

**TESTING FOR MULTIPLE BUBBLES:
HISTORICAL EPISODES OF EXUBERANCE
AND COLLAPSE IN THE S&P 500**

By

Peter C.B. Phillips, Shu-Ping Shi and Jun Yu

September 2013

COWLES FOUNDATION DISCUSSION PAPER NO. 1914



**COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281**

<http://cowles.econ.yale.edu/>

Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500*

Peter C. B. Phillips

*Yale University, University of Auckland,
University of Southampton & Singapore Management University*

Shu-Ping Shi

The Australian National University

Jun Yu

Singapore Management University

August 7, 2013

Abstract

Recent work on econometric detection mechanisms has shown the effectiveness of recursive procedures in identifying and dating financial bubbles. These procedures are useful as warning alerts in surveillance strategies conducted by central banks and fiscal regulators with real time data. Use of these methods over long historical periods presents a more serious econometric challenge due to the complexity of the nonlinear structure and break mechanisms that are inherent in multiple bubble phenomena within the same sample period. To meet this challenge the present paper develops a new recursive flexible window method that is better suited for practical implementation with long historical time series. The method is a generalized version of the sup ADF test of Phillips, Wu and Yu (2011, PWY) and delivers a consistent date-stamping strategy for the origination and termination of multiple bubbles. Simulations show that the test significantly improves discriminatory power and leads to distinct power gains when multiple bubbles occur. An empirical application of the methodology is conducted on S&P 500 stock market data over a long historical period from January 1871 to December 2010. The new approach successfully identifies the well-known historical episodes of exuberance and collapse over this period, whereas the strategy of PWY and a related CUSUM dating procedure locate far fewer episodes in the same sample range.

Keywords: Date-stamping strategy; Flexible window; Generalized sup ADF test; Multiple bubbles, Rational bubble; Periodically collapsing bubbles; Sup ADF test;

JEL classification: C15, C22

*The current paper and its technical companion “Testing for Multiple Bubbles: Limit Theory of Real Time Detectors” build on work that was originally circulated in 2011 in a long paper entitled “Testing for Multiple Bubbles” accompanied by a long supplement of technical results. We are grateful to Heather Anderson, Farshid Vahid, Tom Smith and Anthony Lynch for valuable discussions. Phillips acknowledges support from the NSF under Grant No. SES 09-56687 and SES 12-58258. Shi acknowledges the Financial Integrity Research Network (FIRN) for funding support. Peter C.B. Phillips email: peter.phillips@yale.edu. Shuping Shi, email: shuping.shi@anu.edu.au. Jun Yu, email: yujun@smu.edu.sg.

Economists have taught us that it is unwise and unnecessary to combat asset price bubbles and excessive credit creation. Even if we were unwise enough to wish to prick an asset price bubble, we are told it is impossible to see the bubble while it is in its inflationary phase. (George Cooper, 2008)

1 Introduction

As financial historians have argued recently (Ahamed, 2009; Ferguson, 2008), financial crises are often preceded by an asset market bubble or rampant credit growth. The global financial crisis of 2007-2009 is no exception. In its aftermath, central bank economists and policy makers are now affirming the recent Basil III accord to work to stabilize the financial system by way of guidelines on capital requirements and related measures to control “excessive credit creation”. In this process of control, an important practical issue of market surveillance involves the assessment of what is “excessive”. But as Cooper (2008) puts it in the header cited above from his recent bestseller, many economists have declared the task to be impossible and that it is imprudent to seek to combat asset price bubbles. How then can central banks and regulators work to offset a speculative bubble when they are unable to assess whether one exists and are considered unwise to take action if they believe one does exist?

One contribution that econometric techniques can offer in this complex exercise of market surveillance and policy action is the detection of exuberance in financial markets by explicit quantitative measures. These measures are not simply ex post detection techniques but anticipative dating algorithm that can assist regulators in their market monitoring behavior by means of early warning diagnostic tests. If history has a habit of repeating itself and human learning mechanisms do fail, as financial historians such as Ferguson (2008)¹ assert, then quantitative warnings may serve as useful alert mechanisms to both market participants and regulators.

Several attempts to develop econometric tests have been made in the literature going back some decades (see Gurkaynak, 2008, for a recent review). Phillips, Wu and Yu (2011, PWY hereafter) recently proposed a recursive method which can detect exuberance in asset price series

¹“Nothing illustrates more clearly how hard human beings find it to learn from history than the repetitive history of stock market bubbles.” Ferguson (2008).

during an inflationary phase. The approach is anticipative as an early warning alert system, so that it meets the needs of central bank surveillance teams and regulators, thereby addressing one of the key concerns articulated by Cooper (2008). The method is especially effective when there is a single bubble episode in the sample data, as in the 1990s Nasdaq episode analyzed in the PWY paper and in the 2000s U.S. house price bubble analyzed in Phillips and Yu (2011).

Just as historical experience confirms the existence of many financial crises (Ahamed reports 60 different financial crises since the 17th century²), when the sample period is long enough there will often be evidence of multiple asset price bubbles in the data. The econometric identification of multiple bubbles with periodically collapsing behavior over time is substantially more difficult than identifying a single bubble. The difficulty arises from the complex nonlinear structure involved in the multiple breaks that produce the bubble phenomena. Multiple breaks typically diminish the discriminatory power of existing test mechanisms such as the recursive tests given in PWY. These power reductions complicate attempts at econometric dating and enhance the need for new approaches that do not suffer from this problem. If econometric methods are to be useful in practical work conducted by surveillance teams they need to be capable of dealing with multiple bubble phenomena. Of particular concern in financial surveillance is the reliability of a warning alert system that points to inflationary upturns in the market. Such warning systems ideally need to have a low false detection rate to avoid unnecessary policy measures and a high positive detection rate that ensures early and effective policy implementation.

The present paper responds to this need by providing a new framework for testing and dating bubble phenomena when there may be multiple bubbles in the data. The mechanisms developed here extend those of PWY by allowing for flexible window widths in the recursive regressions on which the test procedures are based. The approach adopted in PWY uses a sup ADF (SADF) test based on sequence of forward recursive right-tailed ADF unit root tests. This procedure also gives rise to a dating strategy which identifies points of origination and termination of a bubble. When there is a single bubble in the data, it is known that this dating strategy is consistent,

² “Financial booms and busts were, and continue to be, a feature of the economic landscape. These bubbles and crises seem to be deep-rooted in human nature and inherent to the capitalist system. By one count there have been 60 different crises since the 17th century.” Ahamed (2009).

as was first shown in an unpublished working paper by Phillips and Yu (2009) whose results are subsumed as a special case within the present work. Other break testing procedures such as Chow tests, model selection, and CUSUM tests may also be applied as dating mechanisms. Extensive simulations conducted recently by Homm and Breitung (2012) indicate that the PWY procedure works satisfactorily against other recursive (as distinct from full sample) procedures for structural breaks and is particularly effective as a real time bubble detection algorithm. Importantly, the procedure can detect market exuberance arising from a variety of sources, including mildly explosive behavior that may be induced by changing fundamentals such as a time-varying discount factor.

When the sample period includes multiple episodes of exuberance and collapse, the SADF test may suffer from reduced power and can be inconsistent, thereby failing to reveal the existence of bubbles. This weakness is a particular drawback in analyzing long time series or rapidly changing market data where more than one episode of exuberance is suspected. To overcome this weakness and deal with multiple breaks of exuberance and collapse, the present paper proposes an alternative approach named the *generalized sup ADF* (GSADF) test. The GSADF test also relies on recursive right-tailed ADF tests but uses flexible window widths in the implementation. Instead of fixing the starting point of the recursion on the first observation, the GSADF test extends the sample coverage by changing both the starting point and the ending point of the recursion over a feasible range of flexible windows. Since the GSADF test covers more subsamples of the data and has greater window flexibility, it is designed to outperform the SADF test in detecting explosive behavior when multiple episodes occur in the data. This expected enhancement in performance by the GSADF test is demonstrated here in simulations which compare the two tests in terms of their size and power in bubble detection. The new procedure delivers a consistent dating mechanism when multiple bubbles occur, in contrast to the original version of the PWY dating strategy which can be inconsistent when multiple bubbles occur. The technique is therefore well suited to analyzing long historical time series.

A modified version of the original PWY dating algorithm is developed in which the detection procedure is repeated sequentially with re-initialization after the detection of each bubble.

Like the GSADF test, this sequential PWY algorithm works with subsamples of the data with different initializations in the recursions and therefore in theory is capable of detecting multiple bubbles. We also consider detection mechanism based on a recursive CUSUM test suggested recently in Homm and Breitung (2012).

An empirical application of these methodologies is conducted to S&P 500 stock market data over the period January 1871 to December 2010. The new GSADF approach successfully identifies all the well-known historical episodes of exuberance and collapse over this period, including the great crash, the post war boom in 1954, Black Monday in October 1987, the dot-com bubble and the subprime mortgage crisis. Several short episodes are also identified, including the famous banking panic of 1907, and the 1974 stock market crash. The strategy of PWY is much more conservative and locates only two episodes over the same historical period, catching the 1990s stock bubble but entirely missing the 2007-2008 subprime crisis. The sequential PWY algorithm is similarly conservative in detecting bubbles in this data set, as is the CUSUM procedure.

The organization of the paper is as follows. Section 2 discusses reduced form model specification issues for bubble testing, describes the new rolling window recursive test, and gives its limit theory. Section 3 proposes date-stamping strategies based on the new test and outlines their properties in single, multiple and no bubble scenarios. Section 4 reports the results of simulations investigating size, power, and performance characteristics of the various tests and dating strategies. In Section 5, the new procedures, the original PWY test, the sequential PWY test, and the CUSUM test are all applied to the S&P 500 price-dividend ratio data over 1871-2010. Section 6 concludes. Proofs are given in the Appendix. A companion paper (Phillips, Shi and Yu, 2013b) develops the limit theory and consistency properties of the dating procedures of the present paper covering both single and multiple bubble scenarios.

2 A Rolling Window Test for Identifying Bubbles

2.1 Models and Specification

A common starting point in the analysis of financial bubbles is the asset pricing equation:

$$P_t = \sum_{i=0}^{\infty} \left(\frac{1}{1+r_f} \right)^i \mathbb{E}_t (D_{t+i} + U_{t+i}) + B_t, \quad (1)$$

where P_t is the after-dividend price of the asset, D_t is the payoff received from the asset (i.e. dividend), r_f is the risk-free interest rate, U_t represents the unobservable fundamentals and B_t is the bubble component. The quantity $P_t^f = P_t - B_t$ is often called the market fundamental and B_t satisfies the submartingale property

$$\mathbb{E}_t (B_{t+1}) = (1+r_f) B_t. \quad (2)$$

In the absence of bubbles (i.e. $B_t = 0$), the degree of nonstationarity of the asset price is controlled by the character of the dividend series and unobservable fundamentals. For example, if D_t is an $I(1)$ process and U_t is either an $I(0)$ or an $I(1)$ process, then the asset price is at most an $I(1)$ process. On the other hand, given (2), asset prices will be explosive in the presence of bubbles. Therefore, when unobservable fundamentals are at most $I(1)$ and D_t is stationary after differencing, empirical evidence of explosive behavior in asset prices may be used to conclude the existence of bubbles.³

The pricing equation (1) is not the only model to accommodate bubble phenomena and there is continuing professional debate over how (or even whether) to include bubble components in asset pricing models (see, for example, the discussion in Cochrane, 2005, pp. 402-404) and their relevance in empirical work (notably, Pástor and Veronesi, 2006, but note also the strong critique

³This argument also applies to the logarithmic asset price and the logarithmic dividend under certain conditions. This is due to the fact that in the absence of bubbles, equation (1) can be rewritten as

$$(1-\rho) p_t^f = \kappa + \rho e^{\bar{d}-\bar{p}} d_t + \rho e^{\bar{u}-\bar{p}} u_t + e^{\bar{d}-\bar{p}} \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t [\Delta d_{t+j}] + e^{\bar{u}-\bar{p}} \sum_{j=1}^{\infty} \rho^j \mathbb{E}_t [\Delta u_{t+j}],$$

where $p_t^f = \log(P_t^f)$, $d_t = \log(D_t)$, $u_t = \log(U_t)$, $\rho = (1+r_f)^{-1}$, κ is a constant, \bar{p} , \bar{d} and \bar{u} are the respective sample means of p_t^f , d_t and u_t . The degree of nonstationarity of p_t^f is determined by that of d_t and u_t . Lee and Phillips (2011) provide a detailed analysis of the accuracy of this log linear approximation under various conditions.

of that view in Cooper, 2008⁴). There is greater agreement on the existence of market exuberance (which may be rational or irrational depending on possible links to market fundamentals), crises and panics (Kindelberger and Aliber, 2005; Ferguson, 2008). For instance, financial exuberance might originate in pricing errors relative to fundamentals that arise from behavioral factors, or fundamental values may themselves be highly sensitive to changes in the discount rate, which can lead to price run ups that mimic the inflationary phase of a bubble. With regard to the latter, Phillips and Yu (2011) show that in certain dynamic structures a time-varying discount rate can induce temporary explosive behavior in asset prices. Similar considerations may apply in more general stochastic discount factor asset pricing equations. Whatever its origins, explosive or mildly explosive (Phillips and Magdalinos, 2007) behavior in asset prices is a primary indicator of market exuberance during the inflationary phase of a bubble and it is this time series manifestation that may be subjected to econometric testing using recursive testing procedures like the right sided unit root tests in PWY. As discussed above, recursive right sided unit root tests seem to be particularly effective as real time detection mechanisms for mildly explosive behavior and market exuberance.

The PWY test is a reduced form approach to bubble detection. In such tests (as distinct from left sided unit root tests), the focus is usually on the alternative hypothesis (rather than the martingale or unit root hypothesis) because of interest in possible departures from fundamentals and the presence of market excesses or mispricing. Right sided unit root tests, as discussed in PWY, are informative about mildly explosive or submartingale behavior in the data and are therefore useful as a form of market diagnostic or warning alert.

As with all testing procedures, model specification under the null is important for estimation purposes, not least because of the potential impact on asymptotic theory and the critical values used in testing. Unit root testing is a well known example where intercepts, deterministic trends, or trend breaks all materially impact the limit theory. Such issues also arise in right-tailed unit root tests of the type used in bubble detection, as studied recently in Phillips, Shi and Yu (2013a;

⁴“People outside the world of economics may be amazed to know that a significant body of researchers are still engaged in the task of proving that the pricing of the NASDAQ stock market correctly reflected the market’s true value throughout the period commonly known as the NASDAQ bubble.... The intellectual contortions required to rationalize all of these prices beggars belief.” (Cooper, 2008, p.9).

