Sequential Detector Statistics for Speculative Bubbles

Jörg Breitung* and Max Diegel

University of Cologne

May 2025

Abstract

We propose a heteroskedasticity-robust locally best invariant (LBI) statistic to test the hypothesis of a unit root against the alternative of an explosive root associated with speculative bubbles. Compared to existing alternatives such as Dickey-Fuller type tests, the LBI statistic has a standard limiting distribution and greater power, particularly in the empirically relevant scenario of a moderately explosive root. Further refinements, such as the point-optimal linear test, approach the power envelope remarkably closely. To detect bubbles with an unknown starting date, we consider sequential (CUSUM) schemes based on constant and time-varying boundary functions, where the exponentially weighted CUSUM detector with constant boundary function turns out to be most powerful. We also propose a simple method for date-stamping the start of the bubble consistently. Finally, we illustrate our methods using two empirical examples.

*Corresponding author: Jörg Breitung: University of Cologne, Institute of Econometrics, Albertus Magnus Platz, 50923 Köln, Germany, email: breitung@statistik.uni-koeln.de., Replication Files are available on Github at `https://github.com/maxcd/SequentialBubbleDetectors`

1 Introduction

Since the publication of the seminal paper by Phillips et al. (2011), testing for speculative bubbles has gained significant interest in financial analysis. Speculative bubbles, characterized by explosive price movements, have spurred the development of statistical tests to detect such patterns. In quasi-efficient asset markets, prices follow a martingale process, implying that asset prices resemble a (geometric) random walk in the fundamental regime. According to the present value model, prices are governed by a difference equation that does not uniquely determine the price *level*. This indeterminacy permits the addition of a bubble component to the fundamental price without violating the dynamics of the model. Although a transversality condition rules out the existence of such bubble components in the long term, speculative bubbles can emerge and dissipate in the short term.

Existing tests for speculative bubbles aim to detect a shift from a random walk to an explosive autoregressive process at an unknown point in time. Phillips et al. (2011, 2015) identify this structural break in the time series by using sequential Dickey-Fuller (DF) tests applied to an expanding window of observations. Alternatively, Homm and Breitung (2012) and Astill et al. (2017) consider *CUSUM* schemes to detect a potential change to an explosive process. In this paper, we argue that the DF statistic is suboptimal and may be replaced by simpler alternatives like the locally best invariant (*LBI*) statistic, which is also robust against potential heteroskedasticity. While the *LBI* statistic allows for further refinements, such as best linear point-optimal tests, these yield only marginal improvements, as the power of the *LBI* statistic is already remarkably close to the envelope.

The *LBI* detector is closely related to the *CUSUM* approach for sequential testing and monitoring. Whereas the original *CUSUM* approach applies a time-varying boundary function, we argue that applying a constant boundary results in a more powerful test that is robust to heteroskedasticity. The power of the test can be further improved by an exponential weighting scheme. Monte Carlo simulations demonstrate significant improvements over existing methods based on the DF statistic. Furthermore, we adapt the (weighted) *LBI* detector for real-time monitoring and propose a simple and consistent estimator for the starting date of the bubble, often referred to as "date-stamping".

The paper is organized as follows. In Section 2 we outline the testing framework that builds on standard models of speculative bubbles. The *LBI* detector and its refinements are considered in Section 3. In Section 4.1 we discuss sequential testing schemes that can be used to test for speculative bubbles with an unknown starting date. The adaptation of the sequential tests for real-time monitoring is discussed in Section 4.2, while Section

4.3 examines properties of existing estimators for date-stamping and proposes our new estimator. The small sample properties of these methods are studied by means of Monte Carlo experiments in Section 5 and two empirical applications are considered in Section 6. Finally, Section 7 concludes.

2 Framework for testing against explosive alternatives

Following the work of Phillips et al. (2011), Homm and Breitung (2012), Astill et al. (2017) and many others, we consider an AR(1) process of the form:

$$P_t = d_t + y_t \tag{1}$$

$$y_t = \begin{cases} y_{t-1} + u_t & \text{for } t = 1, 2, \ldots, [r_e T] \\ \rho y_{t-1} + u_t & \text{for } t = [r_e T] + 1, \ldots, T \end{cases} \tag{2}$$

where P_t denotes the asset price that consists of a deterministic component d_t and an autoregressive component y_t with innovation u_t. Under the null hypothesis of no bubble, we assume that the autoregressive parameter satisfies $\rho = 1$ and r_e is the fraction of the sample generated under the null hypothesis. Under the alternative, the price exhibits explosive dynamics due to $\rho > 1$. Let $[a]$ denote the integer part of a such that $T_e = [r_e T]$ is the last period before the bubble emerges.

It should be noted that (2) does not directly correspond to a typical price process with a speculative bubble which is given by

$$P_t = P_t^f + B_t \tag{3}$$

where P_t^f denotes the fundamental price at period t. P_t^f is typically represented by a (geometric) random walk (possibly with drift), and $B_t = \varrho B_{t-1} + \epsilon_t$ represents the explosive bubble component. Here, the exponential rate $\varrho = (1 + r)/\pi$ is driven by the return of an alternative investment r and the probability that the bubble continues, denoted π, see Blanchard and Watson (1982). As shown by Breitung and Kruse (2013) the process in (3) has an ARIMA(1,1,1) representation which can be well approximated by an explosive AR(1) process. The approximation error only affects the power of the test whereas the statistical properties under the null hypothesis of no bubble remain intact.

The efficient market hypothesis implies that during the fundamental regime, the (logarithm of) prices can be characterized by a martingale process obeying $\mathbb{E}(P_{t+1}|P_t, P_{t-1}, \ldots) =$

$\beta + P_t$, where β is the mean return compensating for opportunity costs and the risk of the investment. This implies that $P_t = \mu + \beta t + y_t$, where y_t can be represented by a pure random walk. In empirical applications the mean return β is typically small. Assume for example an annual mean return of 7 percent. This implies that for monthly logarithmic data the constant is $\beta = 0.0057$. Accordingly, the drift β is often found to be statistically insignificant and setting this parameter to zero does not result in severe size distortions of the bubble test. Phillips et al. (2014) considers an "asymptotically negligible drift" defined as $\kappa T^{-\eta}$ with $\eta > 0.5$ and κ is some positive constant. We therefore focus on the case with a constant mean and refer to the working paper version (Breitung and Diegel 2024) for the treatment of a model with a linear time trend.

We make the following assumption for the series y_t:

Assumption 1: *Under the null hypothesis* $y_t = y_{t-1} + u_t$ *for* $t = 1, \ldots, T$ *and* $y_0 = o_p(T^{1/2})$, *where* u_t *is a martingale difference process with* $\mathbb{E}(u_t | u_{t-1}, u_{t-2}, \ldots) = 0$, $\sigma^2 = \lim_{T\to\infty} \mathbb{E}(T^{-1} y_T^2)$ *exists, and* $\mathbb{E}\,|u_t|^{2+\eta} < \infty$ *for some* $\eta > 0$ *and all* t.

Note that this assumption allows u_t to be heteroskedastic which is an important feature of asset returns. As argued in the working paper version it is possible to allow for (weak) autocorrelation in u_t. Empirically, however, the autocorrelation in asset returns is typically negligible.

3 Tests with known starting date of the bubble

In this section, we consider a scenario where the bubble is known to start at the beginning of the sample, i.e. $r_e = 0$ (resp. $T_e = 0$). Accordingly, we are interested in testing the hypothesis $\rho = 1$ for the the entire sample $t = 1, 2, \ldots, T$, against the alternative $\rho > 1$. For ease of exposition, we start by assuming $\mathbb{E}(P_t) = \mathbb{E}(y_t) = 0$, i.e. $d_t = 0$ in (1), with initial value $y_0 = 0$. These assumptions will be relaxed below. The standard testing procedures for speculative bubbles such as Phillips et al. (2011) are based on the DF statistic

$$\tau_T = \frac{\sum_{t=1}^{T} y_{t-1} \Delta y_t}{\widehat{\sigma} \sqrt{\sum_{t=1}^{T} y_{t-1}^2}},$$

where $\widehat{\sigma}^2$ denotes the usual variance estimator of the residuals $\widehat{u}_t = y_t - \widehat{\rho} y_{t-1}$ and $\widehat{\rho}$ is the OLS estimator of ρ.

An important drawback of this test statistic is that under the alternative of an explosive process the denominator diverges towards infinity at a rate higher than T, which may result in a loss of power. It is therefore desirable to replace the denominator $\sum_{t=1}^{T} y_{t-1}^2$ by its expectation $\mathbb{E}\left(\sum_{t=1}^{T} y_{t-1}^2\right) \simeq \sigma^2 T^2/2$. Furthermore, it is well known that

$$2\sum_{t=1}^{T} \Delta y_t y_{t-1} = y_T^2 - T\widetilde{\sigma}^2 \tag{4}$$

where $\widetilde{\sigma}^2 = T^{-1}\sum_{t=1}^{T}(\Delta y_t)^2$ is the estimator of the residual variance under the null hypothesis. Replacing σ^2 and $\widehat{\sigma}^2$ by the estimate $\widetilde{\sigma}^2$ yields the simple test statistic

$$LBI_T^2 = \frac{y_T^2}{\widetilde{\sigma}^2 T} = \frac{\left(\sum_{t=1}^{T} \Delta y_t\right)^2}{\widetilde{\sigma}^2 T} \tag{5}$$

where we ignore the irrelevant factor and the additional constant that do not affect the power of the test. Note that this statistic is equivalent to the *locally best invariant statistic* (*LBI*) for testing the unit root hypothesis against stationary alternatives.[1]

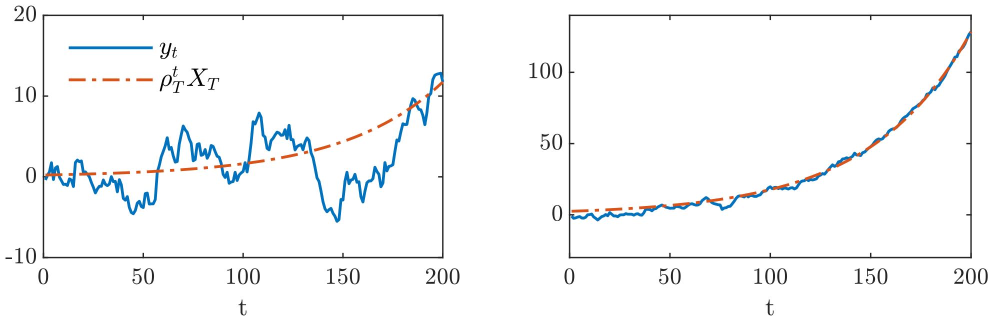
It is important to note that the *LBI* statistic is invariant with respect to heteroskedasticity of u_t, see Cavaliere (2005). This is an important advantage when applied to financial time series such as stock prices or exchange rates. Another advantage of the *LBI* statistic is that it is able to accommodate one-sided alternatives such as *positive* bubbles only. If $\rho > 1$ then the series may tend to plus or minus infinity. On financial markets it is implausible to allow for negative bubbles as it does not make sense to assume that investors are willing to buy a risky asset with negative expected return. By applying the DF statistic, it is however not possible to distinguish between positive or negative bubbles. Using the *LBI* statistic we are able to focus on positive bubbles by applying a right-sided critical value to $LBI_T = \left(\widetilde{\sigma}\sqrt{T}\right)^{-1}\sum_{t=1}^{T}\Delta y_t$. We present the limiting distribution for local alternatives of the form $\rho_T = 1 + c/T$, where c is some fixed constant in the working paper version, cf. Breitung and Diegel (2024). Under the null hypothesis with $c = 0$ the LBI_T statistic has a standard normal limiting distribution.

