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

COWLES FOUNDATION PAPER NO. 1498

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

2015

http://cowles.yale.edu/

INTERNATIONAL ECONOMIC REVIEW

November 2015
Vol. 56, No. 4

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

BY PETER C. B. PHILLIPS, SHUPING SHI, AND JUN YU[1]

Yale University, U.S.A., University of Auckland, New Zealand, University of Southampton, U.K., and Singapore Management University, Singapore; Macquarie University and CAMA, Australia; Singapore Management University, Singapore

Recent work on econometric detection mechanisms has shown the effectiveness of recursive procedures in identifying and dating financial bubbles in real time. 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, this article 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 augmented Dickey–Fuller (ADF) test of Phillips et al. ("Explosive behavior in the 1990s NASDAQ: When did exuberance escalate asset values?" *International Economic Review* 52 (2011), 201–26; PWY) and delivers a consistent real-time 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 collapses over this period, whereas the strategy of PWY and a related cumulative sum (CUSUM) dating procedure locate far fewer episodes in the same sample range.

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 (Ferguson, 2008; Ahamed, 2009), 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 policymakers

*Manuscript received July 2015.

[1] This article and its technical companion "Testing for Multiple Bubbles: Limit Theory of Real Time Detectors" (Phillips et al., 2015a, this issue) 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 the editor and three referees for helpful comments, as well as many colleagues, seminar participants, and central bank economists for valuable discussions. Phillips acknowledges support from the NSF under grant numbers SES 09-56687 and SES 12-58258. Shi acknowledges the Financial Integrity Research Network (FIRN) for funding support. Yu acknowledges support from the Singapore Ministry of Education for Academic Research Fund under grant number MOE2011-T2-2-096. Please address correspondence to: Peter C. B. Phillips, Cowles Foundation for Research in Economics, Yale University, Box 208281, New Haven, CT 06520-8281, U.S.A. E-mail: *peter.phillips@yale.edu*.

have been affirming the Basel 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?

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 algorithms that use data only up to the point of analysis for ongoing assessment, giving an early warning diagnostic that can assist regulators in market monitoring. If history has a habit of repeating itself and human learning mechanisms do fail, as financial historians such as Ferguson (2008)[2] assert, then quantitative warnings may serve as useful alert mechanisms to both market participants and regulators in real time.

Several attempts to develop ex post 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 that can detect exuberance in asset price series during an inflationary phase. The approach is ex ante (or 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 century[3]), 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.

This article 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) to test for the presence of a bubble based on sequence of forward recursive right-tailed ADF unit root tests. PWY then proposed a dating strategy, which identifies points of origination and termination of a bubble based on a backward regression technique. When there is a single bubble in the data, it is known that this dating strategy is consistent, as was first shown in an unpublished working paper by Phillips and Yu (2009) whose results are subsumed as a special case within this work. Other break testing procedures such as Chow tests, model selection, and CUSUM tests may also be applied as dating mechanisms. Extensive simulations conducted

[2] "Nothing illustrates more clearly how hard human beings find it to learn from history than the repetitive history of stock market bubbles" (Ferguson, 2008).

[3] "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).

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.

As shown here, when the sample period includes multiple episodes of exuberance and collapse, the PWY procedures 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, this article proposes a generalized sup ADF (GSADF) method to test for the presence of bubbles as well as a recursive backward regression technique to time-stamp the bubble origination and termination dates. Like PWY, the new procedures rely on recursive right-tailed ADF tests but use flexible window widths in their 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 start point and the endpoint of the recursion over a feasible range of flexible windows. This test is therefore a right-sided double recursive test for a unit root and is analogous to double recursive left-sided ex post tests of persistence such as that considered in Leybourne et al. (2007).

The new dating strategy is an ex ante procedure and extends the dating strategy of PWY by changing the start point in the real-time analysis. Since the new procedures cover more subsamples of the data and have greater window flexibility, they are designed to outperform the PWY procedures in detecting explosive behavior when multiple episodes occur in the data. This expected enhancement in performance by the new procedures is demonstrated here in simulations, which compare the two methods in terms of their size and power in bubble detection. Moreover, 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. Throughout the article *consistency* refers to consistency in determining the relevant sample fraction of the break point instead of the sample observation, as is usual in structural break asymptotic theory.

In addition to the GSADF test and ex ante dating algorithm, a modified version of the original PWY algorithm is developed in which the detection procedure is repeated sequentially with re-initialization after the detection of each bubble. 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 a 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 approach successfully identifies all the well-known historical episodes of exuberance over this period, including the great crash, the postwar boom in 1954, Black Monday in October 1987, and the dot-com bubble. The strategy of PWY is much more conservative and locates only a single episode over the same historical period, catching the 1990s stock bubble. The sequential PWY algorithm is similarly conservative in detecting bubbles in this data set, as is the CUSUM procedure.

The organization of the article 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 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 of the main results under the null are given in the Appendix.

A companion paper in this journal (Phillips et al., 2015a) develops the limit theory and consistency properties of the dating procedures of this article covering both single and

multiple-bubble scenarios. An Online Supplement (Phillips et al., 2015b) provides some further robustness checks on the empirical findings of this article together with additional background results that are needed for the mathematical derivations of the limit theory in this article and the companion paper. Gauss and Matlab computer codes are available online[4] for all the procedures developed herein. An add-in for the Eviews software package is also available (Caspi, 2013) that makes for convenient practical implementation of this article's methods.

2. A ROLLING WINDOW TEST FOR BUBBLES

2.1. *Models and Specification.* A common starting point in the analysis of financial bubbles is the asset pricing equation[5]:

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

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

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

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 (and hence the price–dividend ratio) 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 or the price–dividend ratio may be used to infer the existence of bubbles.[6]

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, e.g., the discussion in Cochrane, 2005, pp. 402–4) and their relevance in empirical work (notably, Pástor and Veronesi, 2006, but note also the strong critique of that view in Cooper, 2008[7]). There is greater agreement on the existence of market exuberance (which may be rational or irrational depending on possible links to market fundamentals), crises,

[4] https://sites.google.com/site/shupingshi/PrgGSADF.zip?attredirects=0

[5] Although it is common to focus on rational bubbles (Blanchard, 1979; Diba and Grossman, 1988), our reduced-form empirical approach accommodates other bubble-generating mechanisms such as intrinsic bubbles (Froot and Obstfeld, 1991), herd behavior (Avery and Zemsky, 1998; Abreu and Brunnermeier, 2003), and time-varying discount factor fundamentals (Phillips and Yu, 2011). Shi (2011) provides a partial overview of the literature.

[6] This 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

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

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, and $\bar{p}$, $\bar{d}$, and $\bar{u}$ are the respective sample means of p_t^f, d_t, and u_t. The degree of nonstationary of p_t^f is determined by that of d_t and u_t. In the absence of bubbles, the log-linear approximation yields a cointegrated relationship between p_t and d_t. Lee and Phillips (2011) provide a detailed analysis of the accuracy of this log-linear approximation under various conditions.

[7] "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).

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 runups 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 (instead of 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. These issues also arise in right-tailed unit root tests of the type used in bubble detection, as studied recently in Phillips et al. (2014). 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:

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

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 $T \to \infty$. 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 of y_t, and when $\eta < \frac{1}{2}$, the standardized output $T^{-1/2}y_t$ behaves asymptotically like a Brownian motion with drift. In this article, we focus on the case of $\eta > \frac{1}{2}$ where the order of magnitude of y_t is the same as that of a pure random walk (i.e., the null of PWY).[8]

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 $r_w > 0$ is the (fractional) window size of the regression. The empirical regression model can then be written as

$$\Delta y_t = \hat{\alpha}_{r_1,r_2} + \hat{\beta}_{r_1,r_2} y_{t-1} + \sum_{i=1}^{k} \hat{\psi}^i_{r_1,r_2} \Delta y_{t-i} + \hat{\varepsilon}_t, \tag{4}$$

[8] The procedure may also be used to detect bubbles in a data series where $\eta \leq \frac{1}{2}$, as shown in Phillips et al. (2014). In this case, the asymptotic distribution of the test statistic and, hence, the test critical values are different. Estimation of the localizing coefficient η is discussed in Phillips et al. (2014). When $\eta > 0.5$, 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 $\eta = \frac{1}{2}$.