PSY1). Their analysis allowed for a martingale null with an asymptotically negligible drift to capture the mild drift in price processes that are often empirically realistic over long historical periods. The prototypical model of this type has the following weak (local to zero) intercept form

$$y_t = dT^{-\eta} + \theta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2), \quad \theta = 1 \quad (3)$$

where d is a constant, T is the sample size, and the parameter η is a localizing coefficient that controls the magnitude of the intercept and drift as $T \rightarrow \infty$. Solving (3) gives $y_t = d\frac{t}{T^\eta} + \sum_{j=1}^t \varepsilon_j + y_0$ revealing the deterministic drift dt/T^η . When $\eta > 0$ the drift is small relative to a linear trend, when $\eta > \frac{1}{2}$, the drift is small relative to the martingale component, and when $\eta = \frac{1}{2}$ the standardized output $T^{-1/2}y_t$ behaves asymptotically like a Brownian motion with drift which suits many macroeconomic and financial time series. The null specification (3) includes the pure random walk null of PWY as a special case when $\eta \rightarrow \infty$ and the order of magnitude of y_t is then identical to that of a pure random walk. Estimation of the localizing coefficient η is discussed in PSY1.⁵

The model specification (3) is usually complemented with transient dynamics in order to conduct tests for exuberance, just as in standard ADF unit root testing against stationarity. The recursive approach that we now suggest involves a rolling window ADF style regression implementation based on such a system. In particular, suppose the rolling window regression sample starts from the r_1^{th} fraction of the total sample (T) and ends at the r_2^{th} fraction of the sample, where $r_2 = r_1 + r_w$ and r_w is the (fractional) window size of the regression. The empirical regression model can then be written as

$$\Delta y_t = \alpha_{r_1, r_2} + \beta_{r_1, r_2} y_{t-1} + \sum_{i=1}^k \psi_{r_1, r_2}^i \Delta y_{t-i} + \varepsilon_t, \quad (4)$$

where k is the lag order and $\varepsilon_t \stackrel{iid}{\sim} (0, \sigma_{r_1, r_2}^2)$. The number of observations in the regression is $T_w = \lfloor Tr_w \rfloor$, where $\lfloor \cdot \rfloor$ is the floor function (giving the integer part of the argument). The ADF statistic (t-ratio) based on this regression is denoted by $ADF_{r_1}^{r_2}$.

⁵When $\eta > 0.5$ the drift component is dominated by the stochastic trend and estimates of η typically converge to $1/2$, corresponding to the order of the stochastic trend. When $\eta \in [0, \frac{1}{2}]$, the parameter is consistently estimable, although only at a slow logarithmic rate when $\eta = \frac{1}{2}$. See PSY1 for details.

We proceed to use rolling regressions of this type to construct a new approach to bubble detection that is particularly useful in the case of multiple bubbles in the sample. The formulation includes the earlier SADF test procedure developed and used in PWY, which we now briefly review together with some other recursive and regression switching procedures.

2.2 The SADF Test of PWY

The SADF test relies on repeated estimation of the ADF model on a forward expanding sample sequence and the test is obtained as the sup value of the corresponding ADF statistic sequence. In this case, the window size r_w expands from r_0 to 1, so that r_0 is the smallest sample window width fraction (initializing computation) and 1 is the largest window fraction (the total sample size) in the recursion. The starting point r_1 of the sample sequence is fixed at 0, so the end point of each sample (r_2) equals r_w , and changes from r_0 to 1. The ADF statistic for a sample that runs from 0 to r_2 is denoted by $ADF_0^{r_2}$. The SADF statistic is defined as

$$SADF(r_0) = \sup_{r_2 \in [r_0, 1]} ADF_0^{r_2}.$$

The SADF test and other right-sided unit root tests are not the only method of detecting explosive behavior. An alternative approach is the two-regime Markov-switching unit root test of Hall, Psaradakis and Sola (1999). While this procedure offers some appealing features like regime probability estimation, recent simulation work by Shi (2012) reveals that the Markov switching model is susceptible to false detection or spurious explosiveness. In addition, when allowance is made for a regime-dependent error variance as in Funke, Hall and Sola (1994) and van Norden and Vigfusson (1998), filtering algorithms can find it difficult to distinguish periods which may appear spuriously explosive due to high variance and periods when there is genuine explosive behavior. Furthermore, the bootstrapping procedure embedded in the Markov switching unit root test is computationally burdensome as Psaradakis, Sola and Spagnolo (2001) pointed out. These pitfalls make the Markov switching unit root test a difficult and somewhat unreliable tool of financial surveillance.

Other econometric approaches may be adapted to use the same recursive feature of the SADF test, such as the modified Bhargava statistic (Bhargava, 1986), the modified Busetti-

Taylor statistic (Busetti and Taylor, 2004), and the modified Kim statistic (Kim, 2000). These tests are considered in Homm and Breitung (2012) for bubble detection and all share the spirit of the SADF test of PWY. That is, the statistic is calculated recursively and then the sup functional of the recursive statistics is calculated for testing. Since all these tests are similar in character to the SADF test and since Homm and Breitung (2012) found in their simulations that the PWY test was the most powerful in detecting multiple bubbles, we focus attention in this paper on extending the SADF test. However, our simulations and empirical implementation provide comparative results with the CUSUM procedure in view of its good overall performance recorded in the Homm and Breitung simulations.

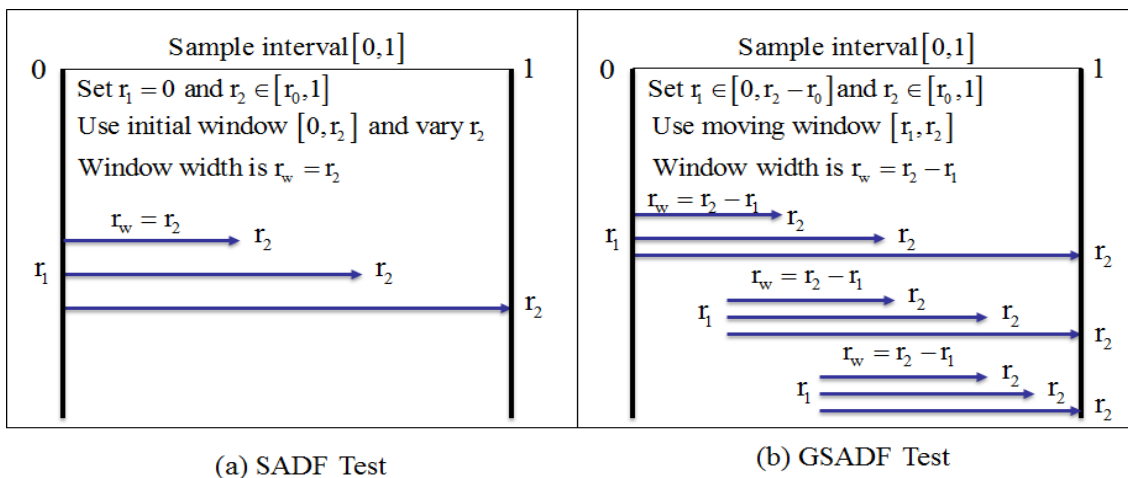


Fig. 1: The sample sequences and window widths of the SADF test and the GSADF test.

2.3 The Rolling Window GSADF Test

The GSADF test developed here continues the idea of repeatedly running the ADF test regression (4) on subsamples of the data in a recursive fashion. However, the subsamples used in the recursion are much more extensive than those of the SADF test. Besides varying the end point of the regression r_2 from r_0 (the minimum window width) to 1, the GSADF test allows the starting point r_1 in (4) to change within a feasible range, i.e. from 0 to $r_2 - r_0$. We define the GSADF statistic to be the largest ADF statistic over all feasible ranges of r_1 and r_2 , and we

denote this statistic by $GSADF(r_0)$. That is,

$$GSADF(r_0) = \sup_{\substack{r_2 \in [r_0, 1] \\ r_1 \in [0, r_2 - r_0]}} \{ADF_{r_1}^{r_2}\}.$$

Fig. 1 illustrates the comparative sample sequences used in the recursive SADF and GSADF procedures.

Theorem 1 *When the regression model includes an intercept and the null hypothesis is a random walk with an asymptotically negligible drift (i.e. $dT^{-\eta}$ with $\eta > 1/2$ and constant d) as in (3), the limit distribution of the GSADF test statistic is:*

$$\sup_{\substack{r_2 \in [r_0, 1] \\ r_1 \in [0, r_2 - r_0]}} \left\{ \frac{\frac{1}{2}r_w [W(r_2)^2 - W(r_1)^2 - r_w] - \int_{r_1}^{r_2} W(r) dr [W(r_2) - W(r_1)]}{r_w^{1/2} \left\{ r_w \int_{r_1}^{r_2} W(r)^2 dr - \left[\int_{r_1}^{r_2} W(r) dr \right]^2 \right\}^{1/2}} \right\} \quad (5)$$

where $r_w = r_2 - r_1$ and W is a standard Wiener process. The limit theory continues to hold when the null is a unit root process with asymptotically negligible drift and innovations satisfying the error condition EC in the Appendix.

The proof of Theorem 1 is given in the Appendix. The limit distribution of the GSADF statistic is identical to the case where the regression model includes an intercept and the null hypothesis is a random walk or unit root process without drift. The usual limit distribution of the ADF statistic is a special case of equation (5) with $r_1 = 0$ and $r_2 = r_w = 1$ while the limit distribution of the SADF statistic is a further special case of (5) with $r_1 = 0$ and $r_2 = r_w \in [r_0, 1]$ (see PSY1).

Similar to the limit theory of the SADF statistic, the asymptotic GSADF distribution depends on the smallest window size r_0 . In practice, r_0 needs to be chosen according to the total number of observations T . If T is small, r_0 needs to be large enough to ensure there are enough observations for adequate initial estimation. If T is large, r_0 can be set to be a smaller number so that the test does not miss any opportunity to detect an early explosive episode. In our empirical application we use $r_0 = 36/1680$, corresponding to around 2% of the data.

Critical values of the SADF and GSADF statistics are displayed in Table 1. The asymptotic critical values are obtained by numerical simulations, where the Wiener process is approximated by partial sums of 2,000 independent $N(0, 1)$ variates and the number of replications is 2,000. The finite sample critical values are obtained from 5,000 Monte Carlo replications. The lag order k is set to zero. The parameters (d and η) in the null hypothesis are set to unity.⁶

Table 1: Critical values of the SADF and GSADF tests against an explosive alternative

<i>(a) The asymptotic critical values</i>						
	$r_0 = 0.4$		$r_0 = 0.2$		$r_0 = 0.1$	
	SADF	GSADF	SADF	GSADF	SADF	GSADF
90%	0.86	1.25	1.04	1.66	1.18	1.89
95%	1.18	1.56	1.38	1.92	1.49	2.14
99%	1.79	2.18	1.91	2.44	2.01	2.57
<i>(b) The finite sample critical values</i>						
	$T = 100$ and $r_0 = 0.4$		$T = 200$ and $r_0 = 0.4$		$T = 400$ and $r_0 = 0.4$	
	SADF	GSADF	SADF	GSADF	SADF	GSADF
90%	0.72	1.16	0.75	1.21	0.78	1.27
95%	1.05	1.48	1.08	1.52	1.10	1.55
99%	1.66	2.08	1.75	2.18	1.75	2.12
<i>(c) The finite sample critical values</i>						
	$T = 100$ and $r_0 = 0.4$		$T = 200$ and $r_0 = 0.2$		$T = 400$ and $r_0 = 0.1$	
	SADF	GSADF	SADF	GSADF	SADF	GSADF
90%	0.72	1.16	0.97	1.64	1.19	1.97
95%	1.05	1.48	1.30	1.88	1.50	2.21
99%	1.66	2.08	1.86	2.46	1.98	2.71

Note: the asymptotic critical values are obtained by numerical simulations with 2,000 iterations. The Wiener process is approximated by partial sums of $N(0, 1)$ with 2,000 steps. The finite sample critical values are obtained from the 5,000 Monte Carlo simulations. The parameters, d and η , are set to unity.

We observe the following phenomena. First, as the minimum window size r_0 decreases, critical values of the test statistic (including the SADF statistic and the GSADF statistic) increase. For instance, when r_0 decreases from 0.4 to 0.1, the 95% asymptotic critical value of the GSADF statistic rises from 1.56 to 2.14 and the 95% finite sample critical value of the test statistic for sample size 400 increases from 1.48 to 2.21. Second, for a given r_0 , the finite sample

⁶From PSY1, we know that when $d = 1$ and $\eta > 1/2$, the finite sample distribution of the SADF statistic is almost invariant to the value of η .

critical values of the test statistic are almost invariant. Notice that they are very close to the corresponding asymptotic critical values, indicating that the asymptotic critical values may well be used in practical work.⁷

Third, critical values for the GSADF statistic are larger than those of the SADF statistic. As a case in point, when $T = 400$ and $r_0 = 0.1$, the 95% critical value of the GSADF statistic is 2.21 while that of the SADF statistic is 1.50. Fig. 2 shows the asymptotic distribution of the ADF , $SADF(0.1)$ and $GSADF(0.1)$ statistics. The distributions move sequentially to the right and have greater concentration in the order ADF , $SADF(0.1)$ and $GSADF(0.1)$.

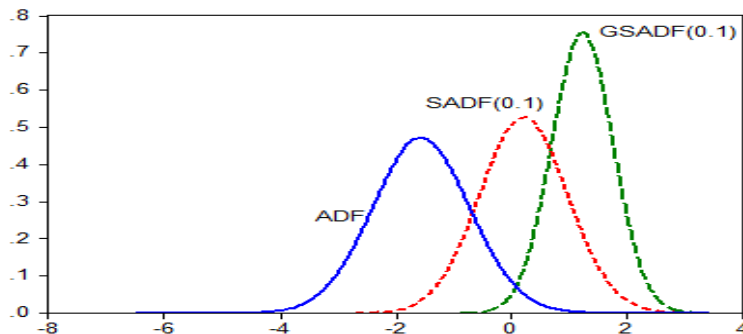


Fig. 2: Asymptotic distributions of the ADF and supADF statistics ($r_0 = 0.1$).

3 Date-stamping Strategies

As discussed in the Introduction, regulators and central banks concerned with practical policy implementation need to assess whether real time data provide evidence of financial exuberance - specifically whether any particular observation such $\tau = \lfloor Tr \rfloor$ belongs to a bubble phase in the overall trajectory. The strategy suggested in PWY is to conduct a right-tailed ADF test using information up to this observation (i.e. information embodied in $I_{\lfloor Tr \rfloor} = \{y_1, y_2, \dots, y_{\lfloor Tr \rfloor}\}$). Since it is possible that the data $I_{\lfloor Tr \rfloor}$ may include one or more collapsing bubble episodes, the ADF test, like earlier unit root/cointegration-based tests for bubbles (e.g., Diba and Grossman, 1988), may result in finding *pseudo stationary* behavior. The strategy recommended here is to

⁷Finite sample critical values are used for accuracy in the simulations and the later empirical applications.

perform a *backward sup ADF test* on $I_{[Tr]}$ to improve identification accuracy. We use a similar flexible window recursion as that described above.