Following Phillips and Magdalinos (2007) it has become popular to consider *moderately* explosive processes that can be represented by letting

$$\rho_T = 1 + c/T^\theta \quad \text{with } 0 < \theta < 1. \tag{6}$$

[1]The LBI_T^2 statistic is equivalent to the LM (score) test statistic based on the expected second derivative of the log-likelihood (Fisher information), see e.g. Solo (1984), Phillips and Schmidt (1989), and Schmidt and Lee (1991).

Figure 1: Decomposition of two selected bubble realizations



This sequence of coefficients is considered to better represent explosive episodes on financial markets. Under this local alternative, the limiting behavior of the process can be represented as in Lemma 1.

Lemma 1 *Let y_t be generated by the AR(1) process in Equation* (2)*, where ρ_T is given by* (6) *and u_t satisfies Assumption 1. If $y_0 = o_p(T^{\theta/2})$ we have*

$$y_t = \rho_T^t X_T + O_p(T^{\theta/2})$$

where $T^{-\theta/2} X_T = T^{-\theta/2} \sum_{t=1}^{T} \rho_T^{-t} u_t \xrightarrow{d} \mathcal{N}(0, \sigma^2/(2c))$.

This representation highlights important properties of moderately explosive processes. The first part $\rho_T^t X_T$ evolves like a deterministic explosive process with stochastic initial value X_T. Due to ρ_T^t this component dominates as t gets large. The remaining component is of similar order of magnitude as X_T and behaves like a (nearly) stationary component. If X_T is sizable and t increases the exponential component eventually dominates. However, for small values of X_T, the bubble process can be hard to distinguish from a nearly stationary process for finite T.

Figure 1 illustrates this feature for $\rho = 1.02$ with a "weak" bubble realization due to a small positive X_T in the left panel and a "strong" bubble with a large positive realization of X_T in the right. The weak bubble is hardly visible even after 200 time periods. In contrast, for a large positive X_T, the series is clearly dominated by the bubble component and follows an exponential path.

Moreover, if X_T is positive, then the process diverges towards plus infinity (positive bubble), whereas for negative X_T the process collapses towards minus infinity (negative bubble). One may argue that negative bubbles can be ruled out, as it does not make sense to invest in an asset that follows an exponential path toward zero. But as short selling strategies become more and more available, investing in a negative bubble may become feasible. There is, however, an important difference to the case of a positive bubble. While the growth of a positive bubble is not limited and may grow to 200 or 300 percent of the fundamental value, a negative bubble has a natural terminal value of zero and is therefore limited to -100 percent. Accordingly, it is less likely to encounter negative bubbles on financial markets.

3.1 Point optimal linear detectors

The *LBI* statistic is based on the (unweighted) sum of first differences $\sum_{t=1}^{T} \Delta y_t$. Now, let us consider a specific alternative with $\rho = \bar{\rho} > 1$. From Lemma 1 we conclude that for large X_T we obtain the representation

$$\begin{aligned} \Delta y_t &= (\bar{\rho}^t - \bar{\rho}^{t-1}) X_T + e_t \\ &= \bar{\rho}^t \delta + e_t \end{aligned}$$

where $\delta = X_T(\bar{\rho} - 1)/\bar{\rho}$ and e_t is $O_p(T^{\theta/2})$, see Lemma 1. The null hypothesis implies $\delta = 0$. To test the null hypothesis we may use the t-statistic of the least-squares estimator of δ which can be written as

$$\tau_\delta = \frac{1}{\kappa(\bar{\rho})} \sum_{t=1}^{T} \rho^t \Delta y_t \tag{7}$$

where $\kappa(\bar{\rho})^2 = \sigma^2 \sum_{t=1}^{T} \bar{\rho}^{2t}$. This suggest that the power of the *LBI* statistic can be improved by exponentially weighting the differences with weights w_t that are proportional to ρ^t such that

$$\phi_T = \frac{1}{\sigma} \sum_{t=1}^{T} w_t \Delta y_t \quad \text{with} \quad \sum_{t=1}^{T} w_t^2 \ = \ 1, \tag{8}$$

Note that we have normalized the weights such that $w_t = O(T^{-1/2})$ and, therefore, the factor $1/\sqrt{T}$ is absorbed in the weight function. Furthermore, the exponential weighting scheme assigns increasing importance to the later part of the sample, where the bubble is more pronounced.

In order to derive an optimal weighting function for some prespecified alternative with

$\bar{\rho} > 1$, define

$$\boldsymbol{R}_{\bar{\rho}} = \begin{pmatrix} 1 & 0 & 0 & \cdot & 0 & 0 \\ \bar{\rho} & 1 & 0 & \cdot & 0 & 0 \\ \bar{\rho}^2 & \bar{\rho} & 1 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \bar{\rho}^{T-1} & \bar{\rho}^{T-2} & \bar{\rho}^{T-3} & \cdot & 1 & 0 \\ \bar{\rho}^{T} & \bar{\rho}^{T-1} & \bar{\rho}^{T-2} & \cdot & \bar{\rho} & 1 \end{pmatrix} \quad \text{and} \quad \boldsymbol{D} = \begin{pmatrix} -1 & 1 & 0 & \cdot & 0 & 0 \\ 0 & -1 & 1 & \cdot & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & 1 & 0 \\ 0 & 0 & 0 & \cdot & -1 & 1 \end{pmatrix}.$$

It follows that under the alternative

$$\mathbb{E}(\phi_T^2) = \boldsymbol{w}'\boldsymbol{\Psi}_{\bar{\rho}}\boldsymbol{w} \quad \text{with} \quad \boldsymbol{\Psi}_{\bar{\rho}} = \boldsymbol{D}\boldsymbol{R}_{\bar{\rho}}\boldsymbol{R}_{\bar{\rho}}'\boldsymbol{D}'$$

where $\boldsymbol{w} = (w_1, w_2, \dots, w_T)'$. The (two-sided) point-optimal linear detector statistic maximizes the rejection probability

$$\text{Prob}\left(\phi_T^2 > cv_\alpha^\chi\right) = \text{Prob}\left(\frac{\phi_T^2}{\mathbb{E}(\phi_T^2)} > cv_\alpha^\chi \frac{1}{\boldsymbol{w}'\boldsymbol{\Psi}_{\bar{\rho}}\boldsymbol{w}}\right), \tag{9}$$

where cv_α^χ is the critical value of the χ^2 distribution. It follows that the optimal weight vector is the eigenvector associated with the largest eigenvalue of the matrix $\boldsymbol{\Psi}_{\bar{\rho}}$.

To get an idea of the optimal weight vector, Figure 2 plots the weights for $T = 100$ with $\bar{\rho} = 1.001$, $\bar{\rho} = 1.02$, $\bar{\rho} = 1.04$ and $\bar{\rho} = 1.06$. Two observations stand out. First, the optimal weights approach the *LBI* detector with a flat weight function as the alternative approaches the null hypothesis $\rho = 1$. The farther the alternative moves away from the null hypothesis, the more weight is assigned to the end of the series.

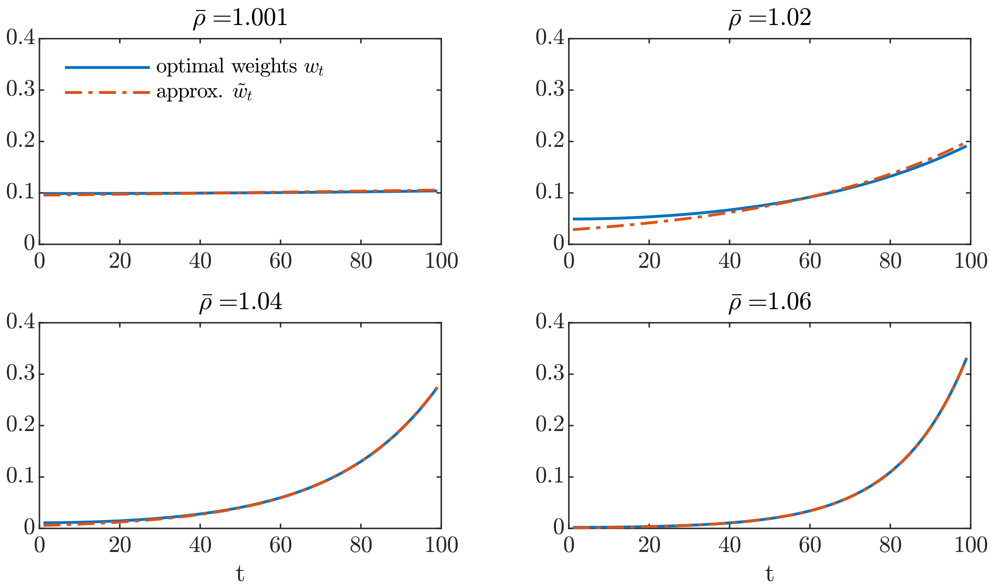
Second, the weight function is well approximated by the sequence

$$\widetilde{w}_t = \frac{\bar{\rho}^t}{\sqrt{\sum_{t=1}^T \bar{\rho}^{2t}}} \tag{10}$$

that is motivated by (7). The approximation improves as ρ increases.

It should be noted that applying the (functional) central limit theorem requires that $\sqrt{T}\,\widetilde{w}_t$ is bounded away from zero for all t. To ensure this, we consider a sequence of weights derived from $\bar{\rho}_T = 1 + \bar{c}/T \simeq e^{\bar{c}/T}$. The corresponding weighted statistic has the limiting

Figure 2: Approximation to optimal weights for different values of $\bar{\rho}$



Notes: The optimal weights w_t are obtained as the eigenvector corresponding to the largest eigenvalue of $\mathbf{\Psi}_{\bar{\rho}}$*. Approximate optimal weights as in* (10)*.*

distribution

$$\phi_T(\bar{c}) = \frac{1}{\sigma} \sum_{t=1}^{T} w^{\bar{c}}_{[t/T]} \Delta \tilde{y}_t \ \xrightarrow{d} \ \mathcal{N}(0,1) \tag{11}$$

with the weights

$$w^{\bar{c}}_r = \frac{\sqrt{2\bar{c}/T}}{\sqrt{e^{2\bar{c}} - 1}}\, e^{\bar{c}r} \ . \tag{12}$$

3.2 Power envelope

So far we considered the point-optimal *linear* detector (denoted *PO-lin*) that ignores the information in the squares of the series. Assuming Gaussian errors we can derive the more general point-optimal test statistic from the Neyman-Pearson lemma. Assuming Gaussian errors and using (4), the point-optimal test statistic against the alternative $\rho = \bar{\rho} > 1$ is given by

$$\phi^*(\bar\rho) = \frac{2}{\sigma^2 T}\sum_{t=1}^{T}\Delta y_t y_{t-1} - \frac{(\bar\rho - 1)}{\sigma^2 T}\sum_{t=1}^{T} y_{t-1}^2$$
$$= \frac{1}{\sigma^2 T} y_T^2 - \frac{\widetilde\sigma^2}{\sigma^2} - \frac{\bar c}{\sigma^2 T^2}\sum_{t=1}^{T} y_{t-1}^2 \qquad (13)$$

where $\bar c = T(\bar\rho - 1)$, see also Elliott et al. (1996) for the point-optimal test against stationary alternatives. Accordingly, for $\bar c \to 0$ the point-optimal test approaches the *LBI* detector. Setting the true parameter equal to $\bar\rho$ enables us to calculate the power envelope, see Elliott et al. (1996). In Section 5 we compare the point-optimal linear test statistics to the power envelope and find that the loss of power due to leaving out the quadratic part is marginal and therefore we focus on (weighted) linear detectors in the following.