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

We proceed to use rolling regressions of this type for bubble detection. The formulation is particularly useful in the case of multiple bubbles and includes the earlier SADF test procedure developed and used in PWY, which we now briefly review.

2.2. *The PWY Test for Bubbles.* The PWY 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 (fraction) r_w expands from r_0 to 1, where r_0 is the smallest sample window width fraction (which initializes computation of the test statistic) 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 endpoint 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 PWY test is then a sup statistic based on the forward recursive regression and is simply 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 approach to detecting explosive behavior. One alternative is the two-regime Markov-switching unit root test of Hall et al. (1999). This procedure offers some appealing features like regime probability estimation. But recent simulation work by Shi (2013) 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 et al. (1994) and van Norden and Vigfusson (1998), filtering algorithms find it difficult to distinguish periods, which may appear spuriously explosive due to high variance and periods where there is genuine explosive behavior. Further, to the best of our knowledge, the asymptotic properties of the Markov switching unit root test are presently unknown and require investigation. A related approach within the Markov switching framework is the use of Bayesian methods to analyze explosive behavior—see, for example, Shi and Song (2014). Analytic and simulation-based comparisons of the methods proposed in this article with Markov switching unit root tests are worthy of further research but are beyond the scope of this article.

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 bubbles, we focus attention in this article on extending the SADF test. However, our simulations and empirical implementation provide some comparative results with the CUSUM procedure in view of its good overall performance recorded in the Homm and Breitung simulations.

2.3. *The Rolling Window GSADF Test for Bubbles.* The GSADF test developed here pursues the idea of repeated ADF test regressions (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 endpoint 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, that is, from 0 to $r_2 - r_0$. We define the GSADF statistic to be the largest ADF statistic in this

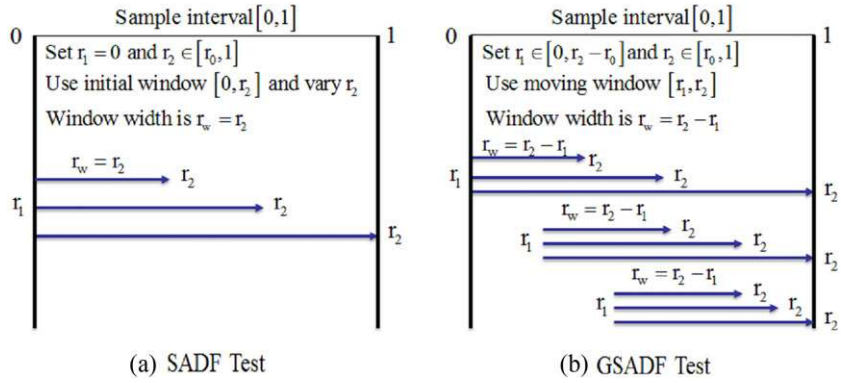

FIGURE 1

THE SAMPLE SEQUENCES AND WINDOW WIDTHS OF THE SADF TEST AND THE GSADF TEST

double recursion over all feasible ranges of r_1 and r_2, and we denote this statistic by $GSADF(r_0)$. That is,

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

Figure 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*

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

where $r_w = r_2 - r_1$ and W is a standard Wiener process.

The proof of Theorem 1 is given in the Appendix. Following Zivot and Andrews (1992) the proof uses a continuous mapping argument for the functional defining $GSADF(r_0)$ in terms of partial sums of the innovations, instead of a conventional "fidi + tightness" argument. This approach more easily accommodates the double recursion-based sup statistic. The argument given in the Appendix is also useful in justifying the limit theory for double recursive left-sided unit root tests against stationary alternatives where the tests involve inf instead of sup statistics, such as that of Leybourne et al. (2007).

The limit distribution (5) 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 (5) with $r_1 = 0$ and $r_2 = r_w = 1$ whereas the limit distribution of the single recursive SADF statistic is a further special case of (5) with $r_1 = 0$ and $r_2 = r_w \in [r_0, 1]$ (see Phllips et al., 2014). We conjecture that the limit theory (5) also continues to hold when the null is a unit root process with asymptotically negligible drift and with innovations that satisfy the error condition (EC) in the Appendix under suitable moment conditions and provided the lag order $k \to \infty$ with $k = o(T^{1/3})$ as $T \to \infty$ in

TABLE 1
THE ASYMPTOTIC AND FINITE SAMPLE CRITICAL VALUES OF THE SADF AND GSADF TESTS AGAINST AN EXPLOSIVE ALTERNATIVE

	SADF			GSADF		
	90%	95%	99%	90%	95%	99%
Asymptotic critical values						
$r_0 = 0.190$	1.10	1.37	1.88	1.67	1.89	2.37
$r_0 = 0.137$	1.12	1.41	2.03	1.78	2.01	2.48
$r_0 = 0.100$	1.20	1.49	2.07	1.97	2.19	2.69
$r_0 = 0.074$	1.21	1.51	2.06	1.99	2.20	2.62
$r_0 = 0.055$	1.23	1.51	2.06	2.08	2.30	2.74
Finite sample critical values						
$T = 100, r_0 = 0.190$	0.98	1.30	1.99	1.65	2.00	2.57
$T = 200, r_0 = 0.137$	1.12	1.40	1.90	1.84	2.08	2.70
$T = 400, r_0 = 0.100$	1.19	1.49	2.05	1.92	2.20	2.80
$T = 800, r_0 = 0.074$	1.25	1.53	2.03	2.10	2.34	2.79
$T = 1600, r_0 = 0.055$	1.28	1.57	2.22	2.19	2.41	2.87

NOTES: The asymptotic critical values are obtained by numerical simulations with 2000 replications. The Wiener process is approximated by partial sums of $N(0, 1)$ with 2000 steps. The finite sample critical values are obtained from Monte Carlo simulation with 2000 replications.

the empirical regression (4) (cf. Berk, 1974; Said and Dickey, 1984; Zivot and Andrews, 1992; Xiao and Phillips, 1999).

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. However, as is commonplace in the asymptotic theory of break-test methodology, the limit theory in Theorem 1 requires r_0 be bounded away from zero as $T \to \infty$. Based on extensive simulation findings, we recommend a rule for choosing r_0 that is based on a lower bound of 1% of the full sample and has the simple functional form $r_0 = 0.01 + 1.8/\sqrt{T}$, which is convenient for implementation. This setting delivers satisfactory size and power performance for both the SADF and GSADF tests in the simulations reported in Section 4.

Table 1 tabulates the asymptotic critical values of the SADF and GSADF statistics (top panel) and their corresponding finite sample critical values (bottom panel). The asymptotic critical values are obtained from 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 Monte Carlo simulations with 2,000 replications. The parameters (d and η) in the null hypothesis are set to one.[9]

We observe the following phenomena: First, as the minimum window size r_0 decreases, critical values of both test statistics increase. For instance, when r_0 decreases from 0.190 to 0.055, the 95% asymptotic critical value of the GSADF statistic rises from 1.89 to 2.30. Second, critical values for the GSADF statistic are larger than those of the SADF statistic. For instance, when $r_0 = 0.10$, the 95% asymptotic critical value of the GSADF statistic is 2.19 whereas that of the SADF statistic is 1.49. Figure 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 $GSADF(0.1)$. Third, the finite sample and asymptotic critical values are almost identical for the case of $T = 400$ and $r_0 = 0.10$. The finite sample critical values go below (above) their asymptotic counterparts when $T < 400$ (when $T > 400$).

[9] From Phillips et al. (2014), 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.