In particular, the backward SADF test performs a sup ADF test on a backward expanding sample sequence where the end point of each sample is fixed at r_2 and the start point varies from 0 to $r_2 - r_0$. The corresponding ADF statistic sequence is $\{ADF_{r_1}^{r_2}\}_{r_1 \in [0, r_2 - r_0]}$. The backward SADF statistic is defined as the sup value of the ADF statistic sequence over this interval, viz.,

$$BSADF_{r_2}(r_0) = \sup_{r_1 \in [0, r_2 - r_0]} \{ADF_{r_1}^{r_2}\}.$$

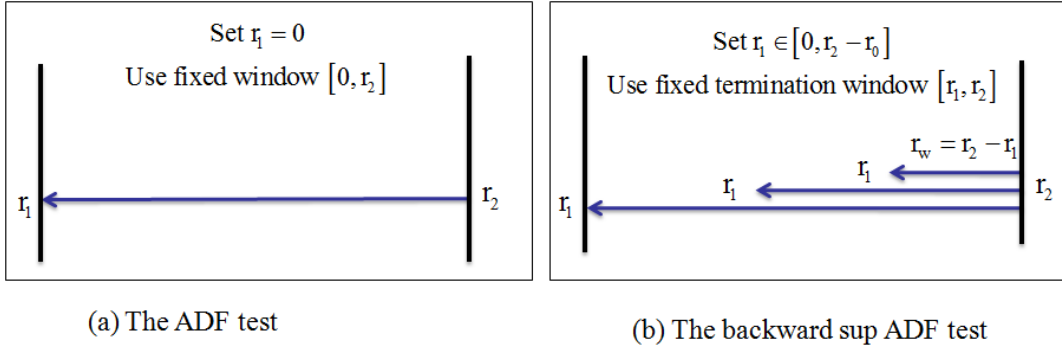


Fig. 3: The sample sequences of the ADF test and the backward SADF test.

The ADF test is a special case of the backward sup ADF test with $r_1 = 0$. We denote the corresponding ADF statistic by ADF_{r_2} . Fig. 3 illustrates the difference between the simple ADF test and the backward SADF test recursion. PWY propose comparing ADF_{r_2} with the (right-tail) critical values of the standard ADF statistic to identify explosiveness at observation $[Tr_2]$. The feasible range of r_2 runs from r_0 to 1. The origination date of a bubble $[Tr_e]$ is calculated as the first chronological observation whose ADF statistic exceeds the critical value. We denote the calculated origination date by $[T\hat{r}_e]$. The estimated termination date of a bubble $[T\hat{r}_f]$ is the first chronological observation after $[T\hat{r}_e] + \log(T)$ whose ADF statistic goes below the critical value. PWY impose a condition that for a bubble to exist its duration must exceed a slowly varying (at infinity) quantity such as $L_T = \log(T)$. This requirement helps to exclude short lived blips in the fitted autoregressive coefficient and, as discussed below, can be adjusted to take into account the data frequency. The dating estimates are then delivered by the crossing

time formulae

$$\hat{r}_e = \inf_{r_2 \in [r_0, 1]} \left\{ r_2 : ADF_{r_2} > cv_{r_2}^{\beta_T} \right\} \text{ and } \hat{r}_f = \inf_{r_2 \in [\hat{r}_e + \log(T)/T, 1]} \left\{ r_2 : ADF_{r_2} < cv_{r_2}^{\beta_T} \right\}, \quad (6)$$

where $cv_{r_2}^{\beta_T}$ is the 100 $(1 - \beta_T)$ % critical value of the ADF statistic based on $[Tr_2]$ observations. The significance level β_T depends on the sample size T and it is assumed that $\beta_T \rightarrow 0$ as $T \rightarrow \infty$. This control ensures that $cv_{r_2}^{\beta_T}$ diverges to infinity and thereby eliminates the type I error as $T \rightarrow \infty$. In empirical applications, however, β_T will often be fixed at some level such as 0.05 rather than using drifting significance levels.

The new strategy suggests that inference about explosiveness of the process at observation $[Tr_2]$ be based on the backward sup ADF statistic, $BSADF_{r_2}(r_0)$. We define the origination date of a bubble as the first observation whose backward sup ADF statistic exceeds the critical value of the backward sup ADF statistic. The termination date of a bubble is calculated as the first observation after $[T\hat{r}_e] + \delta \log(T)$ whose backward sup ADF statistic falls below the critical value of the backward sup ADF statistic. Here it is assumed that the duration of the bubble exceeds $\delta \log(T)$, where δ is a frequency dependent parameter.⁸ The (fractional) origination and termination points of a bubble (i.e. r_e and r_f) are calculated according to the following first crossing time equations:

$$\hat{r}_e = \inf_{r_2 \in [r_0, 1]} \left\{ r_2 : BSADF_{r_2}(r_0) > scv_{r_2}^{\beta_T} \right\}, \quad (7)$$

$$\hat{r}_f = \inf_{r_2 \in [\hat{r}_e + \delta \log(T)/T, 1]} \left\{ r_2 : BSADF_{r_2}(r_0) < scv_{r_2}^{\beta_T} \right\}, \quad (8)$$

where $scv_{r_2}^{\beta_T}$ is the 100 $(1 - \beta_T)$ % critical value of the sup ADF statistic based on $[Tr_2]$ observations. Analogously, the significance level β_T depends on the sample size T and it goes to zero as the sample size approaches infinity.

⁸For instance, one might wish to impose a minimal condition that to be classified as a bubble its duration should exceed a certain period such as one year (which is inevitably arbitrary). Then, when the sample size is 30 years (360 months), δ is 0.7 for yearly data and 5 for monthly data.

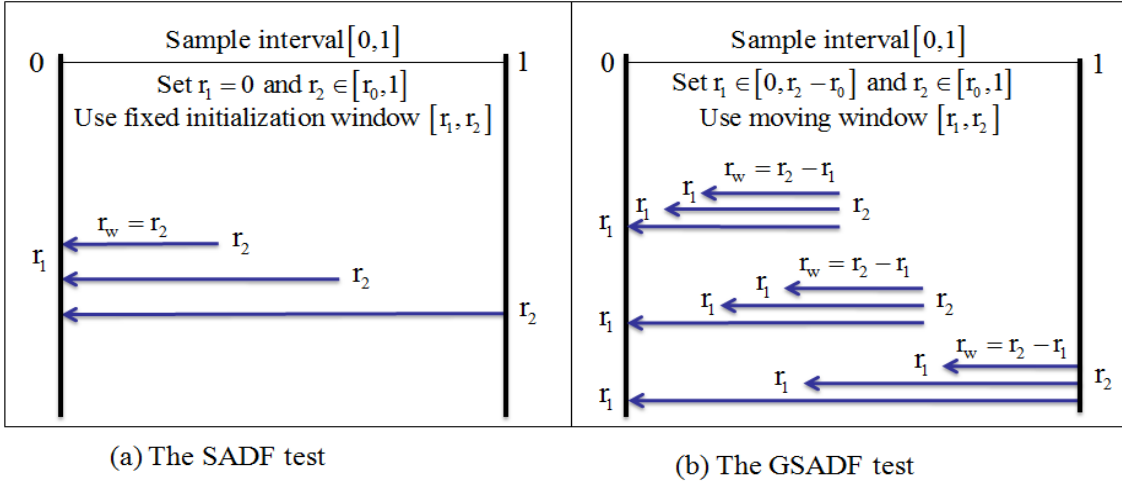


Fig. 4: An alternative illustration of the sample sequences and window widths of the SADF test and the GSADF test.

The SADF test is based on repeated implementation of the ADF test for each $r_2 \in [r_0, 1]$. The GSADF test implements the backward sup ADF test repeatedly for each $r_2 \in [r_0, 1]$ and makes inferences based on the sup value of the backward sup ADF statistic sequence, $\{BSADF_{r_2}(r_0)\}_{r_2 \in [r_0, 1]}$. Hence, the SADF and GSADF statistics can respectively be written as

$$SADF(r_0) = \sup_{r_2 \in [r_0, 1]} \{ADF_{r_2}\},$$

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

Thus, the PWY date-stamping strategy corresponds to the SADF test and the new strategy corresponds to the GSADF test. The essential features of the two tests are shown in stylized form in the diagrams of Fig. 4.

3.1 Asymptotic properties of the dating algorithms

The limit theory of these date-stamping strategies requires very detailed calculations which are provided in our companion paper (Phillips, Shi and Yu, 2013b; PSY2). Additional technical material needed for those derivations is contained in the online supplement to that paper. The

main results and import of the theory for empirical practice are reviewed here. We look in turn at cases where there are no bubbles, a single bubble, and multiple bubbles in the data.

No bubbles Under the null hypothesis of the no bubble episodes in the data the asymptotic distributions of the ADF and SADF statistics follow from Theorem 1. The backward ADF test with observation $[Tr_2]$ is a special case of the GSADF test with $r_1 = 0$ and a fixed r_2 and the backward sup ADF test is a special case of the GSADF test with a fixed r_2 and $r_1 = r_2 - r_w$. Therefore, from the limit theory given in (5), we have the following asymptotic distributions of these two statistics

$$F_{r_2}(W) := \frac{\frac{1}{2}r_2 \left[W(r_2)^2 - r_2 \right] - \int_0^{r_2} W(r) dr W(r_2)}{r_2^{1/2} \left\{ r_2 \int_0^{r_2} W(r)^2 dr - \left[\int_0^{r_2} W(r) dr \right]^2 \right\}^{1/2}},$$

$$F_{r_2}^{r_0}(W) := \sup_{\substack{r_1 \in [0, r_2 - r_0] \\ r_w = r_2 - r_1}} \left\{ \frac{\frac{1}{2}r_w \left[W(r_2)^2 - W(r_1)^2 - r_w \right] - \int_{r_1}^{r_2} W(r) dr [W(r_2) - W(r_1)]}{r_w^{1/2} \left\{ r_w \int_{r_1}^{r_2} W(r)^2 dr - \left[\int_{r_1}^{r_2} W(r) dr \right]^2 \right\}^{1/2}} \right\}.$$

Define cv^{β_T} as the $100(1 - \beta_T)\%$ quantile of $F_{r_2}(W)$ and scv^{β_T} as the $100(1 - \beta_T)\%$ quantile of $F_{r_2}^{r_0}(W)$. We know that $cv^{\beta_T} \rightarrow \infty$ and $scv^{\beta_T} \rightarrow \infty$ as $\beta_T \rightarrow 0$. Given $cv^{\beta_T} \rightarrow \infty$ and $scv^{\beta_T} \rightarrow \infty$ under the null hypothesis of no bubbles, the probabilities of (falsely) detecting the origination of bubble expansion and the termination of bubble collapse using the backward ADF statistic and the backward sup ADF statistic tend to zero, so that both $\Pr\{\hat{r}_e \in [r_0, 1]\} \rightarrow 0$ and $\Pr\{\hat{r}_f \in [r_0, 1]\} \rightarrow 0$.

One bubble PSY2 study the consistency properties of the date estimates \hat{r}_e and \hat{r}_f under various alternatives. The simplest is a single bubble episode, like that considered in PWY. The following generating process used in PWY is an effective reduced form mechanism that switches between a martingale mechanism, a single mildly explosive episode, collapse, and subsequent renewal of martingale behavior:

$$X_t = X_{t-1}1\{t < \tau_e\} + \delta_T X_{t-1}1\{\tau_e \leq t \leq \tau_f\}$$

$$+ \left(\sum_{k=\tau_f+1}^t \varepsilon_k + X_{\tau_f}^* \right) 1\{t > \tau_f\} + \varepsilon_t 1\{j \leq \tau_f\}. \quad (9)$$

In (9) $\delta_T = 1 + cT^{-\alpha}$ with $c > 0$ and $\alpha \in (0, 1)$, $\varepsilon_t \stackrel{iid}{\sim} (0, \sigma^2)$, $X_{\tau_f}^* = X_{\tau_e} + X^*$ with $X^* = O_p(1)$, $\tau_e = \lfloor T r_e \rfloor$ dates the origination of bubble expansion and $\tau_f = \lfloor T r_f \rfloor$ dates the termination of bubble collapse. The pre-bubble period $N_0 = [1, \tau_e)$ is assumed to be a pure random walk process but this is not essential to the asymptotic theory. The bubble expansion period $B = [\tau_e, \tau_f]$ is a mildly explosive process with expansion rate given by the autoregressive (AR) coefficient δ_T . As discussed in PWY, mildly explosive processes are well suited to capturing market exuberance. The process then collapses abruptly to $X_{\tau_f}^*$, which equals X_{τ_e} plus a small perturbation, and continues its random wandering martingale path over the subsequent period $N_1 = (\tau_f, \tau]$. Of course, more general models with various transitional collapse mechanisms can also be considered. The prototypical system (9) captures the main features of interest when there is a single bubble episode and is useful in analyzing test properties for a bubble alternative.

Under (9) and certain rate conditions both the ADF and BSADF detectors provide consistent estimates of the origination and termination dates of the bubble.⁹ When the point estimates \hat{r}_e and \hat{r}_f are obtained as in PWY using the ADF test and the first crossing times (6) then $(\hat{r}_e, \hat{r}_f) \xrightarrow{P} (r_e, r_f)$ as $T \rightarrow \infty$ provided the following rate condition on the critical value cv^{β_T} holds

$$\frac{1}{cv^{\beta_T}} + \frac{cv^{\beta_T}}{T^{1/2}\delta_T^{\tau-\tau_e}} \rightarrow 0, \text{ as } T \rightarrow \infty. \quad (10)$$

Consistency of (\hat{r}_e, \hat{r}_f) was first proved in a working paper (Phillips and Yu, 2009). When the point estimates \hat{r}_e and \hat{r}_f are obtained from the BSADF detector using the crossing time criteria (7) - (8), we again have consistency $(\hat{r}_e, \hat{r}_f) \xrightarrow{P} (r_e, r_f)$ as $T \rightarrow \infty$ under the corresponding rate condition on the critical value scv^{β_T} , viz.,

$$\frac{1}{scv^{\beta_T}} + \frac{scv^{\beta_T}}{T^{1/2}\delta_T^{\tau-\tau_e}} \rightarrow 0, \text{ as } T \rightarrow \infty. \quad (11)$$

Hence both strategies consistently estimate the origination and termination points when there is only a single bubble episode in the sample period. The rate conditions (10) and (11)

⁹Consistent estimation of r_e also requires that the minimum window size $r_0 \leq r_e$, else r_e is unidentified.

require for consistency of (\hat{r}_e, \hat{r}_f) that $(cv^{\beta_T}, scv^{\beta_T})$ pass to infinity and that their orders of magnitude be smaller than $T^{1/2}\delta_T^{\tau-\tau_e}$. It is sufficient for consistency of (\hat{r}_e, \hat{r}_f) that the critical values cv^{β_T} and scv^{β_T} used in the recursions expand slowly as $T \rightarrow \infty$, for example at the slowly varying rate $\log(T)$. The probability of false rejection of normal behavior then goes to zero. The upper rate condition that delimits the rate at which $(cv^{\beta_T}, scv^{\beta_T})$ pass to infinity ensures the successful detection of mildly explosive behavior under the alternative. In effect, the critical values used in the crossing times (6) and (7) must not pass to infinity too fast relative to the strength of exuberance in the data which is governed by the value of the localizing parameter $\alpha < 1$ in the AR coefficient $\delta_T = 1 + cT^{-\alpha}$.