3.3 Extensions

In empirical applications using financial data our model assumptions are often too restrictive. Specifically, the assumptions $y_0 = 0$ and $\mathbb{E}(y_t) = 0$ are unrealistic because risk averse investors expect some positive return for any risky asset. Under the null hypothesis we can relax these assumptions by allowing for $y_0 \neq 0$ and $\mathbb{E}(y_t) \neq 0$. These two assumptions are connected as the mean and starting value are not separately identified. This is due to the fact that under the null hypothesis $P_t = (\mu + y_0) + \sum_{i=1}^{t} u_i$. Note that our *LBI* and *PO-lin* tests rely on the (weighted) sum of the differences and, therefore, these detectors are invariant to μ and y_0 under the null hypothesis. Under the alternative, however, we show in the working paper version (Breitung and Diegel 2024) that a starting value with $y_0 > 0$ has a positive effect on the power of the test which increases with the duration of the bubble.

If price returns have a constant return on investment, then (log-transformed) prices follow a linear trend. Hence, it is important to generalize the analysis to account for a linear time trend. In the working paper version we suggest to apply the weighted version of the test statistic (8) to the mean adjusted differences $\Delta\widetilde y_t = \Delta P_t - \overline{\Delta P}$, where $\overline{\Delta P} = T^{-1}\sum_{i=1}^{T}\Delta P_i$. A heteroskedasticity robust version of the test is obtained by replacing the variance σ^2 by the estimator

$$\widetilde\sigma_w^2 = \sum_{t=1}^{T}(\widetilde w_t\,\Delta\widetilde y_t - \widehat\mu_w)^2 \quad \text{with } \widehat\mu_w \;=\; \frac{1}{T}\sum_{t=1}^{T}\widetilde w_t \Delta y_t\ . \qquad (14)$$

Note that under the null hypothesis it is not necessary to subtract the mean $\widehat\mu_w$ but it improves the power under the alternative.

4 Sequential testing schemes for unknown break dates

As the break date is typically unknown, a sequential testing scheme is required. In this section we therefore consider sequential versions of the (weighted) *LBI* detector. Another popular sequential procedure is that of Phillips et al. (2011) (in short PWY). They proposed a supremum Dickey-Fuller test (henceforth: *supADF*) based on the maximum of a sequence of *ADF* statistics with an expanding window $\mathcal{W}_r = \{y_1, y_2, \dots, y_{[rT]}\}$ with $\in [r_0, 1]$ and $0 < r_0 < 1$ indicates minimum relative window size. This sequential testing scheme is related to the fluctuation test for structural breaks, whereas our *LBI* detector corresponds to the *CUSUM* test, see Homm and Breitung (2012).

4.1 Sequential tests based on the *LBI* detector

To obtain a sequential *LBI* detector, we adopt the classical *CUSUM* approach. To this end we normalize the time index as a fraction of the full sample $r = t/T$, such that $r \in [0, 1]$. The sequential version of the *LBI* detector is based on the normalized partial sum:

$$LBI_{[rT]} = \frac{1}{\sigma\sqrt{T}} \sum_{t=1}^{[rT]} \Delta y_t \;\Rightarrow W(r) \ , \tag{15}$$

where $W(r)$ denotes a standard Brownian motion defined on $[0, 1]$. Brown et al. (1975) suggest a (two-sided) linear boundary function given by $b_\alpha(r) = \pm\gamma_{\alpha/2}(1 + 2r)$, where $\gamma_{\alpha/2}$ is a factor depending on the significance level α. An important advantage of the *CUSUM* detector is that it is able to detect positive bubbles by applying the upper boundary only, whereas the *supADF* test is not able to distinguish between positive and negative bubbles.

Another possibility is to apply a constant boundary function $\bar{b}_\alpha$. Although a constant boundary function is inappropriate for stationary alternatives, as it favors breaks at the end of the sample, this is a desirable feature in the case of detecting explosive alternatives. The reason is that under the explosive alternative the detector tends to be largest at the end of the sample. Sequential test with a constant boundary are indicated by *mCUSUM* (for maximum CUSUM).

As argued in Section 3.1 the power of the (linear) detector may be optimized for a given

Table 1: One-sided asymptotic critical for sequential bubble detectors

	10%	5%	2.5%	1%	0.5%
CUSUM (γ_α)	0.74	0.85	0.95	1.06	1.14
wCUSUM ($b_\alpha^{\bar{c}}$)	1.64	1.95	2.24	2.57	2.80

Notes: Critical values are upper tail quantiles based on 1,000,000 Monte Carlo replications with simulated times series of length $T = 10,000$. The critical values do not depend on the value $\bar{c}$. The critical values for the mCUSUM test are obtained by setting $\bar{c} = 0$ and, therefore, the critical values of the wCUSUM test apply also for the mCUSUM test.

alternative $\bar{\rho} = 1 + \bar{c}/T > 1$ by using the weighted statistic

$$\phi_{[rT]}(\bar{c}) = \frac{1}{\sigma} \sum_{t=1}^{[rT]} w_{[t/T]}^{\bar{c}} \Delta y_t \;\Rightarrow\; \frac{\sqrt{2\bar{c}}}{\sqrt{e^{2\bar{c}} - 1}} \int_0^r e^{\bar{c}a} dW(a)$$

see (11). The null hypothesis is rejected if $\phi_{[rT]}(\bar{c}) > b_\alpha^{\bar{c}}$ for some $r \in (0, 1]$, where $b_\alpha^{\bar{c}}$ is the critical value for the significance level α.

Note that $\int_0^r e^{\bar{c}a} dW(a)$ can be seen as a heteroskedastic Brownian motion where the variance of the increment is proportional to $e^{2\bar{c}a}$. As shown by Cavaliere and Taylor (2007) any heteroskedastic unit root process can be represented as a time-transformed Brownian motion. Following that idea, our test statistic can be asymptotically represented as a standard Brownian motion $W^*(r)$ defined on $r \in [0, 1]$:

$$\phi_r^{\bar{c}} \Rightarrow W^*(\eta(r)), \quad \text{where} \quad \eta(r) = \frac{\int_0^r e^{2\bar{c}a} da}{\int_0^1 e^{2\bar{c}a} da}$$

denotes the variance profile. Since $\sup_{r \in [0,1]} \{W^*(\eta(r))\} = \sup_{r \in [0,1]} \{W(r)\}$, it follows that the critical values do not depend on $\bar{c}$ and are identical to the *mCUSUM* test (which results from setting $\bar{c} = 0$).

In the spirit of Elliott et al. (1996), we calibrate the weights for the weighting scheme in (10) such that the test achieves a power of approximately 50% for $\rho = \bar{\rho}$. This results in $\bar{c} = 2$ which implies $\bar{\rho} = 1.02$ asymptotically. The asymptotic critical values for this weighted CUSUM test (henceforth *wCUSUM*) are provided in Table 1. Note that applying a heteroskedasticity robust variance estimator as in (12) makes the *mCUSUM* and *wCUSUM* detectors robust against heteroskedasticity.

In the working paper version (Breitung and Diegel 2024), we show that *backward CUSUM* schemes as in Otto and Breitung (2023), which reverse the order of the cumulation and start from the end of the sample, can further improve the power of the test.

4.2 Real-time monitoring

In this section we adapt our detectors to identify the change to an explosive regime as new data becomes available, making them suitable for real-time monitoring. We assume an initial set of T_0 observations generated under the null hypothesis /the training sample) and intend to monitor the series for $t = T_0 + 1, \ldots, T_0 + T_m$, where the endpoint of the monitoring period T_m is chosen in advance. The standard approach is to normalize the time index to $r^* = t/T_0$ such that the monitoring runs from $r_0^* = 1$ through $r_m^* = (T_0 + T_m)/T_0$, see e.g. Chu et al. (1996) and Kurozumi (2020). An important drawback involved with this normalization is that any boundary function defined on $r^* \in (1, r_m^*]$ depends on the endpoint r_m^*. Hence, in the literature the parameters of the boundary function are typically reported for various values of r_m^*. As a result, the monitoring boundary function differs from the boundary for retrospective testing which is defined on $r \in (0, 1]$. We therefore suggest to normalize the time index for monitoring as $r = (t - T_0)/T_m$ such that $r \in [-r_0, 0]$ with $r_0 = T_0/T_m$ refers to the training sample and $r \in (0, 1]$ indicates the monitoring period. By applying this normalization, the detectors and boundary functions for retrospective testing and monitoring coincide.

The classical monitoring procedure is based on the *CUSUM* scheme that is computed as the partial sum of the forecast errors (i.e. recursive residuals) yielding the detector LBI_{rT_m} for $r \in (0, 1]$. The training set with T_0 observations is used for estimating nuisance parameters such as σ^2. There is no conceptual difference to the situation of a retrospective *CUSUM* and *mCUSUM* test considered in the previous subsections. The only difference is that we do not have available the complete sample but need to wait for the data to arrive period by period. The decision rule is the same as for the retrospective detectors, that is, the null hypothesis is rejected in the first period where $LBI_{[rT_m]} > b_\alpha(r)$ or $LBI_{[rT_m]} > \bar{b}_\alpha$.

Adapting the *wCUSUM* scheme for real-time forecasting requires a modification to avoid large detection delays due to the small weights that are assigned to the early observations. Without any modification, the test tends to reject at later time periods where the weights are sufficiently large. To overcome this problem we consider a time varying boundary of the form $b_{\bar{c}}(r) = \gamma_\alpha^{\bar{c}}\, w_r^{\bar{c}}$ with $w_r^{\bar{c}}$ defined in (12) such that the null hypothesis is rejected if

$$\frac{1}{\sigma w_r^{\bar{c}}} \sum_{i=1}^{[rT_m]} w_{i/T_m}^{\bar{c}} \Delta y_i \Rightarrow \int_0^r e^{(a-r)\bar{c}} dW(a) > \gamma_\alpha^{\bar{c}} \quad \text{for some } r \in (0, 1].$$

Note that $\int_0^r e^{(a-r)\bar{c}} dW(a) = \int_0^r e^{(r-a)(-\bar{c})} dW(a)$ is a (stationary) Ornstein-Uhlenbeck process that depends on $\bar{c}$ due to the time varying boundary function. We calibrate $\bar{c}$ such that the test attains a power of approximately 50% if $\rho = \bar{\rho}$ for $T = 100$ and $r_e = 0.5$. This results in $\bar{c} = 2.1$ which implies $\bar{\rho} = 1.02$ for $T = 100$. The corresponding one-sided 5% critical value is 1.25. In our simulations we have found, that different choices of $\bar{c}$ have only a negligible effect on the power of the test. Moreover, in a monitoring setting where the full sample is not yet available, it is not possible to estimate the parameter $\sigma^2 = T_m^{-1} \sum_{t=1}^{T_m} \mathbb{E}(u_t^2)$ consistently under heteroskedasticity.