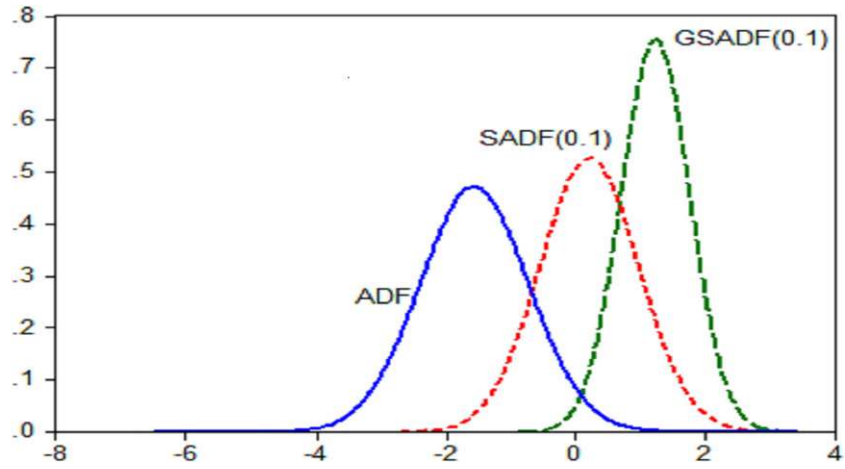

FIGURE 2

ASYMPTOTIC DISTRIBUTIONS OF THE ADF AND SUPADF STATISTICS ($r_0 = 0.1$)

3. DATE-STAMPING STRATEGIES

3.1. *The Method.* 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 as $\tau = \lfloor Tr \rfloor$ belongs to a bubble phase in the overall trajectory. The strategy suggested in PWY is to conduct a right-tailed recursive ADF test using data from the origination of the sample up to the present observation τ (i.e., the information embodied in $I_{\lfloor Tr \rfloor} = \left\{ y_1, y_2, \cdots, y_{\lfloor Tr \rfloor} \right\}$). Since it is possible that the data $I_{\lfloor Tr \rfloor}$ may include one or more collapsing bubble episodes, this ADF test, like earlier unit root and cointegration-based tests for bubbles (such as those studied in Diba and Grossman, 1988), may result in finding *pseudo stationary* behavior. As a result, the method is exposed to the criticism of Evans (1991) and is typically less successful in identifying subsequent bubbles after the first. The strategy recommended here is to perform a double recursive test procedure that we call a *backward sup ADF test* on $I_{\lfloor Tr \rfloor}$ to enhance identification accuracy. We use a flexible window in the double recursion similar to that described above.

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

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

The PWY procedure (i.e., the recursive ADF test) for bubble identification is a special case of the backward sup ADF test in which $r_1 = 0$, and in this case the sup operation is superfluous. We denote the corresponding ADF statistic by ADF_{r_2}. Figure 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 $\lfloor Tr_2 \rfloor$. The backward SADF test provides more information and improves detective capacity for bubbles within the sample because the subsample that gives rise to the maximum ADF statistic may not have the same generating mechanism as other observations within the

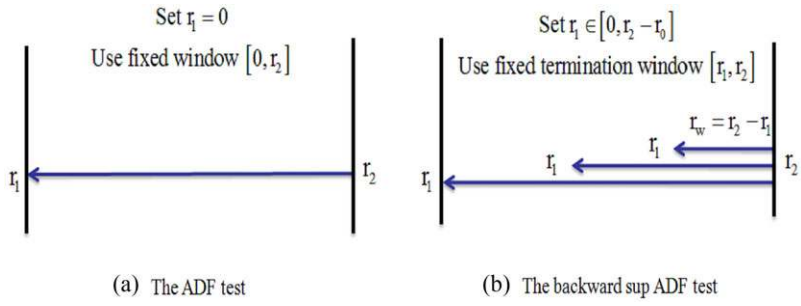

FIGURE 3

THE SAMPLE SEQUENCES OF THE ADF TEST AND THE BACKWARD SADF TEST

full sample from $r_1 = 0$ to r_2. This approach therefore has greater flexibility in the detection of multiple bubbles.

Like the PWY procedure, the feasible range of r_2 itself runs in a recursion from r_0 to 1. The origination date of a bubble $\lfloor Tr_e \rfloor$ is calculated as the first chronological observation whose ADF statistic exceeds the critical value. We denote the calculated origination date by $\lfloor T\hat{r}_e \rfloor$. The estimated termination date of a bubble $\lfloor T\hat{r}_f \rfloor$ is the first chronological observation after $\lfloor T\hat{r}_e \rfloor + L_T$ 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 formulas

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

where $cv_{r_2}^{\beta_T}$ is the $100(1 - \beta_T)\%$ critical value of the ADF statistic based on $\lfloor Tr_2 \rfloor$ observations. The significance level β_T depends on the sample size T and it is assumed that $\beta_T \to 0$ as $T \to \infty$. This control ensures that $cv_{r_2}^{\beta_T}$ diverges to infinity and thereby eliminates the type I error as $T \to \infty$. In practice, one can select a critical value $cv_{r_2}^{\beta_T}$ that diverges slowly to infinity like a slowly varying function (as in Phillips and Yu, 2011). In empirical applications, it is also often convenient to fix β_T at some predetermined level such as 0.05 instead of using a drifting significance level.

Under the new identification strategy, inference about potential explosiveness of the process at observation $\lfloor Tr_2 \rfloor$ is based on the backward sup ADF statistic $BSADF_{r_2}(r_0)$. Accordingly, 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 $\lfloor T\hat{r}_e \rfloor + \delta \log(T)$ whose backward sup ADF statistic falls below the critical value of the backward sup ADF statistic. For a bubble to be defined, it is here assumed that its duration should exceed a minimal period represented by $\delta \log(T)$, where δ is a frequency-dependent parameter.[10] 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:

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

[10] For instance, one may wish to impose a minimal duration condition that, to be classified as a bubble, duration should exceed a 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.

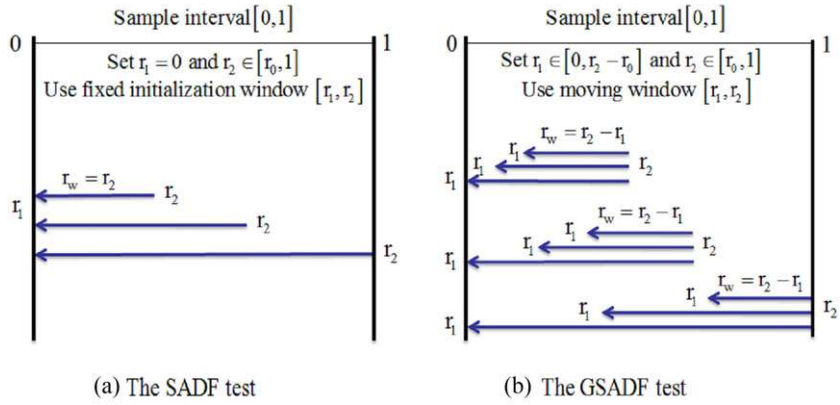

FIGURE 4

AN ALTERNATIVE ILLUSTRATION OF THE SAMPLE SEQUENCES AND WINDOW WIDTHS OF THE SADF TEST AND THE GSADF TEST

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

where $scv_{r_2}^{\beta_T}$ is the $100(1-\beta_T)\%$ critical value of the sup ADF statistic based on $\lfloor Tr_2 \rfloor$ observations. As in the PWY dating procedure and as discussed in Section 3, the significance level β_T may depend on the sample size T and pass to zero as the sample size approaches infinity.

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, $\left\{BSADF_{r_2}(r_0)\right\}_{r_2 \in [r_0,1]}$. Hence, the SADF and GSADF statistics can, respectively, be written as

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

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

Thus, the PWY date-stamping algorithm 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 Figure 4. Importantly, the new date-stamping strategy may be used as an ex ante real-time dating procedure, whereas the GSADF test is an ex post statistic used for analyzing a given data set for bubble behavior.