Multiple bubbles Multiple bubble episodes may be analyzed in a similar way using more complex alternative models and more detailed calculations, which are reported in PSY2. The key outcomes are revealed in the case of two bubble episodes, which are generated in the following system extending the prototypical model (9):

$$\begin{aligned}
X_t = & X_{t-1}1\{t \in N_0\} + \delta_T X_{t-1}1\{t \in B_1 \cup B_2\} + \left(\sum_{k=\tau_{1f}+1}^t \varepsilon_k + X_{\tau_{1f}}^* \right) 1\{t \in N_1\} \\
& + \left(\sum_{l=\tau_{2f}+1}^t \varepsilon_l + X_{\tau_{2f}}^* \right) 1\{t \in N_2\} + \varepsilon_t 1\{j \in N_0 \cup B_1 \cup B_2\}, \tag{12}
\end{aligned}$$

In (12) we use the notation $N_0 = [1, \tau_{1e})$, $B_1 = [\tau_{1e}, \tau_{1f}]$, $N_1 = (\tau_{1f}, \tau_{2e})$, $B_2 = [\tau_{2e}, \tau_{2f}]$ and $N_2 = (\tau_{2f}, \tau]$. The observations $\tau_{1e} = \lfloor Tr_{1e} \rfloor$ and $\tau_{1f} = \lfloor Tr_{1f} \rfloor$ are the origination and termination dates of the first bubble; $\tau_{2e} = \lfloor Tr_{2e} \rfloor$ and $\tau_{2f} = \lfloor Tr_{2f} \rfloor$ are the origination and termination dates of the second bubble; and τ is the last observation of the sample. After the collapse of the first bubble, X_t resumes a martingale path until time $\tau_{2e} - 1$ and a second episode of exuberance begins at τ_{2e} . The expansion process lasts until τ_{2f} and collapses to a value of $X_{\tau_{2f}}^*$. The process then continues on a martingale path until the end of the sample period τ . The expansion duration of the first bubble is assumed to be longer than that of the second bubble, namely $\tau_{1f} - \tau_{1e} > \tau_{2f} - \tau_{2e}$. Obvious extensions of (12) include models where the mildly explosive coefficient δ_T takes different values in regimes B_1 and B_2 , and models where

the transition mechanisms to martingale behavior over N_1 and N_2 take more graduated and possibly different forms, thereby distinguishing the bubble mechanisms in the two cases.

The date-stamping strategy of PWY suggests calculating r_{1e} , r_{1f} , r_{2e} and r_{2f} from the following equations (based on the ADF statistic):

$$\hat{r}_{1e} = \inf_{r_2 \in [r_0, 1]} \left\{ r_2 : ADF_{r_2} > cv_{r_2}^{\beta_T} \right\} \text{ and } \hat{r}_{1f} = \inf_{r_2 \in [\hat{r}_{1e} + \log(T)/T, 1]} \left\{ r_2 : ADF_{r_2} < cv_{r_2}^{\beta_T} \right\}, \quad (13)$$

$$\hat{r}_{2e} = \inf_{r_2 \in [\hat{r}_{1f}, 1]} \left\{ r_2 : ADF_{r_2} > cv_{r_2}^{\beta_T} \right\} \text{ and } \hat{r}_{2f} = \inf_{r_2 \in [\hat{r}_{2e} + \log(T)/T, 1]} \left\{ r_2 : ADF_{r_2} < cv_{r_2}^{\beta_T} \right\}, \quad (14)$$

where the duration of the bubble periods is restricted to be longer than $\log(T)$. The new strategy recommends using the backward sup ADF test and calculating the origination and termination points according to the following equations:

$$\hat{r}_{1e} = \inf_{r_2 \in [r_0, 1]} \left\{ r_2 : BSADF_{r_2}(r_0) > scv_{r_2}^{\beta_T} \right\}, \quad (15)$$

$$\hat{r}_{1f} = \inf_{r_2 \in [\hat{r}_{1e} + \delta \log(T)/T, 1]} \left\{ r_2 : BSADF_{r_2}(r_0) < scv_{r_2}^{\beta_T} \right\}, \quad (16)$$

$$\hat{r}_{2e} = \inf_{r_2 \in [\hat{r}_{1f}, 1]} \left\{ r_2 : BSADF_{r_2}(r_0) > scv_{r_2}^{\beta_T} \right\}, \quad (17)$$

$$\hat{r}_{2f} = \inf_{r_2 \in [\hat{r}_{2e} + \delta \log(T)/T, 1]} \left\{ r_2 : BSADF_{r_2}(r_0) < scv_{r_2}^{\beta_T} \right\}. \quad (18)$$

An alternative implementation of the PWY procedure is to use that procedure sequentially, namely to detect one bubble at a time and sequentially re-apply the algorithm. The dating criteria for the first bubble remain the same (i.e. equation (13)). Conditional on the first bubble having been found and terminated at \hat{r}_{1f} , the following dating criteria are used to date stamp a second bubble:

$$\hat{r}_{2e} = \inf_{r_2 \in (\hat{r}_{1f} + \varepsilon_T, 1]} \left\{ r_2 : \hat{r}_{1f} ADF_{r_2} > cv_{r_2}^{\beta_T} \right\} \text{ and } \hat{r}_{2f} = \inf_{r_2 \in [\hat{r}_{2e} + \log(T)/T, 1]} \left\{ r_2 : \hat{r}_{1f} ADF_{r_2} < cv_{r_2}^{\beta_T} \right\}, \quad (19)$$

where $\hat{r}_{1f} ADF_{r_2}$ is the ADF statistic calculated over $(\hat{r}_{1f}, r_2]$. This sequential application of the PWY procedure requires a few observations in order to re-initialize the test process (i.e. $r_2 \in (\hat{r}_{1f} + \varepsilon_T, 1]$ for some $\varepsilon_T > 0$) after a bubble.

The asymptotic behavior of these various dating estimates is developed in PSY2 and summarized as follows.¹⁰

(i) The PWY procedure: Under (12) and the rate condition (10) the ADF detector provides consistent estimates $(\hat{r}_{1e}, \hat{r}_{1f}) \xrightarrow{P} (r_{1e}, r_{1f})$ of the origination and termination of the first bubble, but does not detect the second bubble when the duration of the first bubble exceeds that of the second bubble ($\tau_{1f} - \tau_{1e} > \tau_{2f} - \tau_{2e}$). If the duration of the first bubble is shorter than the second bubble $\tau_{1f} - \tau_{1e} \leq \tau_{2f} - \tau_{2e}$, then under the rate condition

$$\frac{1}{cv^{\beta T}} + \frac{cv^{\beta T}}{T^{1-\alpha/2}} \rightarrow 0 \text{ as } T \rightarrow \infty, \quad (20)$$

PWY consistently estimates the first bubble and detects the second bubble but with a delay that misdates the bubble – specifically $(\hat{r}_{2e}, \hat{r}_{2f}) \xrightarrow{P} (r_{2e} + r_{1f} - r_{1e}, r_{2f})$.

(ii) The BSADF procedure: Under (12) and the rate condition (11) the BSADF detector provides consistent estimates $(\hat{r}_{1e}, \hat{r}_{1f}, \hat{r}_{2e}, \hat{r}_{2f}) \xrightarrow{P} (r_{1e}, r_{1f}, r_{2e}, r_{2f})$ of the origination and termination points of the first and second bubbles.

(iii) The sequential PWY procedure: Under (12) and the rate condition (10), sequential application (with re-initialization) of the ADF detector used in PWY provides consistent estimates $(\hat{r}_{1e}, \hat{r}_{1f}, \hat{r}_{2e}, \hat{r}_{2f}) \xrightarrow{P} (r_{1e}, r_{1f}, r_{2e}, r_{2f})$ of the origination and termination points of the first and second bubbles.

When the sample period includes successive bubble episodes the detection strategy of PWY consistently estimates the origination and termination of the first bubble but does not consistently date stamp the second bubble when the first bubble has longer duration. The new BSADF procedure and repeated implementation (with re-initialization) of the PWY strategy both provide consistent estimates of the origination and termination dates of the two bubbles. PSY2 also examine the consistency properties of the date-stamping strategies when the duration of the first bubble is shorter than the second bubble. In this case, the PWY procedure fails

¹⁰As mentioned earlier, the condition $r_0 \leq r_{1e}$ is needed for consistent estimation of r_{1e} .

to fully consistently date-stamp the second bubble whereas the new strategy again succeeds in consistently estimating both the origination and termination dates of the two bubbles.

The reason for detection failures in the original PWY procedure lies in the asymptotic behavior of the recursive estimates of the autoregressive coefficient. Under data generating mechanisms such as (12), a recursive estimate $\hat{\delta}_{0,t}$ of $\delta_T = 1 + \frac{c}{T^\alpha}$ that is based on data up to observation $t \in B_2$ is dominated by data over the earlier domain $N_0 \cup B_1 \cup N_1$ and it turns out that $\hat{\delta}_{0,t} \sim 1 - \frac{c}{T^\alpha} < 1$. It follows that right sided unit root tests generally will not detect explosive behavior with such asymptotic behavior in the coefficient estimate. This difficulty is completely avoided by flexible rolling window methods such as the new BSADF test or by repeated use of the original PWY procedure with re-initialization that eliminates the effects of earlier bubble episodes.¹¹

4 Simulations

Simulations were conducted to assess the performance of the PWY and sequential PWY procedures as well as the CUSUM approach and the new moving window detection procedure developed in the present paper. We look at size, power, and detection capability for multiple bubble episodes. The data generating process for size comparisons is the null hypothesis in (3) with $d = \eta = 1$. Discriminatory power in detecting bubbles is determined for two different generating models – the Evans (1991) collapsing bubble model (see (21) - (24) below) and an extended version of the PWY bubble model (given by (9) and (12)).

4.1 Size Comparisons

We concentrate on the SADF and GSADF tests. Size is calculated based on the asymptotic critical values displayed in Table 1 using a nominal size of 5%. The number of replications is 5,000. From Table 2, it is clear that with $k = 0$ (no additional transient dynamic lags in the system), size performance of both tests is satisfactory. We observe that size distortion in the

¹¹To consistently estimate the second bubble using PSY and sequential PWY detectors, the minimum window size needs to be small enough to distinguish the different episodes. In particular, r_0 should be less than the distance separating the two bubbles, i.e. $r_0 < r_{2e} - r_{1f}$. See PSY for more discussion.

GSADF test is smaller than that of the SADF test. For example, when $T = 200$ and $r_0 = 0.2$, size distortion of the GSADF test is 0.6% whereas that of the SADF test is 1.2%.

In addition, we explore the effect of fixed and variable transient dynamic lag length selection in the tests, using a fixed lag, BIC order selection, and sequential significance testing (e.g. Campbell and Perron, 1991) with maximum lag 12. First, as evident in Table 2, use of significance testing to determine order leads to non negligible size distortion particularly when the maximum lag length (k_{\max}) is large. For instance, when $T = 200, r_0 = 0.2$, and $k_{\max} = 12$, sizes of the SADF and GSADF tests are 0.116 and 0.557 (for a nominal size of 5%), indicating distortion in both tests, particularly GSADF which is vulnerable because of the short sample sizes that arise in the implementation of the flexible window width method. Second, there are downward size distortions for both tests when using a fixed lag order ($k = 3$). Third, BIC provides satisfactory sizes for SADF but less so for GSADF, where size distortion is positive and increases with sample size.

Overall, the magnitude of the size distortion seems smallest when a fixed lag length is used in the recursive tests. The tests are conservative in this case and GSADF is noticeably less affected than SADF. There are advantages to conservative testing because size must go to zero for consistent date stamping of bubbles. So conservative testing helps to reduce the false detection probability. We therefore recommend using fixed lag length methods in the GSADF testing and dating algorithms. This approach is used later in the paper in the empirical application.

4.2 Power Comparisons

4.2.1 Collapsing Bubble Alternatives

We first simulate asset price series based on the Lucas asset pricing model and Evans (1991) collapsing bubble model. The simulated asset prices consist of a market fundamental component P_t^f , which combines a random walk dividend process and equation (1) with $U_t = 0$ and $B_t = 0$

Table 2: Sizes of the SADF and GSADF tests with asymptotic critical values. The data generating process is equation (3) with $d = \eta = 1$. The nominal size is 5%.

	$k = 0$	$k = 3$	BIC	Significance Test
$T = 100$ and $r_0 = 0.4$				
SADF	0.043	0.008	0.040	0.115
GSADF	0.048	0.021	0.064	0.378
$T = 200$ and $r_0 = 0.2$				
SADF	0.038	0.007	0.050	0.116
GSADF	0.044	0.024	0.105	0.557
$T = 400$ and $r_0 = 0.1$				
SADF	0.034	0.007	0.056	0.137
GSADF	0.059	0.037	0.131	0.790

Note: size calculations are based on 5000 replications.

for all t to obtain¹²

$$D_t = \mu + D_{t-1} + \varepsilon_{Dt}, \quad \varepsilon_{Dt} \sim N(0, \sigma_D^2) \quad (21)$$

$$P_t^f = \frac{\mu\rho}{(1-\rho)^2} + \frac{\rho}{1-\rho}D_t, \quad (22)$$

and the Evans bubble component

$$B_{t+1} = \rho^{-1}B_t\varepsilon_{B,t+1}, \quad \text{if } B_t < b \quad (23)$$

$$B_{t+1} = \left[\zeta + (\pi\rho)^{-1}\theta_{t+1}(B_t - \rho\zeta) \right] \varepsilon_{B,t+1}, \quad \text{if } B_t \geq b. \quad (24)$$

This series has the submartingale property $\mathbb{E}_t(B_{t+1}) = (1 + r_f)B_t$. Parameter μ is the drift of the dividend process, σ_D^2 is the variance of the dividend, ρ is a discount factor with $\rho^{-1} = 1 + r_f > 1$ and $\varepsilon_{B,t} = \exp(y_t - \tau^2/2)$ with $y_t \sim N(0, \tau^2)$. The quantity ζ is the re-initializing value after the bubble collapse. The series θ_t follows a Bernoulli process which takes the value 1 with probability π and 0 with probability $1 - \pi$. Equations (23) - (24) state that a bubble

¹²An alternative data generating process, which assumes that the logarithmic dividend is a random walk with drift, is as follows:

$$\ln D_t = \mu + \ln D_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_d^2)$$

$$P_t^f = \frac{\rho \exp(\mu + \frac{1}{2}\sigma_d^2)}{1 - \rho \exp(\mu + \frac{1}{2}\sigma_d^2)} D_t.$$

grows explosively at rate ρ^{-1} when its size is less than b while if the size is greater than b , the bubble grows at a faster rate $(\pi\rho)^{-1}$ but with a $1 - \pi$ probability of collapsing. The asset price is the sum of the market fundamental and the bubble component, namely $P_t = P_t^f + \kappa B_t$, where $\kappa > 0$ controls the relative magnitudes of these two components.