It is also straightforward to use the DF statistic for real-time monitoring, cf. Homm and Breitung (2012), Phillips et al. (2015) and Kurozumi (2020). Since for this detector the training sample is included when computing the DF statistic, we normalize the time index such that the respective boundary function corresponds to the boundary function for retrospective testing. Accordingly, $r = t/(T_0 + T_m)$ such that the monitoring period refers to $r \in (r_0, 1]$ with $r_0 = T_0/(T_0 + T_m)$. The null hypothesis is rejected if the DF statistic exceeds some critical value $cv_\alpha^{df}(r_0)$ for the first time. The critical value is obtained from the $(1 - \alpha)$-quantile of the distribution of

$$\sup_{r \in [r_0, 1]} \frac{\int_0^r W(a) \mathrm{d}W(a)}{\sqrt{\int_0^r W(a)^2 \mathrm{d}a}} \ ,$$

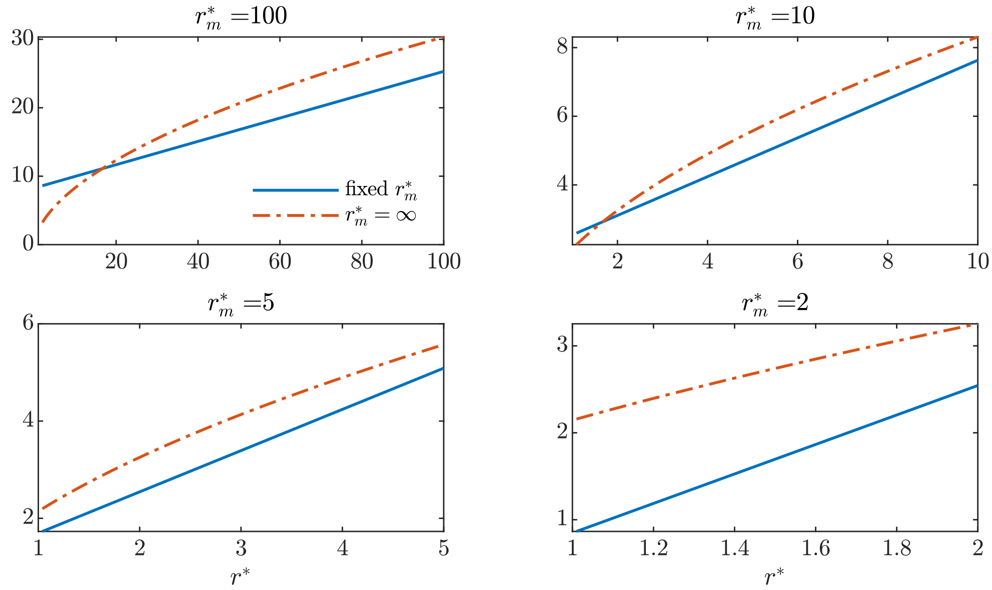
Applying a monitoring scheme, requires choosing a fixed number of monitoring T_m in advance. In contrast, Chu et al. (1996) and Otto and Breitung (2023) discuss monitoring procedures without a fixed terminal date. The idea behind infinite-horizon monitoring is to apply a time-varying (non-linear) boundary function to distribute the size of the test over an infinite monitoring horizon such that the power is larger at the beginning and slowly tends to zero as the number of monitoring periods tends to infinity. From a practical perspective, one may argue that there should not be a significant difference between a very large monitoring horizon and an infinite horizon. However, the choice of boundary function becomes important when the effective number of monitoring periods is “moderate”.

To illustrate these differences, we compare the nonlinear boundary for infinite-horizon monitoring to our linear boundary function for fixed-horizon monitoring with the CUSUM detector. Chu et al. (1996, Equation 8) propose the infinite-horizon boundary

$$d_\infty(r^*) = \sqrt{r^*} \sqrt{\theta_\alpha + \log(r^*)} \tag{16}$$

with time normalization $r^* = t/T_0 \in (1, \infty)$ and θ_α is a constant depending on the significance level α. For the one-sided test, we follow Homm and Breitung (2012) and set $\theta_{0.05} = 4.6$.

Figure 3: CUSUM boundary functions for infinite vs. fixed horizon monitoring



Note: x-axis is normalized by the relative time index $r^* = (t - T_0)/T_0$, where T_0 is the number of training sample periods. Hence $r^* > 1$ refers to the monitoring sample.

How does this boundary function compares to the linear boundary function applied to the fixed-end monitoring scheme considered above? To this end we need to adapt the normalization of the time scale r^* for infinite monitoring scheme to the linear boundaries, yielding a linear function $\tilde{b}_\alpha(r^*)$ for $r^* \in [1, r_m]$. Figure 3 illustrates the respective boundaries for infinite-horizon monitoring (16) alongside the linear boundary for fixed-horizon monitoring with the *CUSUM* detector for various endpoints $r_m^* \in \{100, 10, 5, 2\}$. For large values, $r_m^* = 100$ and $r_m^* = 10$, the infinite-horizon boundary $d_\infty(r^*)$ is initially below the fixed-horizon boundary, leading to more frequent rejections at the start of the monitoring sequence. In contrast, the linear fixed-horizon boundary tends to reject more frequently in later time periods. Thus, the main difference between these two approaches lies in how size and power are distributed across the monitoring periods.

In practice, realistic monitoring horizons are unlikely to exceed five times the size of the training sample, i.e. $r_m^* \leq 5$. In the more realistic scenarios, $r_m^* = 5$ and $r_m^* = 2$, the infinite-horizon boundary function $d_\infty(r^*)$ lies above the linear fixed-horizon boundary for all periods. As a result, the infinite-horizon boundary leads to a test that is unnecessarily conservative when applied to a moderate number of monitoring periods. Therefore, we do not recommend infinite-horizon monitoring in practice, whenever monitoring is not expected to continue for a very long time , say $r_m^* > 10$.

4.3 Date-stamping

Whenever the null hypothesis is rejected in favor of a bubble process, it is interesting to know when the bubble has emerged. Phillips et al. (2011) (in short PWY) propose estimating the bubble's starting date as the first time a sequence of ADF tests, computed using a forward-expanding window, crosses a boundary function from below. Denote $ADF(r)$ the ADF test based on the sub sample from $t = 1$ to $t = [rT]$ where $r \in [r_0, 1]$ is the relative sample length and r_0 is the minimum size. Then, the estimated starting date of the bubble is obtained as

$$\widehat{r}_e^{PWY} = \inf_{r \in [r_0,1]} \{r : \ \mathrm{ADF}(r) > b_{PWY}(r)\}$$

with the boundary function

$$b_{PWY}(r) = \log(\log(rT))/100 \tag{17}$$

that corresponds to a pointwise significance level of roughly 4 percent. Phillips et al. (2011) show that this estimator for the break-date is consistent in the sense that the estimated break-date converges in probability to the true break-date r_e as $T \to \infty$.

However, this approach has several drawbacks. First, the PWY date-stamping estimator exhibits a positive bias in small samples due to the inherent delay of boundary-crossing methods, which require sufficient observations from the bubble regime before crossing the boundary. Second, the probability of crossing the boundary depends on the arbitrarily chosen significance level. Finally, as discussed in Section 3, the DF statistic is not an optimal detector for explosive episodes, particularly for one-sided hypotheses.

Phillips et al. (2015) (in short PSY) propose a backward expanding supremum ADF detector (*BSADF*), where the (relative) endpoint of the test window is fixed at r and the sup value of the ADF sequence is computed over the interval $[0, r - r_0]$, where r_0 ensures a sufficient sample size for computing the ADF statistic. The resulting statistic is defined as

$$\mathrm{BSADF}(r) = \sup_{r_1 \in (0, r - r_0]} \{\mathrm{ADF}(r_1, r)\},$$

where $\mathrm{ADF}(r_1, r)$ denotes the ADF test statistic applied to the subsample $t = [r_1 T] +$

$1, \ldots, [rT]$. Phillips et al. (2015) estimate the starting date of the bubble as

$$\hat{r}_e^{PSY} = \inf_{r \in [r_0, 1]} \{ r : \text{BSADF}(r) > b^{PSY}(r) \},$$

that is, bubble emergence is estimated as the time period at which the $BSADF(r)$ sequence rises above the boundary for the first time. The boundary function b_{PSY} is obtained by simulation to ensure a certain pointwise significance level.

In contrast, Homm and Breitung (2012) propose to estimate the break-point as the *maximum* of all *Chow* statistics. Using the maximum as a date-stamping estimator prevents us from defining some mainly arbitrary boundary function according to some arbitrary significance level. Moreover, it avoids the systematic small sample delay of date-stamping based on boundary-crossing. The *Chow* t-statistic is given by

$$\tau_\lambda(r) = \frac{\sum_{t=[rT]+1}^{[r_f T]} \Delta y_t y_{t-1}}{\sigma \sqrt{\sum_{t=[rT]+1}^{[r_f T]} y_{t-1}^2}} .$$

As the maximum is invariant with respect to σ we can drop this parameter. Maximizing the square of $\tau_\lambda(r)$ is equivalent to the least-squares estimator considered in Bai (1994). Accordingly, maximizing the (one-sided) *Chow* t-statistic imposes the additional information that under the alternative $\rho > 1$. Hence, ignoring σ our proposed *Chow* estimator is

$$\widehat{r}_e = \underset{r \in (0, r_f)}{\operatorname{argmax}} \left(\frac{\sum_{t=[rT]+1}^{[r_f T]} \Delta y_t y_{t-1}}{\sqrt{\sum_{t=[rT]+1}^{[r_f T]} y_{t-1}^2}} \right).$$

In the following theorem we show that $\widehat{r}_e$ is a consistent estimator for r_e as $T \to \infty$.

Theorem 1 *Let y_t be generated by an explosive AR(1) model as in* (2) *with autoregressive parameter as in* (6)*, $y_0 = o_p(T^\theta)$ and the errors satisfy Assumption 1. Then $\widehat{r}_e$ is a consistent estimator for $0 < r_e < r_f$ as $T \to \infty$.*

For date-stamping in a monitoring setting we can choose r_f equal to the time period when the bubble is first detected (which typically comes with a delay of several time periods). For date-stamping in a retrospective setting, the natural choice would be to use the full sample, i.e. $r_f = 1$. To improve the performance of the estimator, we recommend selecting r_f "not far away" from r_e as otherwise the risk increases that a second maximum occurs towards the end of the sample. We therefore suggest cutting the sample for some periods after the detector exceeds the boundary the first time.