3.2. *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 et al., 2015a; PSY2). Additional technical material needed for those derivations is contained in the Online Supplement (Phillips et al., 2015b). The main results and import of the theory for empirical practice are reviewed here to make this article self-contained and to assist readers who wish to implement the test algorithms in practical work. 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 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 $\lfloor Tr_2 \rfloor$ is a special case of the GSADF test with $r_1 = 0$ and fixed r_2, and the backward sup ADF test is a special case of the GSADF test with 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:

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

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

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 $F_{r_2}(W, r_0)$. We know that $cv^{\beta_T} \to \infty$ and $scv^{\beta_T} \to \infty$ as $\beta_T \to 0$. Given $cv^{\beta_T} \to \infty$ and $scv^{\beta_T} \to \infty$ 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 $\Pr\left\{\hat{r}_e \in [r_0, 1]\right\} \to 0$ and $\Pr\left\{\hat{r}_f \in [r_0, 1]\right\} \to 0$.

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:

$$\begin{aligned} X_t = {} & X_{t-1}1\{t < \tau_e\} + \delta_T X_{t-1}1\left\{\tau_e \le t \le \tau_f\right\} \\ & + \left(\sum_{k=\tau_f+1}^{t} \varepsilon_k + X^*_{\tau_f}\right)1\left\{t > \tau_f\right\} + \varepsilon_t 1\left\{t \le \tau_f\right\}. \end{aligned} \tag{9}$$

In (9), $\delta_T = 1 + cT^{-\alpha}$ with $c > 0$ and $\alpha \in (0, 1)$, $\varepsilon_t \sim^{iid}(0, \sigma^2)$, $X^*_{\tau_f} = X_{\tau_e} + X^*$ with $X^* = O_p(1)$, $\tau_e = \lfloor Tr_e \rfloor$ dates the origination of bubble expansion, and $\tau_f = \lfloor Tr_f \rfloor$ 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, as discussed in PWY and Phillips and Shi (2014). 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.[11] 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 $T \to \infty$ provided the following rate condition on the critical value cv^{β_T} holds:

$$\frac{1}{cv^{\beta_T}} + \frac{cv^{\beta_T}}{T^{1-\alpha/2}} \to 0, \text{ as } T \to \infty. \tag{10}$$

[11] Consistent estimation of r_e also requires that the minimum window size $r_0 \le r_e$, else r_e is unidentified.

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) and (8), we again have consistency $(\hat{r}_e, \hat{r}_f) \xrightarrow{p} (r_e, r_f)$ as $T \to \infty$ under the corresponding rate condition on the critical value scv^{β_T}, viz.,

$$\frac{1}{scv^{\beta_T}} + \frac{scv^{\beta_T}}{T^{1-\alpha/2}} \to 0, \text{ as } T \to \infty. \tag{11}$$

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) 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 $T^{1-\alpha/2}$. 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 $T \to \infty$, 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 that extends the prototypical model (9):

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

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 $\tau_{1f} = \left\lfloor Tr_{1f} \right\rfloor$ are the origination and termination dates of the first bubble; $\tau_{2e} = \lfloor Tr_{2e} \rfloor$ and $\tau_{2f} = \left\lfloor Tr_{2f} \right\rfloor$ 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 (see Phillips and Shi, 2014).

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):

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

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

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:

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

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

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

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

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:

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

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 is summarized here as follows[12]:

(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 (10) 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 PSY 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 (Seq. 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.

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

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 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 $\hat{\delta}_{1,t}$ 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 $N_0 \cup B_1 \cup N_1$, which can lead to estimates $\hat{\delta}_{1,t} < 1$. 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.[13]

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 this article. We look at size, power, and detection capability for multiple-bubble episodes.

4.1. *Size Comparisons.* We concentrate on the SADF and GSADF tests. The two data-generating processes for size comparison are the unit root null hypothesis in (3) and a unit root process with a GARCH volatility structure, as suggested by a referee. Size is calculated based on the asymptotic critical values displayed in Table 1 using a nominal size of 5%. The number of replications is 2,000. We consider cases with sample sizes $T = \{100, 200, 400, 800, 1{,}600\}$. The minimum window size is chosen based on the rule $r_0 = 0.01 + 1.8/\sqrt{T}$.

4.1.1. *Data-generating process I: The unit root null* (3). We first examine how frequently the SADF and GSADF tests reject the null of a unit root in favor of an explosive alternative (giving a false positive result) if we simulate data from a unit root process with no conditional heteroskedasticity. In particular, simulations are based on the null model of (3) with $d = \eta = 1$. Table 2 reports actual test size when the nominal size is 5%. Evidently, when $k = 0$ (no transient dynamics are present in the system), there are no serious size distortions in the tests.

We further explore the effects of incorporating transient dynamic lag length selection in the test procedures. We consider fixed lags (1, 3, and 6), BIC order selection, and sequential significance testing. Maximum lag lengths for both BIC and the significance test are set at 3 and 6. Simulated sizes for these cases are reported in Table 2. Evidently, size increases with lag length (k) for both tests. In particular, when $T = 400$ and $r_0 = 0.100$, the size of the SADF (GSADF) test rises from 4.2% to 9% (5.5% to 40.1%, respectively) when k increases from 0 to 6. Thus, size distortion is more severe for GSADF than SADF when lag order is overspecified. The greater distortion for the GSADF test is explained by the use of the smaller sample sizes that occur with this flexible window procedure. Use of BIC lag order selection satisfactorily controls size for SADF but is less adequate for GSADF where the test is oversized, again no

[13] To 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, that is, $r_0 < r_{2e} - r_{1f}$. See PSY for more discussion.

TABLE 2
SIZES OF THE SADF AND GSADF TESTS WITH ASYMPTOTIC CRITICAL VALUES

	Fixed Lag				BIC		Significance Test	
	$k = 0$	$k = 1$	$k = 3$	$k = 6$	Max 3	Max 6	Max 3	Max 6
$T = 100$ and $r_0 = 0.190$								
SADF	0.041	0.065	0.092	0.184	0.055	0.066	0.077	0.145
GSADF	0.061	0.130	0.290	0.787	0.189	0.197	0.260	0.697
$T = 200$ and $r_0 = 0.137$								
SADF	0.049	0.051	0.082	0.134	0.051	0.064	0.073	0.095
GSADF	0.058	0.096	0.260	0.569	0.177	0.180	0.231	0.465
$T = 400$ and $r_0 = 0.100$								
SADF	0.042	0.049	0.066	0.090	0.046	0.048	0.060	0.085
GSADF	0.055	0.103	0.198	0.401	0.148	0.154	0.158	0.340
$T = 800$ and $r_0 = 0.074$								
SADF	0.054	0.046	0.064	0.084	0.050	0.049	0.054	0.076
GSADF	0.069	0.091	0.188	0.339	0.133	0.147	0.168	0.308
$T = 1600$ and $r_0 = 0.055$								
SADF	0.060	0.072	0.064	0.081	0.059	0.059	0.075	0.078
GSADF	0.070	0.101	0.163	0.262	0.129	0.133	0.150	0.252

NOTES: The data-generating process is (3) with $d = \eta = 1$. The nominal size is 5%. Size calculations are based on 5,000 replications.

doubt because of the shorter samples involved in the use of the flexible window. The significance test procedure leads to even greater size distortion, particularly when the maximum lag length ($k_{\max}$) is large. For instance, when $T = 100, r_0 = 0.190$, and $k_{\max} = 6$, the size of SADF and GSADF is 0.145 and 0.697, respectively, indicating significant distortion in both tests.

Overall, size is reasonably well controlled when a small fixed lag length is used in the recursive tests. This approach is therefore recommended for empirical use of the SADF and GSADF test procedures as well as the dating algorithms that are implemented in the application later in the article.