The parameter settings used by Evans (1991) are displayed in the top line of Table 3 and labeled *yearly*. The parameter values for μ and σ_D^2 were originally obtained by West (1988), by matching the sample mean and sample variance of first differenced real S&P 500 stock price index and dividends from 1871 to 1980. The value for the discount factor ρ is equivalent to a 5% yearly interest rate.

Table 3: Parameter settings

	μ	σ_D^2	D_0	ρ	b	B_0	π	ζ	τ	κ
<i>Yearly</i>	0.0373	0.1574	1.3	0.952	1	0.50	0.85	0.50	0.05	20
<i>Monthly</i>	0.0024	0.0010	1.0	0.985	1	0.50	0.85	0.50	0.05	50

In our empirical application of the SADF and GSADF tests to S&P 500 data we use monthly data. Correspondingly in our simulations, the parameters μ and σ_D^2 are set to match the sample mean and sample variance of the first differenced monthly real S&P 500 stock price index and dividend series described in the application section below. The discount value ρ equals 0.985 (we allow ρ to vary from 0.975 to 0.999 in the power comparisons). The new setting is labeled *monthly* in Table 3.

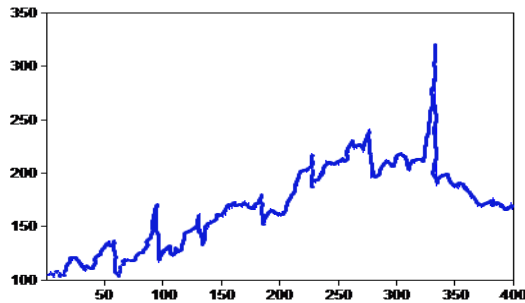


Fig. 5: Simulated time series of $P_t = P_t^f + \kappa B_t$ using the Evans collapsing bubble model (21) - (24) with sample size 400 and monthly parameter settings.

Fig. 5 depicts one realization of the data generating process with the monthly parameter settings. As is apparent in the figure, there are several collapsing episodes of different magnitudes within this particular sample trajectory. Implementation of the SADF and GSADF tests on this particular realization reveals some of the advantages and disadvantages of the two approaches.

We first implement the SADF test on the whole sample range of this trajectory. We then repeat the test on a sub-sample which contains fewer collapsing episodes. The smallest window size considered in the SADF test for the whole sample contains 40 observations (setting $r_0 = 0.1, T = 400$). The SADF statistic for the full trajectory is 0.71,¹³ which is smaller than the 10% finite sample critical value 1.19 (from Table 1). According to this test, therefore, we would conclude that there are no bubbles in the sample.

Next suppose that the SADF test starts from the 201st observation, and the smallest regression window again contains 40 observations (setting $r_0 = 0.2, T = 200$). The SADF statistic obtained from this sample is 1.39¹⁴, which is greater than the 5% finite sample critical value 1.30 (from Table 1). In this case, we reject the null hypothesis of no bubble in the data at the 5% level. These conflicting results point to some instability in the SADF test: evidently the SADF test fails to find bubbles when the full sample is utilized whereas when the sample is truncated to exclude some of the collapse episodes the test succeeds in finding supportive evidence of bubbles.

These two experiments can be viewed as particular (fixed) component runs within the flexible window GSADF test. In the first experiment, the sample starting point of the GSADF test r_1 is set to 0, while in the second experiment the sample starting point r_1 is fixed at 0.5. The conflicting results obtained from these two experiments demonstrate the importance of allowing for variable starting points in the implementation of the test, as is done in the GSADF test. When the GSADF procedure is applied to the data, the test statistic value is 8.59,¹⁵ which substantially exceeds the 1% finite sample critical value 2.71 (from Table 1). Thus, the GSADF

¹³The SADF statistic is obtained from the subsample regression running from the first observation to the peak of the most significant explosive episode within the sample period (i.e. the 333th observation).

¹⁴This value comes from the subsample regression starting with the 201st observation up to the 333th observation.

¹⁵This value is obtained from the subsample regression which covers the most significant expansion period, spanning from the 289th observation to the 333th observation.

test finds strong evidence of bubbles in the simulated data. Compared to the SADF test, the GSADF identifies bubbles without having to arbitrarily re-select sample starting points, giving an obvious improvement that is useful in empirical applications.¹⁶

We now proceed to discuss the results for the full simulation of $R = 5000$ replications. The powers shown in Tables 4 and 5 are calculated using 95% quantiles of the finite sample distributions (Table 1). The smallest window size for both the SADF test and the GSADF test has 40 observations. The data generating process is the periodically collapsing explosive process given in (21) - (24). For comparison with the literature, we first set the parameters in the DGP as in Evans (1991) with sample sizes of 100 and 200. From the left panel of Table 4 (labeled *yearly*), the powers of the GSADF test are respectively 7% and 15.2% higher than those of the SADF test when $T = 100$ and 200.¹⁷

Table 4: Powers of the SADF and GSADF tests. The data generating process is equation (21)-(24).

	<i>Yearly</i>		<i>Monthly</i>	
	SADF	GSADF	SADF	GSADF
$T = 100$ and $r_0 = 0.4$	0.408	0.478	0.509	0.556
$T = 200$ and $r_0 = 0.2$	0.634	0.786	0.699	0.833
$T = 400$ and $r_0 = 0.1$	-	-	0.832	0.977

Note: power calculations are based on 5000 replications.

Table 4 also displays powers of the SADF and GSADF tests under the DGP with monthly parameter settings and sample sizes 100, 200 and 400. From the right panel of the table, when the sample size $T = 400$, the GSADF test raises test power from 83.2% to 97.7%, giving a 14.5% improvement. The power improvement of the GSADF test is 4.7% when $T = 100$ and 13.4% when $T = 200$. For any given bubble collapsing probability π in the Evans model, the sample period is more likely to include multiple collapsing episodes the larger the sample size. Hence,

¹⁶Similar phenomena (not reported in detail here) were observed with an alternative data generating process where the logarithmic dividend is a random walk with drift. Parameters in the alternative data generating process (monthly) were set as follows: $B_0 = 0.5$, $b = 1$, $\pi = 0.85$, $\zeta = 0.5$, $\rho = 0.985$, $\tau = 0.05$, $\mu = 0.001$, $\ln D_0 = 1$, $\sigma_{\ln D}^2 = 0.0001$, and $P_t = P_t^f + 500B_t$.

¹⁷We also considered test results when the lag order is determined by significance testing as in Campbell and Perron (1991) with a maximum lag order of 12. When $T = 200$ and $r_0 = 0.2$, the powers of the SADF test and the GSADF test are 0.565 and 0.661, which are smaller than those in Table 4.

the advantages of the GSADF test are more evident for large T .

In Table 5 we compare powers of the SADF and GSADF tests with the discount factor ρ varying from 0.975 to 0.990, under the DGP with the monthly parameter setting. First, due to the fact that the rate of bubble expansion in this model is inversely related to the discount factor, powers of both the SADF test and GSADF tests are expected to decrease as ρ increases. The power of the SADF (GSADF) test declines from 84.5% to 76.9% (99.3% to 91.0%, respectively) as the discount factor rises from 0.975 to 0.990. Second, as apparent in Table 5, the GSADF test has greater discriminatory power for detecting bubbles than the SADF test. The power improvement is 14.8%, 14.8%, 14.5% and 14.1% for $\rho = \{0.975, 0.980, 0.985, 0.990\}$.

Table 5: Powers of the SADF and GSADF tests. The data generating process is equations (21)-(24) with the *monthly* parameter settings and sample size 400 ($r_0 = 0.1$).

ρ	0.975	0.980	0.985	0.990
SADF	0.845	0.840	0.832	0.769
GSADF	0.993	0.988	0.977	0.910

Note: power calculations are based on 5000 replications.

4.2.2 Mildly Explosive Alternatives

We next consider mildly explosive bubble alternatives of the form generated by (9) and (12). These models allow for both single and double bubble scenarios and enable us to compare the finite sample performance of the PWY strategy, the sequential PWY approach, the new dating method and the CUSUM procedure.¹⁸ The CUSUM detector is denoted by $C_{r_0}^r$ and defined as

$$C_{r_0}^r = \frac{1}{\hat{\sigma}_r} \sum_{j=[Tr_0]+1}^{[Tr]} \Delta y_j \text{ with } \hat{\sigma}_r^2 = ([Tr] - 1)^{-1} \sum_{j=1}^{[Tr]} (\Delta y_j - \hat{\mu}_r)^2,$$

where $[Tr_0]$ is the training sample,¹⁹ $[Tr]$ is the monitoring observation, $\hat{\mu}_r$ is the mean of $\{\Delta y_1, \dots, \Delta y_{[Tr]}\}$, and $r > r_0$. Under the null hypothesis of a pure random walk, the recursive

¹⁸Simulations in Homm and Breitung (2012) show that the PWY strategy has higher power than other procedures in detecting periodically collapsing bubbles of the Evans (1991) type, the closest rival being the CUSUM procedure.

¹⁹It is assumed that there is no structural break in the training sample.

statistic $C_{r_0}^r$ has the following asymptotic property (see Chu, Stinchcombe and White (1996))

$$\lim_{T \rightarrow \infty} P \left\{ C_{r_0}^r > c_r \sqrt{[Tr]} \text{ for some } r \in (r_0, 1] \right\} \leq \frac{1}{2} \exp(-\kappa_\alpha/2),$$

where $c_r = \sqrt{\kappa_\alpha + \log(r/r_0)}$.²⁰ For the sequential PWY method, we use an automated procedure to re-initialize the process following bubble detection. Specifically, if the PWY strategy identifies a collapse in the market at time t (i.e. $ADF_{t-1} > cv_{t-1}^{0.95}$ and $ADF_t < cv_t^{0.95}$)²¹ we re-initialize the test from observation t .

We set the parameters $y_0 = 100$ and $\sigma = 6.79$ so that they match the initial value and the sample standard deviation of the differenced series of the normalized S&P 500 price-dividend ratio described later in our empirical application. The remaining parameters are set to $c = 1$, $\alpha = 0.6$ and $T = 100$. For the one bubble experiment, we set the duration of the bubble to be 15% of the total sample and let the bubble originate 40% into the sample (i.e. $\tau_f - \tau_e = [0.15T]$ and $\tau_e = [0.4T]$). For the two-bubble experiment, the bubbles originate 20% and 60% into the sample, and the durations are $[0.20T]$ and $[0.10T]$, respectively. Fig. 6 displays typical realizations of these two data generating processes.

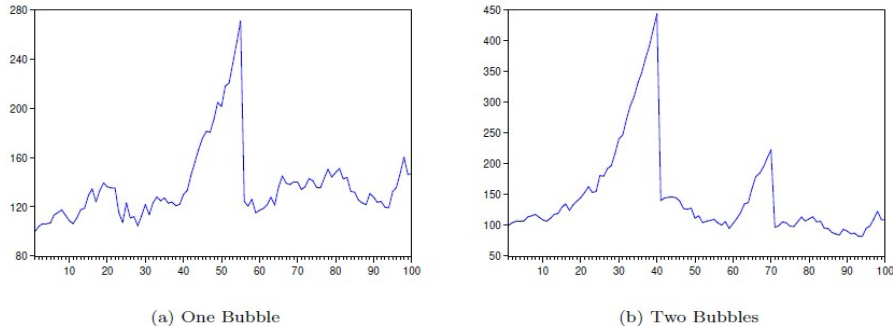


Fig. 6: Typical sample paths generated according to (9) for panel (a) and (12) for panel (b).

We report here a summary of the simulation findings for the main experimental designs. In simulations, we allow τ_e , the bubble location parameter in the single bubble process (13), to be $[0.2T]$, $[0.4T]$ and $[0.6T]$ and the duration of bubbles to vary from $[0.10T]$ to $[0.20T]$. Other

²⁰When the significance level $\alpha = 0.05$, for instance, $\kappa_{0.05}$ equals 4.6.

²¹We impose the additional restriction of successive realizations $ADF_{t+1} < cv_{t+1}^{0.95}$ and $ADF_{t+2} < cv_{t+2}^{0.95}$ to confirm a bubble collapse.

parameter configurations were considered and led to broadly similar findings, so are not reported here. For each parameter constellation, 5,000 replications were used. Bubbles were identified using respective finite sample 95% quantiles, obtained from Monte Carlo simulations with 5,000 replications. The minimum window size has 12 observations. We report the empirical mean and standard deviation (in parentheses) of the number of bubbles identified within the sample range, along with the proportions of sample paths identified with bubbles (in squared brackets). Sample paths with no evidence of bubbles found are omitted from the calculation of the mean and the standard deviation. Tables 6 - 7 provide a selection of the results. The main findings are as follows.

Table 6: Number of bubbles identified for the single bubble DGP with different bubble durations and locations. Parameters are set as: $y_0 = 100, c = 1, \sigma = 6.79, \alpha = 0.6, T = 100$. Figures in parentheses and square brackets are standard deviations and powers of the tests, respectively.