5 Small sample properties

In this section we study the small sample properties of the bubble detectors discussed in the previous sections. Throughout, the data are generated by the process in (1) and (2) with standard normally distributed error u_t. The bubble regime starts at $[r_e T] + 1$ and is characterized by $\rho = 1.05$. Unless stated otherwise the initial value is $y_0 = 0$ and all Monte Carlo experiments are based on 10,000 replications and tests are performed at the nominal 5% significance level. In the interest of brevity, we only report results for the case with a constant mean $E[P_t] = \mu$ and without trend. The results for a model with trend are available in the working paper version, see Breitung and Diegel (2024).

5.1 Known starting date of the bubble

For bubbles that are known to emerge at the beginning of the sample. we compare the tests that are applied to the full sample $t = 1, \ldots, T$ from Section 3. We focus on fairly small samples with $T = 50$ because speculative bubbles typically run less than 5 years (60 months). For the case $\mu = 0$ we also present the power envelope as in Elliott et al. (1996) but for explosive alternatives with $\rho > 1$. Note that our (two-sided) LBI^2 statistic is invariant with respect to the parameter μ, whereas the DF test without constant lacks power for $\mu \neq 0$. To compute the *PO-lin*(50%) statistic, we follow the suggestion of Elliott et al. (1996) and chose $\bar{\rho}$ such that the test results attains a power of 0.5. For $T = 50$, we obtain a value $\bar{\rho} = 1.0335$ which is also a reasonable value for a speculative bubble. As another benchmark, we also include the infeasible point-optimal linear test that uses the true ρ for the weights, denoted *PO-lin*(ρ).

The first panel of Table 2 reports rejection rates for tests without deterministics and a zero starting value, i.e. $y_0 = \mu = 0$. As expected, in this case the DF statistic comes close to the power envelope. More surprisingly, the simple linear detectors LBI^2, *PO-lin*(50%) and the infeasible *PO-lin*(ρ) statistics also approach the power envelope and are even more powerful than the DF statistic in the vicinity of the null hypothesis. For $\rho \leq 1.04$, the LBI^2 detector turns out to be somewhat more powerful than the DF(const) statistic.

In the second panel of Table 2 the DGP includes a constant mean $\mu = 10$ and cases (ii) and (iii) distinguish between two different starting values $y_0 = 5$ and $y_0 = 10$. In these cases, no power envelope is available and, therefore, we focus on the test statistics *DF(const)*, LBI^2, *PO-lin*(50%) that are valid and feasible in this scenario. The rejection rates suggest that a larger initial value has a strong positive effect on the power of all tests. This is due to the fact that under the alternative the initial condition shifts the non-centrality parameter

Table 2: Rejection frequencies for alternative detectors

	case (i) $\mu = 0$ and $y_0 = 0$					
ρ	DF	DF(const)	LBI^2	PO-lin(50%)	PO-lin(ρ)	Envelope
1.00	0.050	0.049	0.050	0.050	0.050	0.050
1.01	0.129	0.084	0.131	0.126	0.132	0.135
1.02	0.268	0.161	0.263	0.265	0.269	0.275
1.03	0.435	0.315	0.422	0.440	0.440	0.449
1.04	0.592	0.534	0.574	0.604	0.605	0.616
1.05	0.716	0.707	0.698	0.732	0.737	0.748
1.06	0.811	0.810	0.792	0.823	0.830	0.841
1.07	0.876	0.885	0.859	0.884	0.891	0.901
1.08	0.917	0.930	0.905	0.924	0.931	0.939
1.09	0.944	0.954	0.936	0.950	0.956	0.962
1.10	0.965	0.972	0.957	0.968	0.972	0.977

	case (ii) $\mu = 10$ and $y_0 = 5$			**case (iii)** $\mu = 10$ and $y_0 = 10$		
ρ	DF(const)	LBI^2	PO-lin(50%)	DF(const)	LBI^2	PO-lin(50%)
1.00	0.050	0.050	0.050	0.052	0.050	0.050
1.01	0.096	0.154	0.149	0.126	0.219	0.215
1.02	0.236	0.361	0.371	0.423	0.588	0.610
1.03	0.529	0.596	0.625	0.852	0.865	0.889
1.04	0.789	0.773	0.803	0.980	0.966	0.976
1.05	0.907	0.879	0.901	0.997	0.992	0.995

Notes: DF denotes the Dickey Fuller test with the specification indicated in parenthesis. LBI^2 denotes the locally best invariant test defined in (5)*. PO-lin(50%) is the point-optimal test tailored to achieve a power of 50% at a pre-specified $\bar{\rho}$, whereas PO-lin(ρ) uses the true ρ to compute the weights.*

of the χ^2 distribution away from zero. Moreover, the results in (ii) and (iii) indicate that the linear detectors are much more powerful than the *DF(const)* test for $\rho = 1.02$ and $\rho = 1.03$. This gap closes for larger values of ρ.

All in all, these results show that linear statistics can have advantages over DF-type tests, particularly in the empirically relevant cases with a constant in the data and for detecting moderately explosive roots.[2]

[2] In the working paper version, we document that these findings carry over to the case with a trend in the data and tests that are modified accordingly.

Table 3: Rejection Frequencies of Sequential Bubble Detectors

	supADF	*CUSUM*	*mCUSUM*	*wCUSUM*	$AHLT_{10}$
r_e			**(i) two-sided tests**		
1.0	0.056	0.040	0.044	0.037	0.071
0.8	0.354	0.239	0.359	0.500	0.487
0.6	0.702	0.614	0.696	0.787	0.674
0.4	0.870	0.824	0.864	0.908	0.791
0.2	0.935	0.915	0.933	0.954	0.757
r_e			**(ii) one-sided tests**		
1.0	0.056	0.041	0.046	0.041	0.081
0.8	0.362	0.308	0.432	0.569	0.519
0.6	0.701	0.658	0.732	0.814	0.686
0.4	0.867	0.845	0.883	0.921	0.798
0.2	0.936	0.926	0.946	0.963	0.758

Notes: Monte Carlo rejection rates for $T = 100$. Case (i) allows positive and negative bubbles and all tests employ a two sided alternative. Case (ii) retains only positive bubbles realizations and one-sided tests for "positive" bubbles are used, except for the supADF test.

5.2 Unknown starting date of the bubble

This subsection examines the performance of the sequential testing schemes that are suitable in the more realistic setting where the starting date of the bubble is unknown. Table 3 presents the relative rejection frequencies of the sequential bubble detectors based on $T = 100$ for various sample fractions $r_e \in \{1.0, 0.8, 0.6, 0.4, 0.2\}$ generated under the null hypothesis before the bubble starts. Hence, $r_e = 1$ implies that no bubble exists and the full sample is generated under the null hypothesis. For $r_e = 0.8$ the first 80 observations follow a random walk and the following 20 time periods are generated under the explosive alternative.

We consider three variants of the *CUSUM* approach. The *CUSUM* detector is based on a linear time-varying boundary function, whereas *mCUSUM* and *wCUSUM* use constant boundaries as provided in Table 1. The *wCUSUM* test applies a weighting scheme with $\bar{c} = 2$ such that the test achieves a power of approximately 50% if $\rho = 1 + \bar{c}/T$. In addition to our *CUSUM* type tests and the *supADF* test of Phillips et al. (2011) we include the end-of-sample test of Astill et al. (2017) with their preferred window size of 10 observations, denoted $AHLT_{10}$, in our comparison. For the *supADF* test we use the asymptotic critical values provided by Homm and Breitung (2012).

The first panel of Table 3 is based on two-sided variants of all tests. The first row

($r_e = 1.0$) shows that all tests except for the $AHLT_{10}$ maintain the nominal the size well. The $AHLT_{10}$ is slightly oversized in small samples because it uses a sub-sampling procedure for approximating the actual null distribution.[3]

It turns out that applying a constant boundary improves the power of the $CUSUM$-type tests. This is due to the fact that under the explosive alternative the detector tends to yield the largest value at the end of the sample and, therefore, adjusting the boundary in earlier time periods does not improve the power of the test. As expected, $wCUSUM$ has the highest power regardless of the starting point of the bubble. The improvement in power of $wCUSUM$ over $mCUSUM$ can be attributed to the weighting scheme.

As argued in Section 3, it might be more empirically relevant to apply the tests in a one-sided manner focusing on positive explosive paths (i.e. "positive bubbles"). Therefore, we repeat the previous simulation experiment using the same DGP but retain only realizations with a positive bubble, meaning those with a positive value of X_T in Lemma 1. It is straightforward to conduct one-sided versions of the tests by applying only the positive boundary, except for the $supADF$ test where no one-sided variant exists. We therefore use the two-sided version in this simulation experiment as well. The results in panel (ii) of Table 3 show that focusing on positive bubbles increases the power of the test substantially relative to the two-sided variants. This demonstrates the gains in power that can be achieved by adapting the tests toward a more realistic scenario when testing for bubbles.

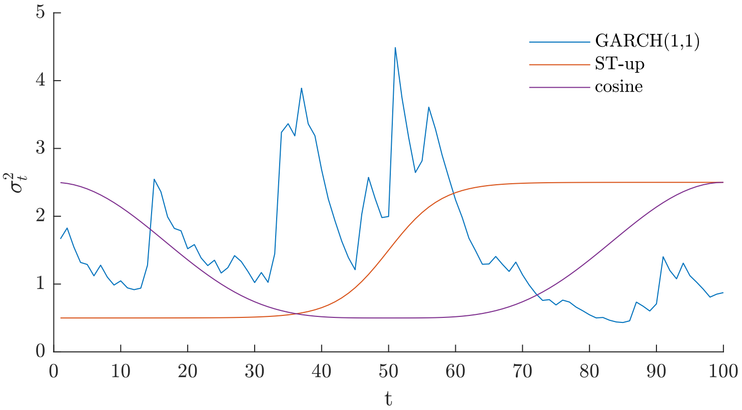
5.3 Heteroskedastic errors

To investigate the performance of the sequential detectors under heteroskedasticity we use the same DGP as in the previous simulations but with a time-varying variance of the normally distributed errors $u_t \overset{iid}{\sim} \mathcal{N}(0, \sigma_t^2)$. We consider three types of heteroscedasticity to accomodate the most empirically relevant cases: First, a stationary GARCH(1,1) model with $\sigma_t^2 = \omega + \alpha u_{t-1}^2 + \beta \sigma_{t-1}^2$ and parameter values $\omega = 0.05$, $\alpha = 0.15$ and $\beta = 0.82$. Second, we use a smooth-transition (ST-up) function inspired by Astill et al. (2023) given by $\sigma_t^2 = 0.5 + \frac{2}{1+e^{-a(t-0.5T)}}$ with $a = -0.25$ to shift the variance upwards from $\sigma_1^2 = 0.5$ to $\sigma_T^2 = 2.5$ in the middle of the sample. Third, we use a cosine function $\sigma_t^2 = 0.5 + 0.5(1 + cos(2\pi t/T))^2$ to implement a temporary drop in variance. Figure 4 displays the time profiles of these variance processes alongside a selected realization of the GARCH(1,1) process, parameterized to produce variances of roughly similar magnitudes.

We robustify the $mCUSUM$ and $wCUSUM$ tests by using the robust variance estimator

[3] When increasing the number of observations the size distortion vanishes quickly.