4.1.2. *Data-generating process II: A unit root model with GARCH errors.* We next examine how frequently the tests reject the null of a unit root in favor of an explosive root (giving a false positive result) when data are simulated with conditional heteroskedasticity. The following unit root model with GARCH errors is used:

$$y_t = dT^{-\eta} + y_{t-1} + \varepsilon_t, \tag{20}$$

$$\varepsilon_t = v_t\sqrt{h_t}, h_t \sim^{iid} (0, 1), \tag{21}$$

$$h_t = \omega + \alpha\varepsilon_{t-1}^2 + \beta h_{t-1}. \tag{22}$$

We set $d = \eta = 1$, $y_0 = 376.8$, $\omega = 30.69, \alpha = 0$, and $\beta = 0.61$. The parameters used in the volatility equation are maximum likelihood estimates obtained from the S&P 500 price–dividend ratio (described in the application below) over the sample period January 2004 to December 2007, a relatively tranquil period that reduces the influence of potential structural changes on the fitted GARCH parameters. The sample size T varies from 100 to 1,600. We set the transient lag length k to the correct value zero. The empirical sizes of the SADF and GSADF tests are reported in Table 3. As is evident in the table, the presence of conditional heterogeneity of this

TABLE 3
SIZES OF THE SADF AND GSADF TESTS WITH GARCH ERRORS USING ASYMPTOTIC CRITICAL VALUES

	$T = 100$ $r_0 = 0.190$	$T = 200$ $r_0 = 0.137$	$T = 400$ $r_0 = 0.100$	$T = 800$ $r_0 = 0.074$	$T = 1,600$ $r_0 = 0.055$
SADF	0.047	0.051	0.049	0.045	0.062
GSADF	0.066	0.066	0.060	0.063	0.072

NOTES: The data-generating process is (20) and (22). Nominal size is 5%. Size calculations are based on 2,000 replications.

magnitude in the asset price dynamics does not have a substantial impact on the size of either test.[14]

4.2. *Power Comparisons.* Discriminatory power in detecting bubbles is determined for two different generating models—the Evans (1991) collapsing bubble model (see 23–26 below) and an extended version of the PWY bubble model (given by (9) and (12)).

4.2.1. *Collapsing bubble alternatives.* We first simulate asset price series based on the Lucas asset pricing model and the 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 $B_t = 0$ for all t to obtain[15]

$$D_t = \mu + D_{t-1} + \varepsilon_{Dt}, \varepsilon_{Dt} \sim N\left(0, \sigma_D^2\right), \tag{23}$$

$$P_t^f = \frac{\mu\rho}{(1-\rho)^2} + \frac{\rho}{1-\rho} D_t, \tag{24}$$

and the Evans bubble component

$$B_{t+1} = \rho^{-1} B_t \varepsilon_{B,t+1}, \qquad \text{if} \quad B_t < b, \tag{25}$$

$$B_{t+1} = \left[\zeta + (\pi\rho)^{-1} \theta_{t+1} (B_t - \rho\zeta)\right] \varepsilon_{B,t+1}, \qquad \text{if} \quad B_t \geq b. \tag{26}$$

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 (25) and (26) state that a bubble grows explosively at rate ρ^{-1} when its size is less than b, whereas if the size is greater than b, the

[14] A comprehensive analysis of the impact of different volatility structures on the size of the SADF test can be found in Harvey et al. (2014). They recommend using a wild bootstrap procedure to reduce size distortion when volatility is nonstationary. The wild bootstrap also seems to be effective in controlling the size of unit root tests against explosive behavior when there are near IGARCH effects in conditional volatility.

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

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

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

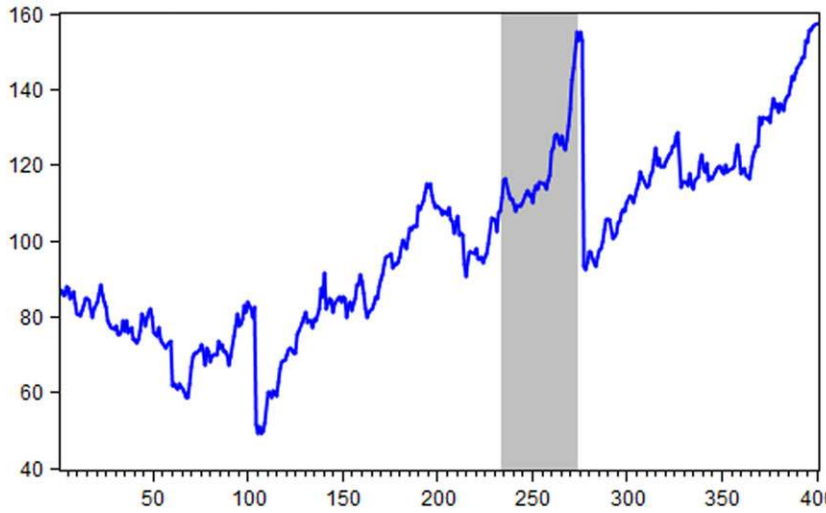

FIGURE 5

SIMULATED TIME SERIES OF $P_t = P_t^f + \kappa B_t$ USING THE EVANS COLLAPSING BUBBLE MODEL (23) TO (26) WITH SAMPLE SIZE 400

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 are as follows: $\mu = 0.0024$, $\sigma_D^2 = 0.0010$, $D_0 = 1.0$, $\rho = 0.985$, $b = 1$, $B_0 = 0.50$, $\pi = 0.85$, $\zeta = 0.50$, $\tau = 0.05$, and $\kappa = 20$. The parameter values for 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 later. We allow the discount factor ρ to vary from 0.975 to 0.999 and the probability of bubble survival π to vary from 0.99 to 0.25 in the power comparisons.[16] Figure 5 depicts a realization of this data-generating process. 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 subsample, which contains fewer collapsing episodes. The lag order k is set to zero. The smallest window size considered in the SADF test for the whole sample contains 40 observations (setting $T = 400$, $r_0 = 0.1$). The SADF statistic for the full trajectory is 0.86,[17] which is smaller than the 10% asymptotic critical value 1.20. 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 contains 27 observations (setting $T = 200$, $r_0 = 0.137$). The SADF statistic obtained from this sample is 2.54,[18] which is greater than the 5% asymptotic critical value 1.49. 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.

[16] The yearly parameter setting in Evans (1991) was also considered. Similar results were obtained and are omitted for brevity.

[17] The 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 273rd observation).

[18] This value comes from the subsample regression starting with the 201st observation up to the 273rd observation.

TABLE 4
POWERS OF THE SADF AND GSADF TESTS

	$T = 100$ $r_0 = 0.190$	$T = 200$ $r_0 = 0.137$	$T = 400$ $r_0 = 0.100$	$T = 800$ $r_0 = 0.074$	$T = 1{,}600$ $r_0 = 0.055$
SADF	0.589	0.738	0.829	0.926	0.885
GSADF	0.697	0.870	0.941	0.993	1.000

NOTES: The data-generating process is (23) and (26). Power calculations are based on 2,000 replications.

TABLE 5
POWERS OF THE SADF AND GSADF TESTS

ρ	0.975	0.980	0.985	0.990
SADF	0.927	0.872	0.829	0.727
GSADF	0.984	0.970	0.941	0.850
π	0.95	0.85	0.75	0.25
SADF	0.891	0.829	0.784	0.462
GSADF	0.957	0.941	0.926	0.517

NOTES: The data-generating process is (23) and (26) with sample size 400 ($r_0 = 0.1$). Power calculations are based on 2,000 replications.