	PWY	PSY	Seq	CUSUM
$\tau_e = [0.2T]$				
$\tau_f - \tau_e = [0.10T]$	1.05 (0.27) [0.81]	1.25 (0.51) [0.84]	1.36 (0.62) [0.80]	1.09 (0.34) [0.71]
$\tau_f - \tau_e = [0.15T]$	1.04 (0.20) [0.92]	1.23 (0.50) [0.95]	1.36 (0.61) [0.91]	1.04 (0.24) [0.90]
$\tau_f - \tau_e = [0.20T]$	1.04 (0.20) [0.96]	1.22 (0.48) [0.98]	1.35 (0.61) [0.95]	1.03 (0.20) [0.96]
$\tau_e = [0.4T]$				
$\tau_f - \tau_e = [0.10T]$	1.18 (0.45) [0.73]	1.24 (0.48) [0.81]	1.35 (0.60) [0.73]	1.17 (0.45) [0.76]
$\tau_f - \tau_e = [0.15T]$	1.17 (0.43) [0.88]	1.22 (0.47) [0.91]	1.33 (0.58) [0.88]	1.13 (0.40) [0.91]
$\tau_f - \tau_e = [0.20T]$	1.17 (0.44) [0.94]	1.22 (0.47) [0.96]	1.31 (0.56) [0.93]	1.13 (0.38) [0.96]
$\tau_e = [0.6T]$				
$\tau_f - \tau_e = [0.10T]$	1.32 (0.64) [0.72]	1.26 (0.51) [0.79]	1.37 (0.62) [0.73]	1.26 (0.58) [0.78]
$\tau_f - \tau_e = [0.15T]$	1.30 (0.61) [0.86]	1.25 (0.50) [0.90]	1.32 (0.58) [0.86]	1.23 (0.55) [0.90]
$\tau_f - \tau_e = [0.20T]$	1.30 (0.61) [0.92]	1.26 (0.50) [0.95]	1.28 (0.55) [0.92]	1.23 (0.55) [0.94]

Note: Calculations are based on 5,000 replications. The minimum window has 12 observations.

1. For the single bubble case, powers of the PWY strategy, the sequential PWY approach and the CUSUM procedure are similar and are slightly lower than that of the new strategy (Table 6).
2. The power of the tests increase with the duration of bubble expansion, shown in square

brackets in Table 6, and with the value of the autoregressive coefficient δ_T , although this is not reported here. Hence, bubble detection is more successful when duration of the bubble is longer (and its expansion rate faster).

3. Bubble location has little impact on the accuracy of the PSY and sequential PWY estimators. But the PWY and CUSUM estimators both become less accurate if the bubble originates at a later stage of the sample period, as shown in parentheses in Table 6. Overall, in the one bubble scenario, the sequential PWY procedure tends to over-estimate the bubble number, the PSY estimator to slightly overestimate bubble number, and the PWY and CUSUM estimators to be more accurate.
4. In the two-bubble scenario, bubble duration can have an especially large impact on the PWY strategy, as is clear in Table 7. When the duration of the first bubble is longer than the second bubble, the mean values of the PWY bubble number estimates are far from the true value and close to one, indicating substantial underestimation. This is consistent with the asymptotic theory which shows that when the duration of the first bubble is longer than the second bubble, the PWY strategy consistently identifies the first bubble but not the second bubble. When the duration of the second bubble is longer than the first bubble, the bias of the estimate is much smaller. For instance, when $\tau_{1f} - \tau_{1e} = [0.1T]$ and $\tau_{2f} - \tau_{2e} = [0.2T]$, the mean value of the PWY estimates is 1.71, which is much closer to the true value. This simulation finding corroborates the asymptotic theory, which shows that the PWY strategy can detect both bubbles under these conditions.
5. Similar to the weakness of the PWY strategy, when the duration of first bubble is longer than that of the second bubble, the performance of the CUSUM procedure is also biased downwards to selecting a single bubble. Also, like the PWY procedure, there is obvious improvement in the performance of the CUSUM procedure when the second bubble lasts longer (Table 7 final column).
6. As expected, the sequential PWY procedure performs nearly as well as the PSY strategy in the two bubble case but tends to have higher variation and less power than PSY.

Estimation accuracy of both estimators improves with the durations of bubbles. (Table 7).

- Overall, substantially better performance in the two bubble case is delivered by the PSY and sequential PWY estimators, with higher power and much greater accuracy in determining the presence of more than one bubble (Table 7 column 2 and 3).

Table 7: Number of bubbles identified for the two-bubble DGP with varying bubble duration. Parameters are set as: $y_0 = 100, c = 1, \sigma = 6.79, \alpha = 0.6, \tau_{1e} = \lfloor 0.20T \rfloor, \tau_{2e} = \lfloor 0.60T \rfloor, T = 100$. Figures in parentheses and square brackets are standard deviations and powers of the tests, respectively.

	PWY	PSY	Seq	CUSUM
$\tau_{1f} - \tau_{1e} = \lfloor \mathbf{0.10T} \rfloor$				
$\tau_{2f} - \tau_{2e} = \lfloor 0.10T \rfloor$	1.25 (0.46) [0.87]	1.77 (0.57) [0.92]	1.85 (0.63) [0.86]	1.41 (0.53) [0.84]
$\tau_{2f} - \tau_{2e} = \lfloor 0.15T \rfloor$	1.55 (0.53) [0.91]	1.84 (0.54) [0.96]	1.89 (0.57) [0.91]	1.62 (0.52) [0.93]
$\tau_{2f} - \tau_{2e} = \lfloor 0.20T \rfloor$	1.71 (0.49) [0.95]	1.86 (0.52) [0.97]	1.85 (0.53) [0.95]	1.68 (0.50) [0.97]
$\tau_{1f} - \tau_{1e} = \lfloor \mathbf{0.15T} \rfloor$				
$\tau_{2f} - \tau_{2e} = \lfloor 0.10T \rfloor$	1.08 (0.30) [0.94]	1.83 (0.54) [0.97]	1.87 (0.61) [0.94]	1.15 (0.39) [0.93]
$\tau_{2f} - \tau_{2e} = \lfloor 0.15T \rfloor$	1.28 (0.47) [0.95]	1.94 (0.47) [0.98]	1.95 (0.53) [0.96]	1.48 (0.53) [0.96]
$\tau_{2f} - \tau_{2e} = \lfloor 0.20T \rfloor$	1.60 (0.53) [0.97]	1.97 (0.43) [0.99]	1.93 (0.47) [0.97]	1.75 (0.48) [0.98]
$\tau_{1f} - \tau_{1e} = \lfloor \mathbf{0.20T} \rfloor$				
$\tau_{2f} - \tau_{2e} = \lfloor 0.10T \rfloor$	1.05 (0.23) [0.97]	1.83 (0.53) [0.99]	1.89 (0.60) [0.97]	1.06 (0.26) [0.97]
$\tau_{2f} - \tau_{2e} = \lfloor 0.15T \rfloor$	1.10 (0.31) [0.97]	1.97 (0.43) [0.99]	1.96 (0.52) [0.98]	1.16 (0.39) [0.97]
$\tau_{2f} - \tau_{2e} = \lfloor 0.20T \rfloor$	1.29 (0.48) [0.98]	2.01 (0.38) [0.99]	1.95 (0.45) [0.98]	1.46 (0.54) [0.98]

Note: Calculations are based on 5,000 replications. The minimum window has 12 observations.

5 Empirical Application

We consider a long historical time series in which many crisis events are known to have occurred. The data comprise the real S&P 500 stock price index and the real S&P 500 stock price index dividend, both obtained from Robert Shiller’s website. The data are sampled monthly over the period from January 1871 to December 2010, constituting 1,680 observations and are plotted in Fig. 6 by the solid (blue) line, which shows the price-dividend ratio over this period to reflect asset prices in relation to fundamentals, according to the pricing equation (1). One might allow

also for a time-varying discount factor in that equation. If there were no unobservable component in fundamentals, it follows from the pricing equation that in the absence of bubbles the price-dividend ratio is a function of the discount factor and the dividend growth rate (e.g., Cochrane, 1992; Ang and Bekaert, 2006). In such cases, tests for a unit root in the price-dividend ratio do not preclude the presence of a (stationary or nonstationary) time-varying discount factor influencing the ratio.

Table 8: The SADF test and the GSADF test of the S&P500 price-dividend ratio

	Test Stat.	Finite Sample Critical Values		
		90%	95%	99%
SADF	3.30	1.45	1.70	2.17
GSADF	4.21	2.55	2.80	3.31

Note: Critical values of both tests are obtained from Monte Carlo simulation with 2,000 replications (sample size 1,680). The smallest window has 36 observations.

We first apply the summary SADF and GSADF tests to the price-dividend ratio. Table 8 presents critical values for these two tests obtained by simulation with 2,000 replications (sample size 1,680). In performing the ADF regressions and calculating the critical values, the smallest window comprised 36 observations. From Table 8, the SADF and GSADF statistics for the full data series are 3.30 and 4.21, obtained from subsamples 1987M01-2000M07 and 1976M04-1999M06, respectively. Both exceed their respective 1% right-tail critical values (i.e. $3.30 > 2.17$ and $4.21 > 3.31$), giving strong evidence that the S&P 500 price-dividend ratio had explosive subperiods. We conclude from both summary tests that there is evidence of bubbles in the S&P 500 stock market data. These calculations used a transient dynamic lag order $k = 0$. The findings are robust to other choices. For example, the same conclusion applies when $k = 3$, where the SADF and GSADF tests for the full data series are 2.16 and 3.88 with corresponding 5% critical values of 1.70 and 3.40.

To locate specific bubble periods, we compare the backward SADF statistic sequence with the 95% SADF critical value sequence, which were obtained from Monte Carlo simulations with 2,000 replications. The top panel of Fig. 7 displays results for the date-stamping strategy over the period from January 1871 to December 1949 and the bottom panel displays results over the

rest of the sample period.

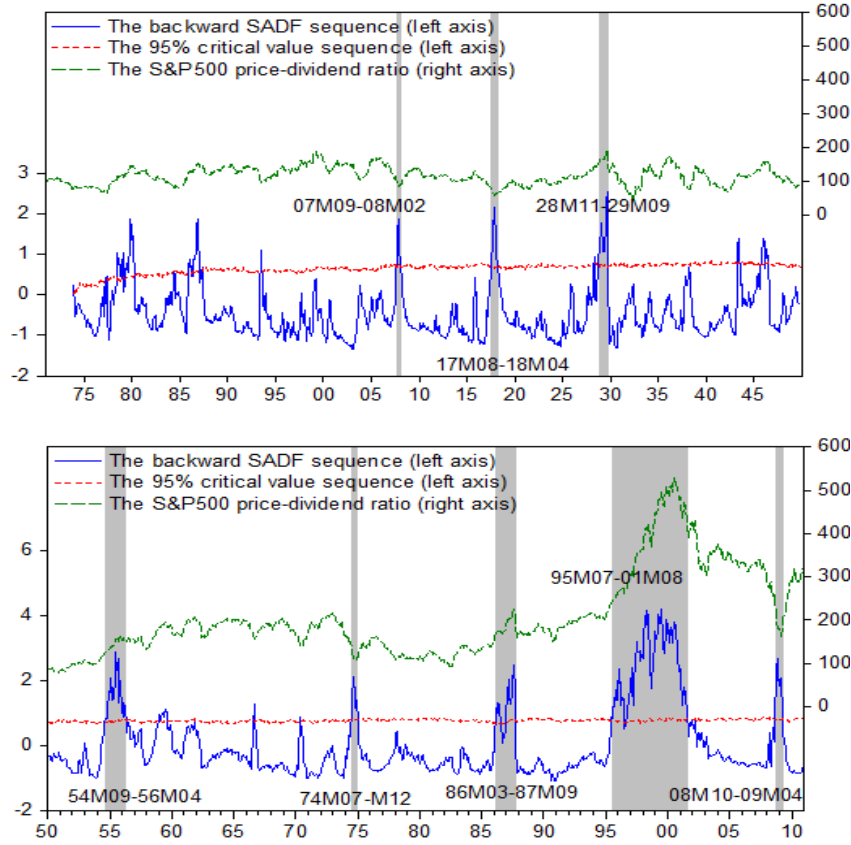


Fig. 7: Date-stamping bubble periods in the S&P 500 price-dividend ratio: the GSADF test.

The identified exuberance and collapse periods after 1900 include *the banking panic of 1907* (1907M09-1908M02), *the 1917 stock market crash* (1917M08-1918M04), *the great crash episode* (1928M11-1929M09), *the postwar boom in 1954* (1954M09-1956M04), *black Monday in October 1987* (1986M03-1987M09), *the dot-com bubble* (1995M07-2001M08) and *the subprime mortgage crisis* (2008M10-2009M04). The durations of those episodes are greater than or equal to half a year. This strategy also identifies several episodes of explosiveness and collapse whose durations are shorter than six months – including *the 1974 stock market crash* (1974M07-M12). Importantly, the new date-stamping strategy not only locates explosive expansion periods but also identifies collapse episodes. Market collapses have occurred in the past when bubbles in

other markets crashed and contagion spread to the S&P 500 as occurred, for instance, during the *dot-com bubble collapse* and *the subprime mortgage crisis*.

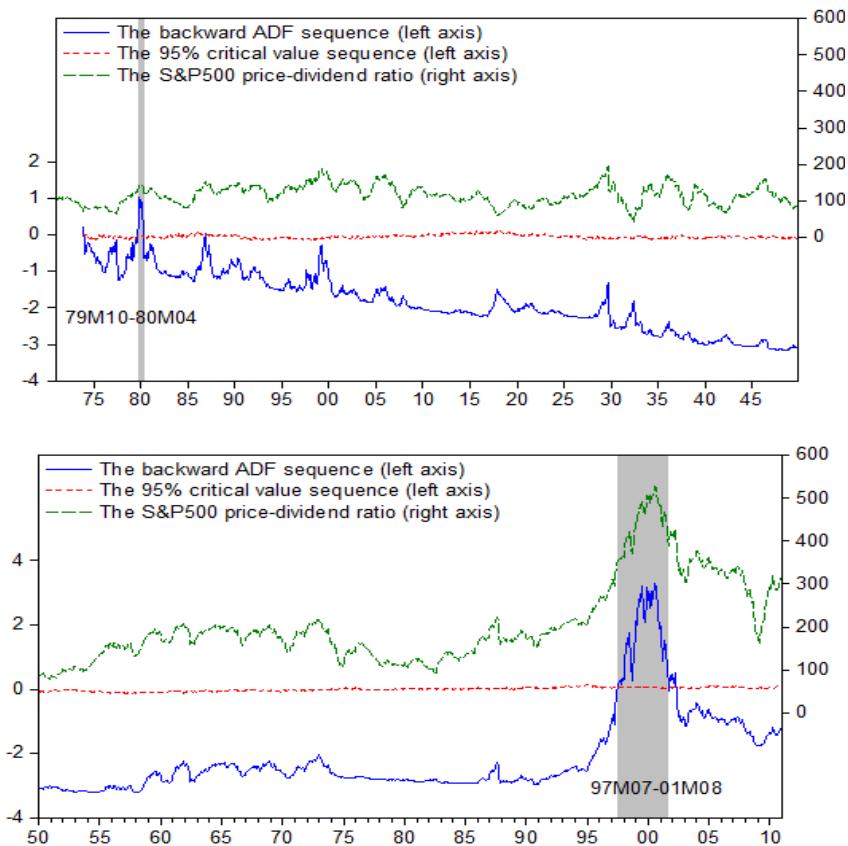


Fig. 8: Date-stamping bubble periods in the S&P 500 price-dividend ratio: the SADF test.

