybourne SJ, Zu Y (2019) Testing explosive bubbles with time-varying volatility. *Econom Rev* 38:1131–1151
- Harvey DI, Leybourne SJ, Zu Y (2020) Sign-based unit root tests for explosive financial bubbles in the presence of deterministically time-varying volatility. *Econom Theory* 36:122–169
- Homm U, Breitung J (2012) Testing for speculative bubbles in stock markets: a comparison of alternative methods. *J Financ Econ* 10:198–231
- Hu Y, Oxley L (2017) Are there bubbles in exchange rates? Some new evidence from G10 and emerging market economies. *Econ Model* 64:419–442
- Hu Y, Oxley L (2018a) Bubbles in US regional house prices: evidence from house price-income ratios at the state level. *Appl Econ* 50:3196–3229
- Hu Y, Oxley L (2018b) Do 18th century ‘bubbles’ survive the scrutiny of 21st century time series econometrics? *Econ Lett* 162:131–134
- Kurozumi E (2020) Asymptotic properties of bubble monitoring tests. *Econom Rev* 39:510–538
- Long W, Li D, Li Q (2016) Testing explosive behavior in the gold market. *Empir Econ* 51:1151–1164
- Pan WF (2019) Detecting bubbles in China’s regional housing markets. *Empir Econ* 56:1413–1432
- Phillips PCB, Shi S (2018) Financial bubble implosion and reverse regression. *Econom Theory* 34:705–753
- Phillips PCB, Shi S (2020) Real time monitoring of asset markets: bubbles and crises. *Handb Stat*. <https://doi.org/10.1016/bs.host.2018.12.002>
- Phillips PCB, Wu Y, Yu J (2011) Explosive behavior in the 1990s NASDAQ: when did exuberance escalate asset values? *Int Econ Rev* 52:201–226
- Phillips PCB, Shi S, Yu J (2014) Specification sensitivity in right-tailed unit root testing for explosive behaviour. *Oxf Bull Econ Stat* 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. *Int Econ Rev* 56:1043–1078
- Phillips PCB, Shi S, Yu J (2015b) Testing for multiple bubbles: limit theory of real-time detectors. *Int Econ Rev* 56:1043–1078
- Reher G, Wilfling B (2016) A nesting framework for Markov-switching GARCH modelling with an application to the German stock market. *Quant Finance* 16:411–426
- Rotermann B, Wilfling B (2014) Periodically collapsing Evans bubbles and stock-price volatility. *Econ Lett* 123:383–386
- Rotermann B, Wilfling B (2018) A new stock-price bubble with stochastically deflating trajectories. *Appl Econ Lett* 25:1091–1096
- Schwert GW (1989) Why does stock market volatility change over time? *J Finance* 44:1115–1153
- Zakoian JM (1994) Threshold heteroscedastic models. *J Econ Dyn Control* 18:931–955

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.