These two experiments can be viewed as particular (fixed starting point) 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, whereas 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 3.37, which substantially exceeds the 1% asymptotic critical value 2.69. This value of the statistic is obtained from the subsample regression, which covers the most significant expansionary period (the one that spans the 234th to 273rd observations). This subsample differs from those used to obtain the earlier SADF statistics and leads to a far larger value of the test statistic. This is unsurprising because the statistic is based on the sample period that displays the strongest evidence of explosive behavior in the sample, as shown in the shaded area of Figure 5. Thus, the GSADF test finds strong evidence of bubbles in this simulated trajectory. Compared to the SADF test, the GSADF identifies bubbles through endogenous subsample determination, giving an obvious improvement over SADF that is useful in empirical applications.[19]

We now discuss results from the full simulation with 2,000 replications. The data-generating process is the periodically collapsing explosive process given in (23)–(26). We calculate the powers of both SADF and GSADF tests under this DGP with $T = \{100, 200, 400, 800, 1{,}600\}$ using asymptotic critical values (95th percentiles of the asymptotic distributions). The results are reported in Table 4.

As evident in Table 4, there are uniform improvements in power from using the GSADF test. When $T = 400$, the GSADF test raises the power of SADF from 82.9% to 94.1%, an 11.2% gain. Similar gains in power of 10.8%, 13.2%, 6.7%, and 11.5% occur for $T = 100$, 200, 800, 1,600. Interestingly, whereas the power of GSADF rises uniformly with the sample size, the powers of SADF do not rise uniformly with the sample size, showing increases for $T = 100$ to $T = 800$ but decreasing from $T = 800$ to $T = 1{,}600$.

In Table 5, we compare powers of the SADF and GSADF tests with the discount factor ρ varying from 0.975 to 0.990 and the probability of bubble survival π varying from 0.95 to 0.25 in the DGP. The expected number of bubble collapses in the sample period is always greater than

[19] Similar 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 this alternative data-generating process had the following settings: $B_0 = 0.5$, $b = 1$, $\pi = 0.85$, $\zeta = 0.5$, $\rho = 0.985$, $\tau = 0.05$, $\mu = 0.001$, $\ln D_0 = 1$, $\sigma^2_{\ln D} = 0.0001$, and $P_t = P_t^f + 500B_t$.

one and increases as π gets smaller. First, as is apparent in Table 5, the GSADF test has greater discriminatory power for detecting bubbles than the SADF test. The power improvement is 5.7%, 9.8%, 11.2%, and 12.3% for $\rho = \{0.975, 0.980, 0.985, 0.990\}$ and 6.6%, 11.2%, 14.2%, and 5.5% for $\pi = \{0.95, 0.85, 0.75, 0.25\}$. Second, since the rate of bubble expansion in this model is inversely related to the discount factor, powers of both SADF and GSADF are expected to decrease as ρ increases. The power of SADF (GSADF) declines from 92.7% to 72.7% (98.4% to 85%) as the discount factor rises from 0.975 to 0.990. Third, as it is easier to detect bubbles with longer duration, powers of both the SADF and GSADF tests increase with the probability of bubble survival. From Table 5, the powers of SADF and GSADF for $\pi = 0.95$ are almost twice as high as those for $\pi = 0.25$.

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.[20] The CUSUM detector is denoted by $C_{r_0}^r$ and defined as

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

where $\lfloor Tr_0 \rfloor$ is the training sample,[21] $\lfloor Tr \rfloor$ is the current monitoring observation, $\hat{\mu}_r$ is the mean of $\{\Delta y_1, ..., \Delta y_{\lfloor Tr \rfloor}\}$, and $r > r_0$. Under the null hypothesis of a pure random walk, the recursive statistic $C_{r_0}^r$ has the following asymptotic property (see Chu et al., 1996):

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

where $c_r = \sqrt{\kappa_\alpha + \log(r/r_0)}$.[22] 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 $ADF_t < cv_t^{0.95}$),[23] 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 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., $\tau_f - \tau_e = \lfloor 0.15T \rfloor$ and $\tau_e = \lfloor 0.4T \rfloor$). For the two-bubble experiment, the bubbles originate 20% and 60% into the sample, and the durations are $\lfloor 0.20T \rfloor$ and $\lfloor 0.10T \rfloor$, respectively. Figure 6 displays typical realizations of these two data-generating processes.

We summarize the findings for the main experiments based on the following specifications: For the single-bubble process (13) the bubble origination parameter τ_e was set to $\lfloor 0.2T \rfloor$, $\lfloor 0.4T \rfloor$, and $\lfloor 0.6T \rfloor$ and bubble duration varied from $\lfloor 0.10T \rfloor$ to $\lfloor 0.20T \rfloor$. Other parameter configurations were considered and produced broadly similar findings but are not reported here.[24] The simulations involved 5,000 replications. Bubbles were identified using respective finite sample

[20] Simulations 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.

[21] It is assumed that there is no structural break in the training sample.

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

[23] We 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.

[24] A more detailed analysis of the finite sample performance of the tests is given in our companion paper "Testing for Multiple Bubbles: Limit Theory of Real Time Detectors."

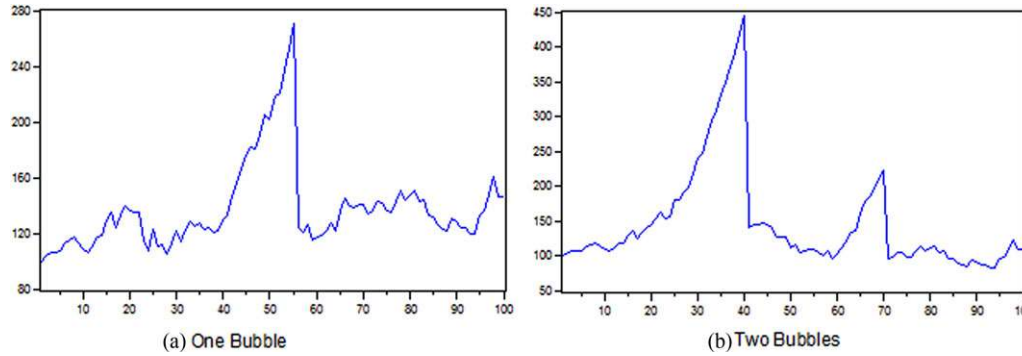

FIGURE 6

TYPICAL SAMPLE PATHS GENERATED ACCORDING TO (9) FOR PANEL (A) AND (12) FOR PANEL (B)

TABLE 6

FREQUENCIES OF DETECTING ONE, TWO, AND MORE BUBBLES FOR THE SINGLE-BUBBLE DGP WITH DIFFERENT BUBBLE DURATIONS AND LOCATIONS

	PWY			PSY			Seq PWY			CUSUM		
	One	Two	More	One	Two	More	One	Two	More	One	Two	More
$\tau_e = \lfloor \mathbf{0.2T} \rfloor$												
$\tau_f - \tau_e = \lfloor 0.10T \rfloor$	**0.73**	0.01	0.00	**0.65**	0.19	0.00	**0.58**	0.21	0.00	**0.41**	0.04	0.00
$\tau_f - \tau_e = \lfloor 0.15T \rfloor$	**0.90**	0.01	0.00	**0.72**	0.20	0.00	**0.64**	0.26	0.00	**0.54**	0.07	0.00
$\tau_f - \tau_e = \lfloor 0.20T \rfloor$	**0.94**	0.01	0.00	**0.75**	0.19	0.00	**0.67**	0.26	0.00	**0.61**	0.08	0.00
$\tau_e = \lfloor \mathbf{0.4T} \rfloor$												
$\tau_f - \tau_e = \lfloor 0.10T \rfloor$	**0.59**	0.06	0.00	**0.61**	0.20	0.00	**0.49**	0.22	0.00	**0.50**	0.10	0.00
$\tau_f - \tau_e = \lfloor 0.15T \rfloor$	**0.77**	0.07	0.00	**0.68**	0.22	0.00	**0.59**	0.27	0.00	**0.70**	0.13	0.00
$\tau_f - \tau_e = \lfloor 0.20T \rfloor$	**0.83**	0.08	0.00	**0.72**	0.23	0.00	**0.64**	0.27	0.00	**0.78**	0.14	0.00
$\tau_e = \lfloor \mathbf{0.6T} \rfloor$												
$\tau_f - \tau_e = \lfloor 0.10T \rfloor$	**0.53**	0.10	0.00	**0.58**	0.21	0.00	**0.49**	0.21	0.00	**0.57**	0.12	0.00
$\tau_f - \tau_e = \lfloor 0.15T \rfloor$	**0.69**	0.12	0.00	**0.64**	0.24	0.00	**0.59**	0.24	0.00	**0.70**	0.14	0.00
$\tau_f - \tau_e = \lfloor 0.20T \rfloor$	**0.76**	0.13	0.00	**0.66**	0.25	0.01	**0.66**	0.24	0.00	**0.75**	0.15	0.00

NOTES: Parameters are set as: $y_0 = 100$, $c = 1$, $\sigma = 6.79$, $\alpha = 0.6$, $T = 100$. Calculations are based on 5,000 replications. The minimum window has 12 observations. PWY is the method proposed by Phillips et al. (2011); PSY is the method proposed in this article. Sequential PWY is the method discussed in Subsection 3.2; CUSUM is the method discussed in Subsection 4.2.2.