Fig. 8 plots the ADF statistic sequence against the 95% ADF critical value sequence. We can see that the strategy of PWY (based on the SADF test) identifies only two explosive periods – the recovery phase of *the panic of 1873* (1879M10-1880M04) and *the dot-com bubble* (1997M07-2001M08). If we restrict the duration of bubbles to be longer than twelve months. The new dating strategy identifies three bubble episodes: *the postwar boom in 1954*, *black Monday in October 1987* and *the dot-com bubble* whereas the strategy of PWY identifies only *the dot-com bubble* in that case.

Empirical results from the sequential PWY procedure are shown in Fig. 9 which plots the ADF statistic sequence against the 95% ADF critical value sequence (as for the PWY dating

strategy). As in the simulation exercise (see Section 4.2.2) we use automated re-initialization in the implementation of sequential PWY. A minimum window size $\lfloor r_0 T \rfloor$ is needed to initiate the recursive regression test, so the sequential PWY procedure is unable to perform detection (and hence will fail to identify any bubbles that may occur) over the intervening period $(t, t + \lfloor r_0 T \rfloor)$ following a re-initialization at time t . Furthermore, if the PWY strategy fails to detect a bubble, no re-initialization occurs and the recursive test continues through the sample until a bubble is detected and a subsequent re-initialization is triggered. Hence, the sequential PWY strategy, just like PWY, has some inherent disadvantages in detecting multiple bubbles. In practice, one could potentially pre-divide the sample period into sub-samples for testing but, as shown in the example of Section 4.2.1, the subsample approach may well be sensitive to the pre-selection of the sample periods.

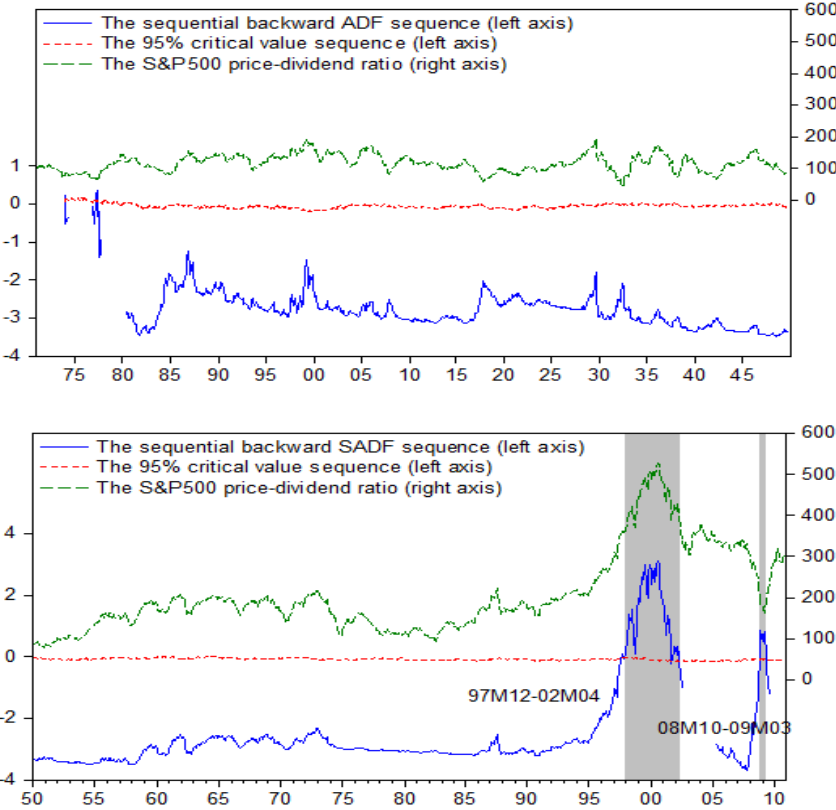


Fig. 9: Date-stamping bubble periods in the S&P 500 price-dividend ratio: the sequential PWY strategy.

The sequential ADF plot has several breaks in the Figure, each corresponding to the re-initialization of the test procedure following a collapse. The findings from the sequential PWY test indicate two bubbles after 1900 – the *dot-com bubble* (1997M12 - 2002M04) and the *subprime mortgage crisis* (2008M10 -2009M03). Interestingly, after excluding data from the dot-com bubble collapse and earlier data, the sequential PWY strategy successfully identifies an additional episode – the *subprime mortgage crisis* – which the PWY strategy fails to catch (Fig. 8).²²

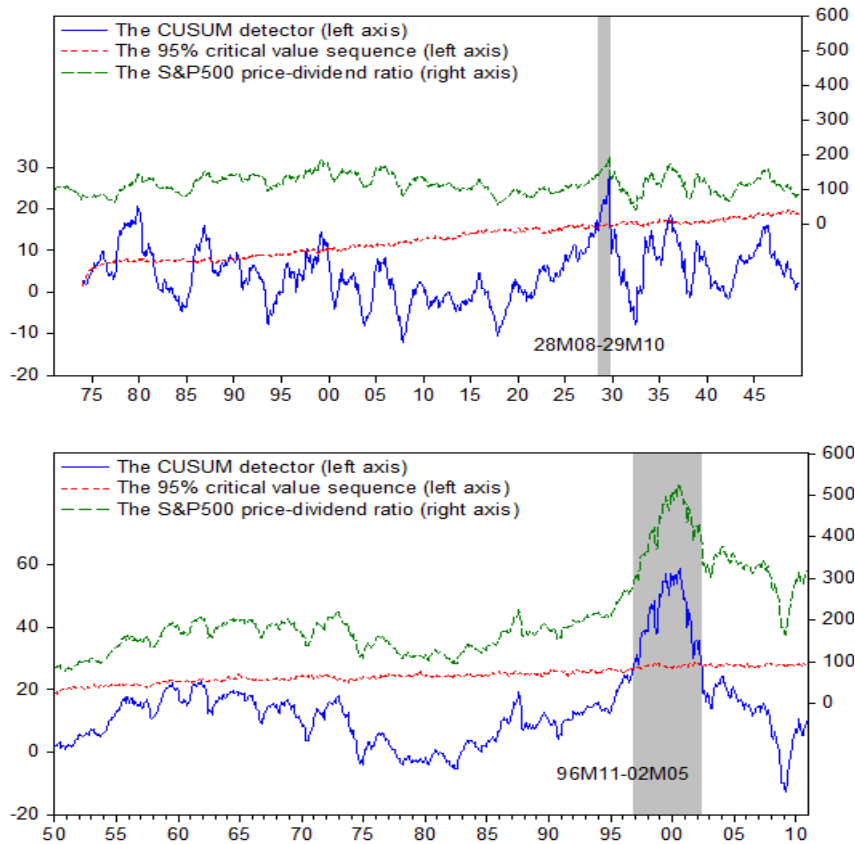


Fig. 10. Date-stamping bubble periods in the S&P 500 price-dividend ratio: the CUSUM monitoring procedure.

For comparison, we applied the CUSUM monitoring procedure to the detrended S&P 500 price-dividend ratio (i.e. to the residuals from the regression of y_t on a constant and a linear time

²²If the transient dynamic lag order is $k = 3$, the backward SADF strategy identifies two additional bubble episodes (namely, 1945M12-1946M07 and 1969M11-1970M12). The PWY and sequential PWY strategies identify the same bubble episodes with slight changes in dates.

trend). To be consistent with the SADF and GSADF dating strategies, we choose a training sample of 36 months. Fig. 8 plots the CUSUM detector sequence against the 95% critical value sequence. The critical value sequence is obtained from Monte Carlo simulation (through application of the CUSUM detector to data simulated from a pure random walk) with 2,000 replications.

As is evident in Fig. 10, the CUSUM test identifies four bubble episodes for periods before 1900. For the post-1900 sample, the procedure detects only the great crash and the dot-com bubble episodes. It does not provide any warning for or acknowledgment of black Monday in October 1987 and the subprime mortgage crisis in 2008, among other episodes identified by the GSADF dating strategy. So CUSUM monitoring may be regarded as a relatively conservative surveillance device.²³

6 Conclusion

This paper introduces a new recursive testing procedure and dating algorithm that is useful in detecting multiple bubble events. The GSADF test is a rolling window right-sided ADF unit root test with a double-sup window selection criteria. The reason for the double sup is that the ADF statistic is computed over feasible ranges of the window start points and over a feasible range of window sizes. As distinct from the SADF test of PWY, the window size is selected using the double-sup criteria and the ADF test is implemented repeatedly on a sequence of samples, which moves the window frame gradually toward the end of the sample. Experimenting on simulated asset prices reveals one of the shortcomings of the SADF test - its limited ability to find and locate bubbles when there are multiple collapsing episodes within the sample range. The GSADF test surmounts this limitation and our simulation findings demonstrate that the GSADF test significantly improves discriminatory power in detecting multiple bubbles. This advantage is particularly important in the empirical study of long historical data series.

The date-stamping strategy of PWY and the new date-stamping strategy are shown to have

²³The conservative nature of the test arises from the fact the residual variance estimate $\hat{\sigma}_r$ (based on the data $\{y_1, \dots, y_{[Tr]}\}$) can be quite large when the sample includes periodically collapsing bubble episodes, which may have less impact on the numerator due to collapses, thereby reducing the size of the CUSUM detector.

quite different behavior under the alternative of multiple bubbles. In particular, when the sample period includes two bubbles the strategy of PWY often fails to identify or consistently date stamp the second bubble, whereas the new strategy consistently estimates and dates both bubbles. The PWY dating algorithm may be applied sequentially by re-initializing the detection process after a bubble is found. This sequential application of the PWY dating algorithm has improved asymptotic properties over PWY in the detection of multiple bubbles but both simulations and empirical applications show its performance to be more limited in this capacity.

We apply both SADF and GSADF tests, the sequential PWY dating algorithm, and the CUSUM monitoring procedure, along with their date-stamping algorithms, to the S&P 500 price-dividend ratio from January 1871 to December 2010. All four tests find confirmatory evidence of multiple bubble existence. The price-dividend ratio over this historical period contains many individual peaks and troughs, a trajectory that is similar to the multiple bubble scenario for which the PWY date-stamping strategy turns out to be inconsistent. The empirical test results confirm the greater discriminatory power of the GSADF strategy found in the simulations and evidenced in the asymptotic theory. The new date-stamping strategy identifies all the well known historical episodes of banking crises and financial bubbles over this long period, whereas all other procedures seem more conservative and locate fewer episodes of exuberance and collapse.

7 References

- Ahamed, L., 2009, *Lords of Finance: The Bankers Who Broke the World*, Penguin Press, New York.
- Bhargava, A. 1986, On the theory of testing for unit roots in observed time series, *Review of Economic Studies*, 53, 369–384.
- Busetti, F., and A. M. R. Taylor, 2004, Tests of stationarity against a change in persistence, *Journal of Econometrics*, 123, 33–66.
- Campbell, J.Y., and Perron, P., 1991, Pitfalls and opportunities: what macroeconomists should know about unit roots. *NBER Macroeconomics Annual*, 6, 141–201.

- Campbell, J.Y., and Shiller R.J., 1989, The dividend-price ratio and expectations of future dividends and discount factors. *The Review of Financial Studies*, 1, 195–228.
- Charemza, W.W., and Deadman, D.F., 1995, Speculative bubbles with stochastic explosive roots: the failure of unit root testing. *Journal of Empirical Finance*, 2, 153–163.
- Chu, C.J., Stinchcombe, M., and White, H., 1996, Monitoring structural change. *Econometrica*, 64:1045-1065.
- Cochrane, J. H., 1992. Explaining the Variance of Price-Dividend Ratios. *The Review of Financial Studies*, 5(2), 243-280.
- Cochrane, J. H., 2005. *Asset Pricing*, Princeton: Princeton University Press.
- Cooper, G., 2008, *The Origin of Financial Crises: Central Banks, Credit Bubbles and the Efficient Market Fallacy*, Vintage Books, New York.
- Diba, B.T., and Grossman, H.I., 1988, Explosive rational bubbles in stock prices? *The American Economic Review*, 78, 520–530.
- Evans, G.W., 1991, Pitfalls in testing for explosive bubbles in asset prices. *The American Economic Review*, 81, 922–930.
- Ferguson, N., 2008, *The Ascent of Money*, Penguin Press, New York.
- Funke, M., Hall, S., and Sola, M., 1994, Rational bubbles during Poland’s hyperinflation: implications and empirical evidence. *European Economic Review*, 38, 1257–1276.
- Gurkaynak, R. S., 2008, Econometric tests of asset price bubbles: taking stock. *Journal of Economic Surveys*, 22, 166–186.
- Hall, S.G., Psaradakis, Z., and Sola M., 1999, Detecting periodically collapsing bubbles: A markov-switching unit root test. *Journal of Applied Econometrics*, 14, 143–154.
- Homm, U., and Breitung, J., 2012. Testing for speculative bubbles in stock markets: a comparison of alternative methods. *Journal of Financial Econometrics*, 10(1), 198-231.

- Kim, J. Y. 2000, Detection of change in persistence of a linear time series, *Journal of Econometrics*, 95, 97–116.
- Kindleberger, C. P., and Aliber, R. Z., 2005, *Manias, Panics and Crashes; A History of Financial Crises*, Hoboken, New Jersey: John Wiley and Sons, Inc.
- Lee, J. and P. C. B. Phillips, 2011, Asset Pricing with Financial Bubble Risk. *Yale University, Working Paper*.
- Pástor, Luboš, and Pietro Veronesi, 2006, Was there a Nasdaq bubble in the late 1990s? *Journal of Financial Economics* 81, 61–100.
- Phillips, P.C.B., 1987, Time series regression with a unit root. *Econometrica* 55, 277-301.
- Phillips, P.C.B., and Perron, P., 1988, Testing for a unit root in time series regression. *Biometrika*, 75, 335–346.
- Phillips, P.C.B., and Magdalinos, T., 2007, Limit theory for moderate deviations from a unit root. *Journal of Econometrics*, 136, 115–130.
- Phillips, P. C. B. and V. Solo, 1992, Asymptotics for Linear Processes. *Annals of Statistics* 20, 971–1001.
- Phillips, P.C.B., Shi, S., and Yu, J., 2013a, Specification Sensitivity in Right-Tailed Unit Root Testing for Explosive Behavior. *Oxford Bulletin of Economics and Statistics*, forthcoming.
- Phillips, P.C.B., Shi, S., and Yu, J., 2013b. Testing for Multiple Bubbles: Limit Theory of Dating Algorithms. Working Paper.
- Phillips, P.C.B., Wu, Y., and Yu, J., 2011, Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review*, 52, 201-226.
- Phillips, P.C.B., and Yu, J., 2009, Limit theory for dating the origination and collapse of mildly explosive periods in time series data. *Singapore Management University, Unpublished Manuscript*.