Figure 4: Different variance profiles



(14). In line with results of Phillips et al. (2015), deviations of the *supADF* test from the nominal 5% level are relatively small under GARCH(1,1) type heteroscedasticity, see the top panel of Table 4 for $r_e = 1$. The same is true for the *CUSUM* test that is also not robust to heteroskedasticity. This is due to the fact that the GARCH process has a constant *unconditional* variance. This does not imply however that GARCH type volatilities are no problem for bubble testing. If the volatility is high in the beginning of the sample, then the *CUSUM* detector yields early rejections with a higher probability. In our simulations, however, GARCH volatilities may be large or small in the first part of the sample resulting in positive and negative size distortions; but *on average* the actual size is close to the nominal size.

The two unconditional types of heteroscedasticity, *ST-up* and *cosine*, are more problematic. The *supADF* test is severely oversized for the *ST-up* variance and undersized for the *cosine* variance. This pattern is reversed for the *CUSUM* test. In contrast the heteroskedasticity-robust tests *mCUSUM*, *wCUSUM*, and $AHLT_{10}$ maintain the sizes quite well. While the $AHLT_{10}$ is robust against heteroscedasticity, it suffers a particularly large loss in power under unconditional heteroscedasticity compared to the homoscedastic case. In contrast, the robust *mCUSUM* and *wCUSUM* detectors imply a much smaller loss of power. Being the most powerful and robust test, the *wCUSUM* appears to be the overall preferred choice.

Table 4: Rejection rates under various variance profiles

	supADF	*CUSUM*	*mCUSUM*	*wCUSUM*	$AHLT_{10}$
			$r_e = 1.0$		
Homoscedasticity	0.052	0.040	0.044	0.037	0.067
GARCH(1,1)	0.068	0.055	0.042	0.038	0.073
ST-up	0.140	0.014	0.042	0.037	0.053
cosine	0.031	0.061	0.045	0.034	0.052
			$r_e = 0.8$		
Homoscedasticity	0.348	0.232	0.352	0.494	0.362
GARCH(1,1)	0.360	0.242	0.357	0.509	0.369
ST-up	0.416	0.231	0.354	0.434	0.187
cosine	0.336	0.229	0.338	0.430	0.178
			$r_e = 0.6$		
Homoscedasticity	0.704	0.616	0.697	0.788	0.592
GARCH(1,1)	0.709	0.618	0.699	0.787	0.569
ST-up	0.678	0.549	0.642	0.701	0.230
cosine	0.692	0.595	0.679	0.745	0.271

Notes: The table shows rejection rates of the bubble tests from 10000 simulated time series with a DGP that switches to the explosive regime with $\rho = 1.05$ *after various fractions of the samples, denoted* r_e*. For* $r_e = 1.0$*, there there is no explosive regime and the rejection rates are the sizes of the tests.*

5.4 Real-time monitoring

Next, we examine the performance of the *supADF*, *CUSUM*, *mCUSUM* and *wCUSUM* detectors for real-time monitoring.[4] We use a fixed number of monitoring periods $T_m = 50$ after a training sample of equal length $T_0 = 50$, both corresponding to roughly 4 years worth of monthly data. This yields a total of $T_0 + T_m = 100$ observations. For monitoring with the *wCUSUM* detector we apply the time-varying bound with $\bar{c} = 2.1$ and simulated 5% critical values of 1.25. For monitoring with the *CUSUM* and *mCUSUM* detectors, the critical values in Table 1 apply.

A bubble is detected when the respective detectors first exceed their respective boundaries. Since the detection typically occurs with a delay, the detection delay that counts the number of periods between the true bubble emergence r_e and the detection r_f, is an additional

[4]Unfortunately, we are not able to include the end-of-sample monitoring procedure suggested by Astill et al. (2018), as in our Monte Carlo experiment our training sample is not large enough for controlling the nominal size of 0.05.

Table 5: Rejection frequency and detection delay for monitoring

	(i) Rejection Frequency				**(ii) Avg. Detection Delay**			
r_e	*CUSUM*	*mCUSUM*	*wCUSUM*	*supADF*	*CUSUM*	*mCUSUM*	*wCUSUM*	*supADF*
1.0	0.047	0.046	0.046	0.047	–	–	–	–
0.8	0.177	0.272	0.290	0.171	1	1	1	1
0.6	0.401	0.508	0.542	0.434	4	5	5	5
0.4	0.609	0.688	0.701	0.636	9	10	10	10
0.2	0.738	0.797	0.800	0.756	13	15	14	15

Notes: In the left panel, figures are relative rejection frequencies and in the right panel the average detection delay (rounded to integers) is reported for the first rejection of the monitoring procedures with $T_m = 50$ *monitoring periods after* $T_0 = 50$ *training periods. Only positive bubble realizations are retained.*

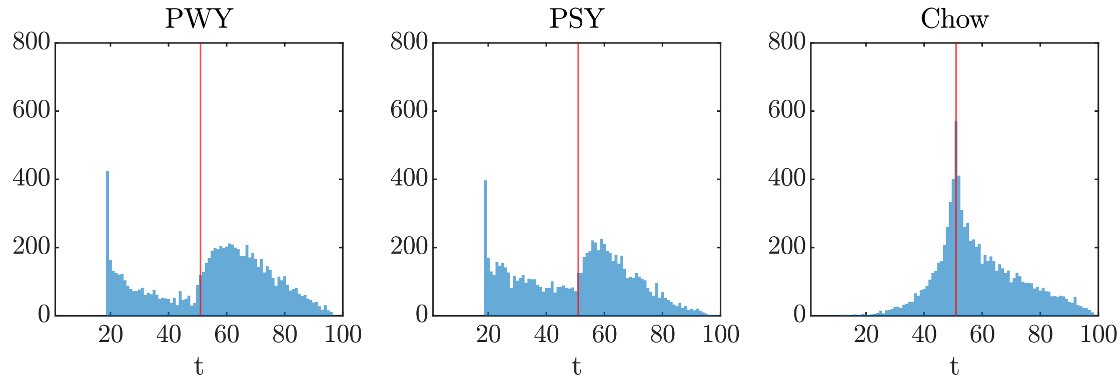
statistic that is important in this setting.

Table 5 confirms the results from the retrospective testing showing that the *wCUSUM* detector is most powerful, closely followed by *mCUSUM*. Both have a particular advantage when the bubble emerges toward the end of the monitoring period. It is reassuring that the adapted version of the *wCUSUM* detector retains its power well in the monitoring setting. The *supADF* and *CUSUM* perform similar for all values of r_e.

Regarding the average detection delays, one observation stands out: The later the bubble emerges, the smaller is the delay of all detectors. There are two explanations for this finding. First, a later bubble leaves fewer periods in the remaining sample, naturally limiting the maximum possible delay. In our setting, the maximum possible delay for $r_e = 0.8$ is 10 periods. Since the reported average delays of one period are significantly smaller than 10 in this case, this obvious explanation cannot fully account for the observed pattern.

The second explanation relates to the "strength" of the bubble, represented by the factor X_T in Lemma 1. In Figure 1 it is apparent that the power and delay of any bubble test depend on the size of X_T. For bubbles starting toward the end of the sample (for instance $r_e = 0.8$) there are fewer periods to build up. Thus, the test requires sizable values of X_T to achieve sufficient power, which typically implies a relatively short detection delay. In contrast, for early emerging bubbles, the test can also detect those with moderate X_T, as the cumulative statistic has more time to accumulate, leading to longer delays. Consequently, the average delay tends to be longer for smaller values of r_e. Overall, for the same r_e detection delays are similar across all tests, with no procedure showing a clear advantage.

Figure 5: Distribution of date-stamping estimators



Notes: Histograms from date-stamping estimates based on 10,000 MC repetitions with $T = 100$. The vertical line at observation $t = 51$ indicates the true emergence of the bubble.

5.5 Performance of the date-stamping procedures

This section compares the performance of the maximum *Chow* estimator (see Theorem 1) for identifying the bubble's starting date with that of the *PWY* and *PSY* estimators. We consider simulated time series with $T \in \{100, 200, 400\}$ observations of the DGP in (1) and (2) with $r_e = 0.5$.

In practice it is common to first establish whether there is a bubble in the present sample and subsequently use the date-stamping estimators to date the start of the bubble. Therefore, to mimic a realistic setup, we first use the *wCUSUM* detector to test for a bubble in a retrospective setting, applying date-stamping methods only if a bubble is detected. As argued in Section 4.3, it is recommended to cut the time series some periods after the first crossing of the boundary, to avoid a possible second maximum at the very end of the sample. We therefore terminate the sample 10 time periods after the detector exceeds the boundary. In a monitoring setting, it is more natural to apply the date-stamping methods just after the first rejection of the null hypothesis. We apply the *PWY* and *PSY* dates-stamping estimators after pretesting for a bubble with the *supADF* and the *generalized supADF* test, respectively.[5]

[5]Date-stamping with the *PWY* and *PSY* procedures requires the choice of the minimum relative window width. We follow Phillips et al. (2015) and use the recommended window length of $r_0 = 0.01 + 1.8/\sqrt{T}$ with corresponding finite sample sample critical values as tabulated in Table 1 of their paper. For our sample sizes of 100 and 200 the minimum window equals 19, 27 and 40 observations. For date-stamping with the *ADF* statistic we use the boundary in (17), and for the *BSADF* we use simulated critical values instead. To implement the *PWY* and *PSY* procedures, and to simulate critical values, we use the MATLAB code provided by the authors at `https://sites.google.com/site/shupingshi/home/codes`.

Table 6: Comparison of date-stamping procedures

	$T=100$, $[r_eT]=51$			$T=200$, $[r_eT]=101$			$T=400$, $[r_eT]=201$		
	PWY	*PSY*	*Chow*	*PWY*	*PSY*	*Chow*	*PWY*	*PSY*	*Chow*
Mode	19	19	51	27	27	101	40	40	201
Mean	55	50	58	96	81	110	162	127	209
Std. Dev.	20	19	14	40	38	18	76	69	18
RMSE	21	19	15	40	42	20	86	101	20
$\mathcal{P}(\hat{r}_e \in [r_e \pm 0.1])$	0.22	0.26	0.49	0.37	0.35	0.76	0.49	0.32	0.93

Notes: Summary statistics of date-stamping procedures for 10,000 replications with $\rho = 1.05$ *and* $r_e = 0.5$ *(bubble starts in the middle of the sample). Only positive bubbles are retained.*

The simulation results confirm the good performance of the maximum *Chow* estimator. The histogram in Figure 5 shows that the simulated distribution for $T = 100$ of the *Chow* estimator concentrates around the true start of the bubble, as indicated by the vertical red line.

In contrast, the histograms of the *PWY* and *PSY* estimators are bimodal. In both cases, the first peak occurs at the first test statistics with $t = [r_0T] = 19$. This first peak is related to the pointwise significance level of the boundary function. For example, the implied significance level for the *ADF* sequence with the boundary (17) for the *PWY* procedure is roughly 4 percent. Hence, we expect 4 percent false detections (i.e. "spurious bubbles emergences") for the first test statistic at $[r_0T]$. Accordingly, the *PWY* date-stamping procedure eventually dates the start of a bubble in 4 percent of the cases at the very beginning of the testing period. By construction, this first peak disappears as T gets large, but as the $\log(\log(\cdot))$ boundary function is varying slowly, the sample needs to become very large to ensure that the estimates concentrates solely around the true starting date.[6] In our simulations, even a sample of $T = 400$ was not enough for the first peak to vanish. The second peak dates the start of the actual bubbles. As expected, this peak occurs with a notable delay.