95% quantiles, obtained from Monte Carlo simulations with 5,000 replications. The minimum window size had 12 observations. Tables 6 and 7 provide a selection of the results, reporting the detection frequencies for one, two, and three or more bubbles.[25] We calculated other summary statistics such as the mean and standard deviation of the number of bubbles identified within the sample range and the proportions of sample paths identified with bubbles, but these are not reported here. The main findings for the four procedures (PWY, PSY, sequential PWY, and CUSUM) are as follows:

1. The frequency of detecting the correct number of bubbles increases with the duration of bubble expansion (Tables 6 and 7) and with the value of the autoregressive coefficient δ_T, although this is not reported here. Hence, bubble detection is more successful when bubble duration is longer and the bubble expansion rate is faster.
2. For the single-bubble case, bubble location has some impact on relative performance among the PWY procedure, the PSY strategy, the sequential PWY approach, and the

[25] We imposed a bubble duration definition that required at least three expansionary observations to confirm the existence of a bubble in these calculations.

TABLE 7
FREQUENCIES OF DETECTING ONE, TWO, AND MORE BUBBLES FOR THE TWO-BUBBLE DGP WITH VARYING BUBBLE DURATION

	PWY			PSY			Seq PWY			CUSUM		
	One	Two	More	One	Two	More	One	Two	More	One	Two	More
$\tau_{1f} - \tau_{1e} = \lfloor \mathbf{0.10T} \rfloor$												
$\tau_{2f} - \tau_{2e} = \lfloor 0.10T \rfloor$	0.67	**0.13**	0.00	0.25	**0.62**	0.01	0.22	**0.55**	0.01	0.48	**0.12**	0.00
$\tau_{2f} - \tau_{2e} = \lfloor 0.15T \rfloor$	0.48	**0.39**	0.00	0.19	**0.69**	0.01	0.17	**0.63**	0.01	0.54	**0.27**	0.00
$\tau_{2f} - \tau_{2e} = \lfloor 0.20T \rfloor$	0.35	**0.57**	0.00	0.18	**0.72**	0.00	0.17	**0.67**	0.00	0.55	**0.34**	0.00
$\tau_{1f} - \tau_{1e} = \lfloor \mathbf{0.15T} \rfloor$												
$\tau_{2f} - \tau_{2e} = \lfloor 0.10T \rfloor$	0.89	**0.03**	0.00	0.21	**0.69**	0.00	0.18	**0.65**	0.01	0.55	**0.08**	0.00
$\tau_{2f} - \tau_{2e} = \lfloor 0.15T \rfloor$	0.76	**0.18**	0.00	0.12	**0.78**	0.00	0.10	**0.72**	0.01	0.54	**0.18**	0.00
$\tau_{2f} - \tau_{2e} = \lfloor 0.20T \rfloor$	0.48	**0.48**	0.00	0.09	**0.81**	0.00	0.08	**0.77**	0.00	0.49	**0.35**	0.00
$\tau_{1f} - \tau_{1e} = \lfloor \mathbf{0.20T} \rfloor$												
$\tau_{2f} - \tau_{2e} = \lfloor 0.10T \rfloor$	0.94	**0.02**	0.00	0.21	**0.71**	0.00	0.16	**0.71**	0.00	0.61	**0.09**	0.00
$\tau_{2f} - \tau_{2e} = \lfloor 0.15T \rfloor$	0.92	**0.05**	0.00	0.10	**0.81**	0.00	0.08	**0.78**	0.00	0.61	**0.10**	0.00
$\tau_{2f} - \tau_{2e} = \lfloor 0.20T \rfloor$	0.79	**0.18**	0.00	0.07	**0.85**	0.00	0.05	**0.83**	0.00	0.57	**0.19**	0.00

NOTES: Parameters are set as: $y_0 = 100, c = 1, \sigma = 6.79, \alpha = 0.6, \tau_{1e} = \lfloor 0.20T \rfloor, \tau_{2e} = \lfloor 0.60T \rfloor, T = 100$. Calculations are based on 5,000 replications. The minimum window has 12 observations. PWY is the method proposed by Phillips, Wu, and Yu (2011); PSY is the method proposed in this article. Sequential PWY is the method discussed in Subsection 3.2; CUSUM is the method discussed in Subsection 4.2.2.

CUSUM procedure (Table 6). When the bubble starts early in the sample and lasts only for a brief interval ($\tau_e = \lfloor 0.20T \rfloor$ and $\tau_f - \tau_e = \lfloor 0.10T \rfloor$), the frequency of detecting the correct bubble number (one) using the PWY procedure is higher than those of the other three procedures. The reason why the PWY procedure can slightly outperform the PSY procedure in this context is explained as follows: As shown in Theorem 1 of PSY2, when there is only one bubble in the sample, the unit root statistic used by PWY and the BSDF statistic used by PSY have the same order of magnitude under the alternative (namely, $T^{1-\alpha/2}$), although the latter quantity (for BSDF) by construction cannot be smaller than the former one. It follows that because the critical value of PWY is smaller than that of PSY (as shown in Table 1), it is easier for PWY to reject the null hypothesis, leading PWY to perform better than PSY when there is only a single bubble in the sample, particularly when it is early in the sample and there is little advantage in a flexible recursion. The advantage of the PSY strategy becomes clear when the bubble occurs later in the sample. For instance, when $\tau_e = \lfloor 0.40T \rfloor$ and $\tau_f - \tau_e = \lfloor 0.10T \rfloor$, the frequency of detecting one bubble using the PSY strategy is 2%, 12%, and 11% higher than that of PWY, sequential PWY, and CUSUM, respectively.[26]

3. The impact of bubble location is not as dramatic as the duration of bubble expansion. As is evident in Table 6, when the bubble expansion duration rises from $\lfloor 0.10T \rfloor$ to either $\lfloor 0.15T \rfloor$ or $\lfloor 0.20T \rfloor$, the performance of the PWY strategy exceeds the other three strategies regardless of the bubble location.
4. Overall, in the one-bubble scenario, the sequential PWY procedure tends to overestimate the bubble number, the PSY procedure slightly overestimates the bubble number, and the PWY and CUSUM estimators tend to be more accurate. For instance, when $\tau_e = \lfloor 0.40T \rfloor$ and $\tau_f - \tau_e = \lfloor 0.10T \rfloor$, the probability of detecting two bubbles is 22%, 20%, 6%, and 10% (the mean value of the bubble number estimates is 1.35, 1.24, 1.18, and 1.17), respectively, for the sequential PWY algorithm, the PSY strategy, the PWY procedure, and the CUSUM algorithm. The PWY procedure is more accurate than CUSUM.

[26] Neither of these results is unexpected and is explained as follows: When the bubble originates early in the sample, the benefit of the PSY procedure searching over a range of starting values is insignificant because the search range is narrow and the critical values of the PSY procedure are larger than those of the PWY strategy. On the other hand, when the bubble starts late in the sample, the search range becomes wider, giving an advantage to the flexible recursion of PSY, and leading to a larger test statistic and hence a significant power gain.