Phillips, P.C.B., and Yu, J., 2011, Dating the Timeline of Financial Bubbles During the Sub-prime Crisis, *Quantitative Economics*, 2, 455-491.

Psaradakis, Z., Sola, M., and Spagnolo, F., 2001, A simple procedure for detecting periodically collapsing rational bubbles. *Economics Letters*, 24, 317–323.

Shi, S., 2012. Specification sensitivities in the Markov-switching unit root test for bubbles. *Empirical Economics*, forthcoming.

van Norden, S. and Vigfusson, R., 1998, Avoiding the pitfalls: can regime-switching tests reliably detect bubbles? *Studies in Nonlinear Dynamics & Econometrics*, 3,1–22.

West, K.D., 1988, Dividend Innovations and Stock Price Volatility. *Econometrica*, 56:37–61.

APPENDIX. Asymptotic Distribution of the GSADF test

Before proving Theorem 1 we give conditions on the innovations and state two preliminary lemmas whose proofs follow directly by standard methods (Phillips, 1987; Phillips and Perron, 1988; Phillips and Solo, 1992).

Assumption (EC) Let $u_t = \psi(L)\varepsilon_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$, where $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and $\{\varepsilon_t\}$ is an i.i.d sequence with mean zero, variance σ^2 and finite fourth moment.

Lemma 7.1 Suppose u_t satisfies error condition EC. Define $M_T(r) = 1/T \sum_{s=1}^{\lfloor Tr \rfloor} u_s$ with $r \in [r_0, 1]$ and $\xi_t = \sum_{s=1}^t u_s$. Let $r_2, r_w \in [r_0, 1]$ and $r_1 = r_2 - r_w$. The following hold:

(1) $\sum_{s=1}^t u_s = \psi(1) \sum_{s=1}^t \varepsilon_s + \eta_t - \eta_0$, where $\eta_t = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{t-j}$, $\eta_0 = \sum_{j=0}^{\infty} \alpha_j \varepsilon_{-j}$ and $\alpha_j = -\sum_{i=1}^{\infty} \psi_{j+i}$, which is absolutely summable.

$$(2) \frac{1}{T} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} \varepsilon_t^2 \xrightarrow{p} \sigma^2 r_w.$$

$$(3) T^{-1/2} \sum_{t=1}^{\lfloor Tr \rfloor} \varepsilon_t \xrightarrow{L} \sigma W(r).$$

$$(4) T^{-1} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} \sum_{s=1}^{t-1} \varepsilon_s \varepsilon_t \xrightarrow{L} \frac{1}{2} \sigma^2 \left[W(r_2)^2 - W(r_1)^2 - r_w \right].$$

$$(5) T^{-3/2} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} \varepsilon_t t \xrightarrow{L} \sigma \left[r_2 W(r_2) - r_1 W(r_1) - \int_{r_1}^{r_2} W(s) ds \right].$$

$$(6) T^{-1} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} (\eta_{t-1} - \eta_0) \varepsilon_t \xrightarrow{p} 0.$$

$$(7) T^{-1/2} \left(\eta_{\lfloor Tr \rfloor} - \eta_0 \right) \xrightarrow{p} 0.$$

$$(8) \sqrt{T} M_T(r) \xrightarrow{L} \psi(1) \sigma W(r).$$

- (9) $T^{-3/2} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} \xi_{t-1} \xrightarrow{L} \psi(1) \sigma \int_{r_1}^{r_2} W(s) ds.$
(10) $T^{-5/2} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} \xi_{t-1} t \xrightarrow{L} \psi(1) \sigma \int_{r_1}^{r_2} W(s) s ds.$
(11) $T^{-2} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} \xi_{t-1}^2 \xrightarrow{L} \sigma^2 \psi(1)^2 \int_{r_1}^{r_2} W(s)^2 ds.$
(12) $T^{-3/2} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} \xi_t u_{t-j} \xrightarrow{p} 0, \forall j \geq 0.$

Lemma 7.2 Define $y_t = \tilde{\alpha}_T t + \sum_{s=1}^t u_s$, $\tilde{\alpha}_T = \tilde{\alpha} \psi(1) T^{-\eta}$ with $\eta > 1/2$ and let u_t satisfy condition EC. Then

- (a) $T^{-1} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} y_{t-1} \varepsilon_t \xrightarrow{L} \frac{1}{2} \sigma^2 \psi(1) \left[W(r_2)^2 - W(r_1)^2 - r_w \right].$
(b) $T^{-3/2} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} y_{t-1} \xrightarrow{L} \psi(1) \sigma \int_{r_1}^{r_2} W(s) ds.$
(c) $T^{-2} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} y_{t-1}^2 \xrightarrow{L} \sigma^2 \psi(1)^2 \int_{r_1}^{r_2} W(s)^2 ds.$
(d) $T^{-3/2} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} y_{t-1} u_{t-j} \xrightarrow{p} 0, j = 0, 1, \dots.$

Proof of Theorem 1. The regression model is

$$\Delta y_t = \alpha_{r_1, r_2} + \beta_{r_1, r_2} y_{t-1} + \sum_{k=1}^{p-1} \phi_{r_1, r_2}^k \Delta y_{t-k} + \varepsilon_t.$$

Under the null hypothesis that $\alpha_{r_1, r_2} = \tilde{\alpha} T^{-\eta}$ and $\beta_{r_1, r_2} = 0$, we have $y_t = \tilde{\alpha}_T t + \sum_{s=1}^t u_s$ and $\Delta y_t = \tilde{\alpha}_T + u_t$, where $\tilde{\alpha}_T = \psi_{r_1, r_2}(1) \alpha_{r_1, r_2}$ and $u_t = \psi_{r_1, r_2}(1) \varepsilon_t$ with $\psi_{r_1, r_2}(1) = (1 - \phi_{r_1, r_2}^1 L - \phi_{r_1, r_2}^2 L^2 - \dots - \phi_{r_1, r_2}^{p-1} L^{p-1})^{-1}$.

The deviation of the OLS estimate $\hat{\theta}_{r_1, r_2}$ from the true value θ_{r_1, r_2} is given by

$$\hat{\theta}_{r_1, r_2} - \theta_{r_1, r_2} = \left[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t' \right]^{-1} \left[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t \varepsilon_t \right], \quad (25)$$

where $X_t = [\tilde{\alpha}_T + u_{t-1} \ \tilde{\alpha}_T + u_{t-2} \ \dots \ \tilde{\alpha}_T + u_{t-p+1} \ 1 \ y_{t-1}]'$, $\theta = [\phi_{r_1, r_2}^1 \ \phi_{r_1, r_2}^2 \ \dots \ \phi_{r_1, r_2}^{p-1} \ \alpha_{r_1, r_2} \ \beta_{r_1, r_2}]'$. The probability limit of $\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t'$ is block diagonal from (d) of Lemma 7.2. Therefore, we

only need to obtain the last 2×2 components of $\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t'$ and the last 2×1 component of $\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t \varepsilon_t$ to calculate the ADF statistics, which are

$$\begin{bmatrix} \Sigma' 1 & \Sigma y_{t-1} \\ \Sigma' y_{t-1} & \Sigma y_{t-1}^2 \end{bmatrix} \text{ and } \begin{bmatrix} \Sigma' \varepsilon_t \\ \Sigma' y_{t-1} \varepsilon_t \end{bmatrix},$$

respectively, where Σ' denotes summation over $t = \lfloor Tr_1 \rfloor, \lfloor Tr_1 \rfloor + 1, \dots, \lfloor Tr_2 \rfloor$. Based on (3) in Lemma 7.1 and (a) in Lemma 7.2, the scaling matrix should be $\Upsilon_T = \text{diag}(\sqrt{T}, T)$. Premultiplying equation (25) by Υ_T , results in

$$\Upsilon_T \begin{bmatrix} \hat{\alpha}_{r_1, r_2} - \alpha_{r_1, r_2} \\ \hat{\beta}_{r_1, r_2} - \beta_{r_1, r_2} \end{bmatrix} = \left\{ \Upsilon_T^{-1} \begin{bmatrix} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t' \end{bmatrix}_{(-2) \times (-2)} \Upsilon_T^{-1} \right\}^{-1} \left\{ \Upsilon_T^{-1} \begin{bmatrix} \sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t \varepsilon_t \end{bmatrix}_{(-2) \times 1} \right\}.$$

Consider the matrix $\Upsilon_T^{-1} \left[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t' \right]_{(-2) \times (-2)} \Upsilon_T^{-1}$, whose partitioned form is

$$\begin{bmatrix} \sqrt{T} & 0 \\ 0 & T \end{bmatrix}^{-1} \begin{bmatrix} \Sigma' 1 & \Sigma' y_{t-1} \\ \Sigma' y_{t-1} & \Sigma' y_{t-1}^2 \end{bmatrix} \begin{bmatrix} \sqrt{T} & 0 \\ 0 & T \end{bmatrix}^{-1} = \begin{bmatrix} T^{-1} \Sigma' 1 & T^{-3/2} \Sigma' y_{t-1} \\ T^{-3/2} \Sigma' y_{t-1} & T^{-2} \Sigma' y_{t-1}^2 \end{bmatrix} \\ \xrightarrow{L} \begin{bmatrix} r_w & \psi_{r_1, r_2}(1) \sigma_{r_1, r_2} \int_{r_1}^{r_2} W(s) ds \\ \psi_{r_1, r_2}(1) \sigma_{r_1, r_2} \int_{r_1}^{r_2} W(s) ds & \sigma_{r_1, r_2}^2 \psi_{r_1, r_2}(1)^2 \int_{r_1}^{r_2} W(s)^2 ds \end{bmatrix}$$

and the matrix $\Upsilon_T^{-1} \left[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t \varepsilon_t \right]_{(-2) \times 1}$, for which

$$\begin{bmatrix} T^{-1/2} \Sigma \varepsilon_t \\ T^{-1} \Sigma y_{t-1} \varepsilon_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma_{r_1, r_2} [W(r_2) - W(r_1)] \\ \frac{1}{2} \sigma_{r_1, r_2}^2 \psi_{r_1, r_2}(1) [W(r_2)^2 - W(r_1)^2 - r_w] \end{bmatrix}.$$

Under the null hypothesis that $\alpha_{r_1, r_2} = T^{-\eta}$ and $\beta_{r_1, r_2} = 0$,

$$\begin{bmatrix} \sqrt{T} (\hat{\alpha}_{r_1, r_2} - \alpha_{r_1, r_2}) \\ T \hat{\beta}_{r_1, r_2} \end{bmatrix} \xrightarrow{L} \begin{bmatrix} r_w & A_{r_1, r_2} \\ A_{r_1, r_2} & B_{r_1, r_2} \end{bmatrix}^{-1} \begin{bmatrix} C_{r_1, r_2} \\ D_{r_1, r_2} \end{bmatrix},$$

where

$$A_{r_1, r_2} = \psi_{r_1, r_2}(1) \sigma_{r_1, r_2} \int_{r_1}^{r_2} W(s) ds, \\ B_{r_1, r_2} = \sigma_{r_1, r_2}^2 \psi_{r_1, r_2}(1)^2 \int_{r_1}^{r_2} W(s)^2 ds, \\ C_{r_1, r_2} = \sigma_{r_1, r_2} [W(r_2) - W(r_1)],$$

$$D_{r_1, r_2} = \frac{1}{2} \sigma_{r_1, r_2}^2 \psi_{r_1, r_2}(1) \left[W(r_2)^2 - W(r_1)^2 - r_w \right].$$

Therefore, $\hat{\beta}_{r_1, r_2}$ converges at rate T to the following limit variate

$$T \hat{\beta}_{r_1, r_2} \xrightarrow{L} \frac{A_{r_1, r_2} C_{r_1, r_2} - r_w D_{r_1, r_2}}{A_{r_1, r_2}^2 - r_w B_{r_1, r_2}}.$$

To calculate the t-statistic $t_{r_1, r_2} = \frac{\hat{\beta}_{r_1, r_2}}{se(\hat{\beta}_{r_1, r_2})}$ of $\hat{\beta}_{r_1, r_2}$, we first find the standard error $se(\hat{\beta}_{r_1, r_2})$. We have

$$\text{var} \left(\begin{bmatrix} \hat{\alpha}_{r_1, r_2} \\ \hat{\beta}_{r_1, r_2} \end{bmatrix} \right) = \sigma_{r_1, r_2}^2 \begin{bmatrix} \Sigma' 1 & \Sigma' y_{t-1} \\ \Sigma' y_{t-1} & \Sigma' y_{t-1}^2 \end{bmatrix}^{-1},$$

so the variance of $T \hat{\beta}_{r_1, r_2}$ can be calculated from

$$\begin{aligned} & \text{var} \left(\begin{bmatrix} \sqrt{T} (\hat{\alpha}_{r_1, r_2} - \alpha_{r_1, r_2}) \\ T \hat{\beta}_{r_1, r_2} \end{bmatrix} \right) \\ &= \sigma_{r_1, r_2}^2 \left\{ \begin{bmatrix} \sqrt{T} & 0 \\ 0 & T \end{bmatrix}^{-1} \begin{bmatrix} \Sigma' 1 & \Sigma' y_{t-1} \\ \Sigma' y_{t-1} & \Sigma' y_{t-1}^2 \end{bmatrix} \begin{bmatrix} \sqrt{T} & 0 \\ 0 & T \end{bmatrix}^{-1} \right\}^{-1} \\ &= \sigma_{r_1, r_2}^2 \begin{bmatrix} T^{-1} \Sigma' 1 & T^{-3/2} \Sigma' y_{t-1} \\ T^{-3/2} \Sigma' y_{t-1} & T^{-2} \Sigma' y_{t-1}^2 \end{bmatrix}^{-1} \xrightarrow{L} \sigma_{r_1, r_2}^2 \begin{bmatrix} r_w & A_{r_1, r_2} \\ A_{r_1, r_2} & B_{r_1, r_2} \end{bmatrix}^{-1}. \end{aligned}$$

It follows that the t-statistic t_{r_1, r_2} of $\hat{\beta}_{r_1, r_2}$ satisfies

$$t_{r_1, r_2} \xrightarrow{L} \frac{\frac{1}{2} r_w \left[W(r_2)^2 - W(r_1)^2 - r_w \right] - \int_{r_1}^{r_2} W(s) ds [W(r_2) - W(r_1)]}{r_w^{1/2} \left\{ r_w \int_{r_1}^{r_2} W(s)^2 ds - \left[\int_{r_1}^{r_2} W(s) ds \right]^2 \right\}^{1/2}}.$$

By continuous mapping the asymptotic distribution of the GSADF statistic is

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

giving the stated result. ■