Table 6 shows that these observations are also quantitatively relevant. Not only does the most frequent value (mode) of the maximum *Chow* estimator correspond exactly to the true starting date of the bubble. It also outperforms the *PWY* and *PSY* estimators in terms of standard deviation and RMSE. For $T = 100$, the table seems to suggest a larger bias in the mean of the the *Chow* estimator. However, the seemingly smaller bias of the *PWY* and *PSY* estimators is an artifact and due to the first peak. For, $T = 200$ and $T = 400$,

[6] The first peak also appears when using simulated critical values instead of the boundary function in (17) for the *ADF* sequence in the *PWY* procedure.

both PWY and PSY estimate the bubble to start too early on average. This does not imply however that these methods detect the bubble even before it occurs but is just a result of its bimodality.

The last row of Table 6, shows that the percentage of estimated dates that fall into an interval of $\pm 0.1 \cdot T$ around the true emergence $r_e T$, denoted as $\mathcal{P}(\hat{r}_e \in [r_e \pm 0.1])$, is at least twice as high for the *Chow* estimator compared to PWY or PSY. To conclude, the *Chow* appears to be a quite accurate estimator for dating the start of a single bubble. That said, there may be cases that we do not consider in which the PWY or PSY have advantages, for instance if there are multiple bubbles.

6 Empirical Applications

Speculative bubbles are more likely to emerge when the fundamental value is uncertain. For start-ups with high initial investment costs and low cash flows, the fundamental value can only be estimated unreliably. As a result, uncertain expectations play a greater role in financial analysts' assessments of newly founded companies, making it easier for speculative bubbles to emerge under such market conditions. A well-known example is the so-called Dotcom bubble at the end of the last century, when numerous internet-related companies went public.

For our empirical analysis, we select two examples of possible speculative bubbles. The first is the hydrogen bubble from 2019 to 2020, represented by the stock price of Plug Power. The second example is the price of Bitcoin, which has shown patterns of speculative bubbles in 2017 and 2021. We analyze whether Bitcoin shows evidence of another speculative bubble since the end of 2023.

In our analysis, we compare two tests from the literature, the *CUSUM* and *supADF* tests with our new variants the *mCUSUM* and *wCUSUM*.[7] For dating the emergence of the bubbles, we use the PWY and the maximum *Chow* procedures. We discuss the findings in detail in the following subsections. Table 7 summarizes the results.

6.1 The hydrogen bubble of 2020

Figure 6a presents the Plug Power stock price series from January 2018 through January 2021. Throughout 2020, the economic prospects of "green" hydrogen technologies were

[7] We apply all tests to the log transformed price series without detrending. For the *supADF* test, we follow Phillips et al. (2011) and Phillips et al. (2015) and use an intercept term in the test regression. The minimum window size is set by $r_0 = 0.01 + 1.8\sqrt{T}$.

Table 7: Results of Empirical Applications

i) Plug Power, 06-Jan-2018 to 30-Jan-2021, $T = 161$						
	supADF	*CUSUM*	*mCUSUM*	*wCUSUM*	date-stamping	
Statistic	2.87	0.81	2.41	2.88	*PWY*	13-Jun-2020
5% critical value	1.34	0.85	1.95	1.95	*Chow*	04-Apr-2020
reject?	yes	no	yes	yes		
ii) Bitcoin, 02-Oct-2022 to 15-Dec-2024, $T = 116$						
	supADF	*CUSUM*	*mCUSUM*	*wCUSUM*	date-stamping	
statistic	0.71	0.77	2.3	2.53	*PWY*	n.a.
5% critical value	1.3	0.85	1.95	1.95	*Chow*	27-Oct-2024
reject?	no	no	yes	yes		

viewed with increasing optimism, and financial analysts competed with in issuing exuberant growth forecasts for this sector. As a result, stock prices for companies like Plug Power sky-rocketed from values below 5$ in 2019 to 75$ in January 2021. By June 2024, the stock price was back down to the level of 2019. During the same period, the fundamental conditions of hydrogen companies changed very little, leaving hardly any doubt that the explosive rise was a speculative bubble.

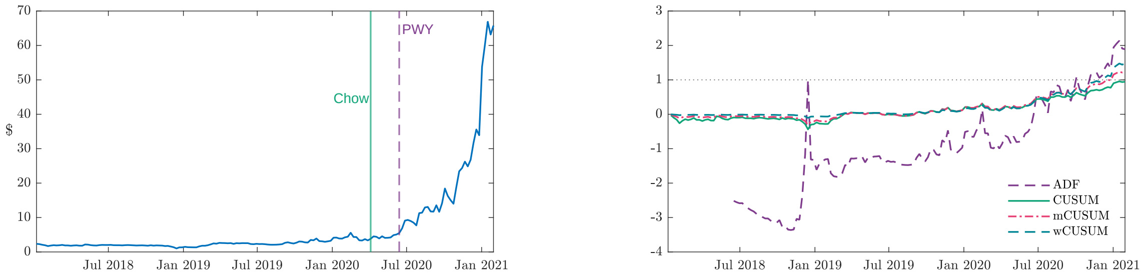
All tests except *CUSUM* find a bubble in the series at a 5% level.[8] Figure 6b plots the underlying sequences of the tests. For an easy comparison, the sequences are divided by the respective boundaries. Therefore, the null hypothesis is rejected if the normalized test statistic exceeds unity. The *ADF*, *mCUSUM*, and *wCUSUM* sequences become significant towards the end of the sample. Due to the weighting scheme of *wCUSUM*, the underlying sequence remains close to zero for the first part of the sample and then rises more quickly than the other *CUSUM* variants, starting in July 2020.

The maximum *Chow* procedure dates the start of the bubble at the beginning of April. Date-stamping based on the *PWY* can be a little more delicate. First, the *PWY* procedure does not distinguish between positive and negative bubbles, and Phillips et al. (2011) note that the procedure can also detect multiple bubbles.[9] Recall that the *PWY* date-stamping estimator is based on the same *ADF* sequence shown in Figure 6b, but with a different boundary function that is generally lower than the critical value for the *supADF* statistic.

[8] At a 10% level also *CUSUM* detects a bubble.

[9] We do not apply the *PSY* procedure that is specifically designed for multiple bubbles, because Phillips et al. (2015) document that the *PWY* procedure works well in cases where the second bubble in the sample runs the longest, which is obviously the case in our example.

Figure 6: Results for Plug Power



(a) Price of Plug Power and Bubble Dates (b) Sequence of normalized bubble detectors

The *ADF* statistic crosses its boundary three times from below, which may imply three different bubbles in the sample. For the first two crossings the *ADF* sequence exceeds the critical boundary only for a single observation. We follow the advice of Phillips et al. (2015) and ignore such "blips". Since we want to date the start of the obvious positive bubble that runs up to the sample end in Figure 6a, we use the last boundary crossing of the *ADF* sequence as the estimated starting date.[10] Hence we conclude that the *ADF* sequence dates the start of the bubble at the beginning of June 2020, a few weeks after the *Chow* estimator. Estimated dates are qualitatively close, this confirms the greater delay of the *PWY* date-stamping procedure that was also documented in our Monte Carlo simulations.

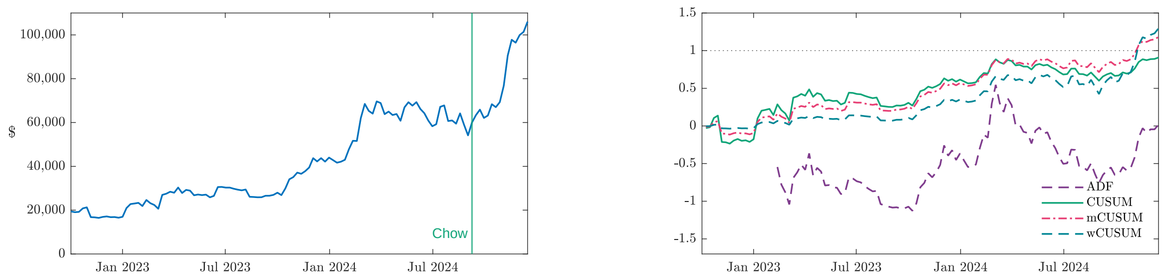
6.2 Is there a bubble in the price of Bitcoin?

From January 2023 to December 2024 the price of Bitcoin has quintupled with particularly steep rice increases in early 2024 and around the US election at the beginning of November that year, see Figure 7a. This behavior might suggest a speculative bubble in recent Bitcoin prices. To test for the presence of a speculative bubble we again apply the bubble detectors to $T = 116$ weekly observations of the log price of Bitcoin from October 2022 to December 2024.

The results in Panel *ii)* of Table 7 show that only the *mCUSUM* and *wCUSUM* detectors are able to indicate the presence of a bubble. Figure 7b shows the normalized sequential test statistics . The *mCUSUM* and *wCUSUM* detectors both enter the rejection region after the US election. Note that the *ADF* sequence increases sharply after January 2024 but it starts from a level that is too low to reach the rejection region.

[10]The first bubble, dated December 2018 at the sharp peak visible in Figure 6b, is due to a sharp drop of the price series that is hardly visible in the original price series.

Figure 7: Results for Bitcoin



(a) Price of Bitcoin and Bubble Dates (b) Sequence of normalized bubble detectors

Since the *supADF* test does not indicate a bubble, we do not use the *PWY* date-stamping procedure. To date the start of the bubble indicated by *mCUSUM* and *wCUSUM*, we apply the maximum *Chow* procedure, which estimates the bubble to emerge in October 27, 2024, shortly before the US election, see Table 7.

All in all, both applications confirm that the modifications of the *CUSUM* type detectors and the maximum *Chow* date-stamping procedure are also empirically relevant. Particularly, their ability to perform one-sided test for “positive” bubbles is an important advantage and our date-stamping procedures estimates the bubble to emerge earlier than the *PSY* date-stamping procedure.

7 Conclusions

Testing for speculative bubbles involves testing the unit root hypothesis against the explosive alternative. One might argue that the results for testing against stationary alternatives (i.e., against negative deviations) can naturally be extended to positive deviations. In this paper, we show that this is not necessarily the case and identify several peculiarities when testing for explosive alternatives.. For instance, using a simple linear (*LBI*) test statistic can significantly improve the power of the test near the null hypothesis and it is possible to distinguish positive and negative bubbles by using a one-sided test. Furthermore, such tests have a standard limit distribution and are robust against heteroskedasticity.

The linear test statistic can be readily adapted for a sequential testing scheme that can be used to test for a bubble with an unknown starting point. This leads to the standard *CUSUM* test proposed by Homm and Breitung (2012). The power of the *CUSUM* test can be further improved by (i) applying a constant boundary function, (ii) adopting an exponential weighting scheme and (iii) accumulating the differences in reversed chronological order (cf. Otto and Breitung 2023; Breitung and Diegel 2024). Another advantage is that the *CUSUM*-type test statistics with constant boundary are robust to heteroskedasticity, a key characteristic of asset returns. Furthermore, the sequential tests can be adapted for real-time monitoring.

Whenever the (retrospective or monitoring) test indicates the presence of a bubble, it is interesting to estimate its starting date. To this end, we consider the maximum *Chow* t-statistic. We argue that this estimator is a one-sided version of the Gaussian maximum likelihood estimator and show that it is consistent as the sample size tends to infinity. Our Monte Carlo experiments suggest that the maximum *Chow* estimator performs much better than estimators based on crossing some boundary for the DF test statistic.

An important limitation of our analysis is the exclusion of cases in which the bubble bursts within the sample. Consequently, we also disregard multiple bubbles, as considered in Phillips et al. (2015). This is not a serious issue if the tests are applied as part of a monitoring exercise, as they typically end immediately after a bubble is identified. Furthermore, crashes are usually easy to detect due to the dramatic drop in the price series, making it straightforward to exclude bubble crashes from the sample.

Appendix: Proofs

Proof of Lemma 1:

PROOF: Following Phillips and Magdalinos (2007) we have

$$\begin{aligned} y_t &= \rho_T^t y_0 + \sum_{i=0}^{t-1} \rho_T^i u_{t-i} \\ &= \rho_T^t \left(y_0 + \rho_T^{-1} u_1 + \rho_T^{-2} u_2 + \rho_T^{-3} u_3 + \cdots + \rho_T^{-t} u_t \right) \\ &= \rho_T^t X_T - \rho_T^{-1} u_{t+1} - \rho_T^{-2} u_{t+2} - \cdots - \rho_T^{t-T} u_T \qquad (18) \\ \text{where} \quad X_T &= y_0 + \rho_T^{-1} u_1 + \rho_T^{-2} u_2 + \rho_T^{-3} u_3 + \cdots + \rho_T^{-T} u_T \end{aligned}$$

and $T^{-\theta/2} X_T \Rightarrow \mathcal{N}(0, \sigma^2/(2c))$. It follows that $(\rho_T^{-1} u_{t+1} + \rho_T^{-2} u_{t+2} + \cdots + \rho_T^{t-T} u_T)$ is also $O_p(T^{\theta/2})$. Since ρ_T^t is $O\left[\exp\left(crT^{1-\theta}\right)\right]$ (see Guo et al. 2019, Lemma A.1) it follows that for $X_T \neq 0$, y_t will be asymptotically dominated by $\rho_T^t X_T$ and the series follows approximately a deterministic exponential trend. ■

Proof of Theorem 1:

To proof this theorem we show that (i) $\text{Prob}\Big(r_e \in (r_e, r_e + \delta]\Big) \to 0$ and (ii) $\text{Prob}\Big(r_e \in [r_e - \delta, r_e)\Big) \to 0$ for all $\delta > 0$ as $T \to \infty$.

Part (i): Let us first consider the case that the maximum occurs at $\widetilde{r}_e = r_e + \delta$ with some $\delta > 0$. This implies that

$$\frac{\sum_{t=[(r_e+\delta)T]+1}^{T} \Delta y_t y_{t-1}}{\sqrt{\sum_{t=[(r_e+\delta)T]+1}^{T} y_{t-1}^2}} > \frac{\sum_{t=[r_e T]+1}^{T} \Delta y_t y_{t-1}}{\sqrt{\sum_{t=[r_e T]+1}^{T} y_{t-1}^2}}.$$

Using $\Delta y_t = (\rho_T - 1)y_{t-1} + u_t$ we obtain

$$(\rho_T - 1)\sqrt{\sum_{t=[(r_e+\delta)T]+1}^{T} y_{t-1}^2} + R_T(r_e + \delta) > (\rho_T - 1)\sqrt{\sum_{t=[(r_e)T]+1}^{T} y_{t-1}^2} + R_T(r_e),$$

where

$$R_T(r) = \frac{\sum_{t=[rT]+1}^{T} u_t y_{t-1}}{\sqrt{\sum_{t=[rT]+1}^{T} y_{t-1}^2}}.$$

As shown by Phillips and Magdalinos (2007)

$$\begin{aligned}
\sum_{t=[rT]+1}^{T} u_t y_{t-1} &= O_p(\rho_T^{(1-r)T} T^{\theta}) \\
\sum_{t=[rT]+1}^{T} y_{t-1}^2 &= O_p(\rho_T^{2(1-r)T} T^{2\theta}) \\
(\rho_T - 1)\sqrt{\sum_{t=[(r_e+\delta)T]+1}^{T} y_{t-1}^2} &= O_p(\rho_T^{2(1-r)T})
\end{aligned}$$

for $r \geq r_e$. It follows that $R_T(r)$ is $O_p(1)$ and, therefore, such terms are asymptotically negligible. Therefore it is sufficient to consider

$$
\begin{aligned}
&\text{Prob}\left(\sqrt{\sum_{t=[(r_e+\delta)T]+1}^{T} y_{t-1}^2} > \sqrt{\sum_{t=[r_eT]+1}^{T} y_{t-1}^2}\right) \\
&= \text{Prob}\left(\sum_{t=[(r_e+\delta)T]+1}^{T} y_{t-1}^2 > \sum_{t=[r_eT]+1}^{[(r_e+\delta)T]} y_{t-1}^2 + \sum_{t=[(r_e+\delta)T]+1}^{T} y_{t-1}^2\right) \\
&= \text{Prob}\left(\sum_{t=[r_eT]+1}^{[(r_e+\delta)T]} y_{t-1}^2 < 0\right) \;=\; 0 \quad \text{for all } \delta > 0
\end{aligned}
$$

It follows that the probability that the test statistic has a maximum at $r_e + \delta$ tends to zero for all $\delta > 0$.

Part (ii): For the case that the maximum occurs at $\tilde{r}_e = r_e - \delta$, the *Chow* t-statistic includes observations from the fundamental regime. In this case we obtain

$$
\begin{aligned}
&\text{Prob}\left((\rho_T - 1)\frac{\sum_{t=[r_eT]+1}^{T} y_{t-1}^2}{\sqrt{\sum_{t=[(r_e-\delta)T]+1}^{T} y_{t-1}^2}} + R_T(r_e-\delta) + o_p(1) > (\rho_T - 1)\sqrt{\sum_{t=[r_eT]+1}^{T} y_{t-1}^2}\right) + R_T(r_e) \\
&= \text{Prob}\left((\rho_T - 1)\frac{A_T}{\sqrt{A_T + B_T}} - (\rho_T - 1)\sqrt{A_T} + R_T(r_e - \delta) - R_T(r_e) > 0\right)
\end{aligned}
$$

where $A_T = \sum_{t=[r_eT]+1}^{T} y_{t-1}^2$ and $B_T = \sum_{t=[(r_e-\delta)T]+1}^{r_eT} y_{t-1}^2$. A first order Taylor expansion yields

$$
\frac{(\rho_T - 1)A_T}{\sqrt{A_T + B_T}} - (\rho_T - 1)\sqrt{A_T} = -\frac{(\rho_T - 1)}{2}\frac{B_T}{\sqrt{A_T}} \;=\; O_p\left(\frac{T^{2(1-\theta)}}{\rho_T^{(1-r_e)T}}\right)
$$

whereas

$$
R_T(r_e - \delta) - R_T(r_e) = O_p\left(\frac{T^(1-\theta)}{\rho_T^{(1-r)T}}\right).
$$

Accordingly, $R_T(r_e - \delta) - R_T(r_e)$ is asymptotically negligible and it is sufficient to consider

$$
\text{Prob}\left(-\frac{1}{2}\frac{B_T}{\sqrt{A_T}} > 0\right) = 0
$$

as both A_T and B_T are positive. ■

References

Astill, S., D. Harvey, S. Leybourne, and R. Taylor (2017). Tests for an end-of-sample bubble in financial time series. *Econometric Reviews 36*(6-9), 651–666.

Astill, S., D. Harvey, S. J. Leybourne, R. Sollis, and A. Robert Taylor (2018). Real-time monitoring for explosive financial bubbles. *Journal of Time Series Analysis 39*(6), 863–891.

Astill, S., D. I. Harvey, S. J. Leybourne, A. R. Taylor, and Y. Zu (2023). Cusum-based monitoring for explosive episodes in financial data in the presence of time-varying volatility. *Journal of Financial Econometrics 21*(1), 187–227.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis 15*(5), 453–472.

Blanchard, O. and M. Watson (1982). Bubbles, rational expectations, and financial markets. In P. Wachtel (Ed.), *Crisis in the economic and financial structure*, pp. 295–315. Lexington Books.

Breitung, J. and M. Diegel (2024). Sequential detector satistics for speculative bubbles. Manuscript. Available at `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5065510`.

Breitung, J. and R. Kruse (2013). When bubbles burst: Econometric tests based on structural breaks. *Statistical Papers 54*(4), 911–930.

Brown, R., J. Durbin, and J. Evans (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society Series B: Statistical Methodology 37*(2), 149–163.

Cavaliere, G. (2005). Unit root tests under time-varying variances. *Econometric Reviews 23*(3), 259–292.

Cavaliere, G. and R. Taylor (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics 140*(2), 919–947.

Chu, J., M. Stinchcombe, and H. White (1996). Monitoring structural change. *Econometrica 65*(5), 1045–1065.

Elliott, G., T. Rothenberg, and J. Stock (1996). Efficient tests for an autoregressive unit root. *Econometrica 64*(4), 813–836.

Guo, G., Y. Sun, and S. Wang (2019). Testing for moderate explosiveness. *The Econometrics Journal 22*(1), 73–94.

Homm, U. and J. Breitung (2012). Testing for speculative bubbles in stock markets: A comparison of alternative methods. *Journal of Financial Econometrics 10*(1), 198–231.

Kurozumi, E. (2020). Asymptotic properties of bubble monitoring tests. *Econometric Reviews 39*(5), 510–538.

Otto, S. and J. Breitung (2023). Backward cusum for testing and monitoring structural change with an application to covid-19 pandemic data. *Econometric Theory 39*(4), 659–692.

Phillips, P. and T. Magdalinos (2007). Limit theory for moderate deviations from a unit root. *Journal of Econometrics 136*(1), 115–130.

Phillips, P. and P. Schmidt (1989). Testing for a unit root in the presence of deterministic trends. Technical report, Cowles Foundation for Research in Economics, Yale University.

Phillips, P., S. Shi, and J. Yu (2014). Specification sensitivity in right-tailed unit root testing for explosive behaviour. *Oxford Bulletin of Economics and Statistics 76*(3), 315–333.

Phillips, P., S. Shi, and J. Yu (2015). Testing for multiple bubbles: Historical episodes of exuberance and collapse in the S&P 500. *International economic review 56*(4), 1043–1078.

Phillips, P., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International Economic Review 52*(1), 201–226.

Schmidt, P. and J. Lee (1991). A modification of the Schmidt-Phillips unit root test. *Economics Letters 36*(3), 285–289.

Solo, V. (1984). The order of differencing in arima models. *Journal of the American Statistical Association 79*(388), 916–921.