5. In the two-bubble scenario, bubble duration has 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 PWY procedure detects only one bubble for a majority of the sample paths, indicating substantial underestimation. For instance, when $\tau_{1f} - \tau_{1e} = \lfloor 0.2T \rfloor > \tau_{2f} - \tau_{2e} = \lfloor 0.1T \rfloor$, the frequency of detecting one bubble is 94%. In addition, the mean values of the PWY bubble number estimates are far from the true value and are close to unity. This is consistent with the asymptotic theory (established in our companion paper), 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. The performance of the PWY strategy improves significantly when the duration of the second bubble is longer than the first bubble. For instance, when $\tau_{1f} - \tau_{1e} = \lfloor 0.1T \rfloor < \tau_{2f} - \tau_{2e} = \lfloor 0.2T \rfloor$, the frequency of detecting two bubbles using the PWY strategy is 57%, which is much higher than when the first bubble has longer duration. This simulation finding again corroborates the asymptotic theory, showing that the PWY strategy can perform satisfactorily in detecting both bubbles under these conditions.
6. 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, just like the PWY procedure, there is obvious improvement in performance of the CUSUM procedure when the second bubble lasts longer (Table 7).
7. As expected, the sequential PWY procedure performs nearly as well as the PSY strategy in the two-bubble case but tends to have lower power and less accuracy than PSY.
8. 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, columns 2 and 3).

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 Figure 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 Equation (1). 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 and can be either I(1) or I(0) (e.g., Cochrane, 1992; Ang and Bekaert, 2007). 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. In addition, the unit root null hypothesis, as opposed to a stationary null, typically leads to relatively conservative outcomes.[27]

We first apply the summary SADF and GSADF tests to the price–dividend ratio. Table 8 presents the two test statistics. Also reported are the finite sample critical values of the two tests obtained obtained from 2,000 replications of 1,680 observations. In performing the ADF regressions and calculating the critical values, the smallest window contains 90 observations of the sample, based on the rule $r_0 = 0.01 + 1.8/\sqrt{1680}$. From Table 8, the SADF and GSADF statistics for the full data series are 3.30 and 4.21, obtained from subsamples 1871M01–2000M07 and 1976M04–1999M06, respectively. Both exceed their respective 1% right-tail critical values (i.e., $3.30 > 2.14$ and $4.21 > 2.74$), giving strong evidence that the S&P 500 price–dividend ratio

[27] It is well known that, ceteris paribus, critical values associated with a unit root null are higher than those from a stationary null.

TABLE 8
THE SADF TEST AND THE GSADF TEST OF THE S&P500 PRICE–DIVIDEND RATIO

		Finite Sample Critical Values		
	Test Stat.	90%	95%	99%
SADF	3.30	1.30	1.59	2.14
GSADF	4.21	2.17	2.34	2.74

NOTES: Critical values of both tests are obtained from Monte Carlo simulation with 2, 000 replications (sample size 1,680). The smallest window has 90 observations.

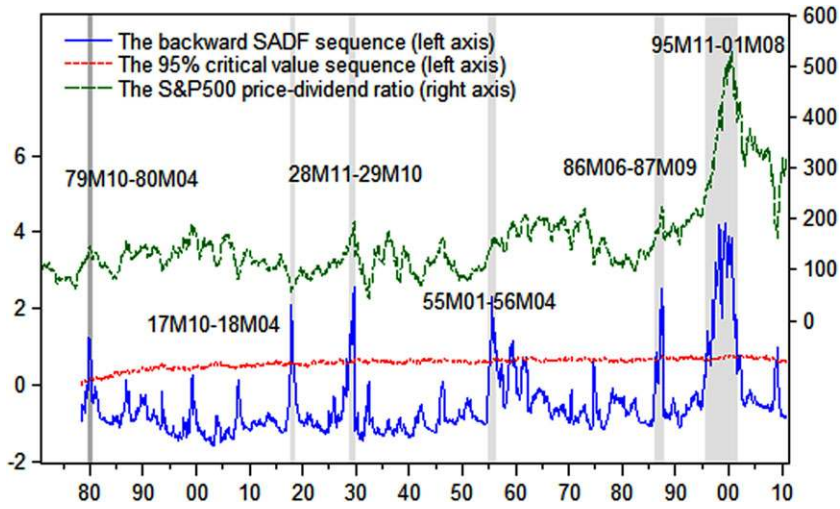

FIGURE 7

DATE-STAMPING BUBBLE PERIODS IN THE S&P 500 PRICE–DIVIDEND RATIO: THE GSADF TEST

had explosive subperiods.[28] 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 $k = 0$.

Next, we conduct a (pseudo) real-time bubble monitoring exercise for the S&P 500 stock market using PSY, PWY, sequential PWY, and CUSUM dating strategies. With a training period of 90 observations, we monitor the time series behavior of the price–dividend ratio for the market from June 1878 until the end of the sample period.

For the PSY real-time dating strategy, we compared the backward SADF statistic with the 95% SADF critical value (obtained from Monte Carlo simulations with 2,000 replications) for each observation of interest. From Figure 7, the identified periods of exuberance in the market include *the so-called post long-depression period* (1879M10–1880M04), *the great crash episode* (1928M11–1929M10), *the postwar boom in 1954* (1955M01–1956M04), *Black Monday in October 1987* (1986M06–1987M09), and *the dot-com bubble* (1995M11–2001M08). With regard to the *dot-com bubble*, the PSY strategy detects mildly explosive market behavior 5 years before the market crashes. The PSY strategy identifies two episodes related to market downturns instead of bubble expansion, namely *the 1917 stock market crash* (1917M08–1918M04) and *the subprime mortgage crisis* (2009M02–M04).

In the supplement (Phillips et al., 2015b), we show the same analysis conducted with a smaller window size of 36 observations.[29] The main episodes identified as bubbles in the data are robust

[28] Although the GSADF test is implemented here to detect explosiveness against I(1) behavior, it is expected to have stronger discriminative power against I(0) behavior.

[29] As a robustness check, we also applied the PSY procedure to the logarithm of the real S&P 500 price index (instead of the price–dividend ratio) and considered minimum window sizes of 48 and 60 observations (equivalent to

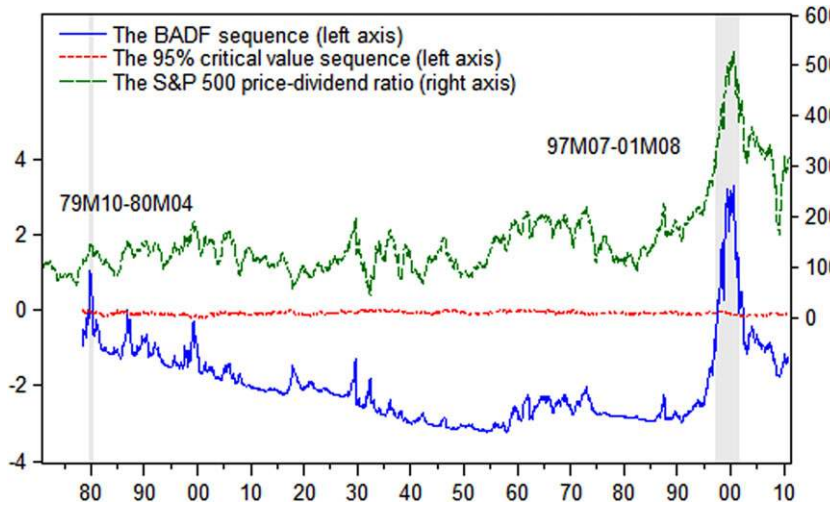

FIGURE 8

DATE-STAMPING BUBBLE PERIODS IN THE S&P 500 PRICE–DIVIDEND RATIO: THE SADF TEST

to this smaller value of r_0. Interestingly, the method identified several short crisis periods, including the banking panic of 1907 and the 1974 stock market crash, with the use of the smaller window size. The identification of crashes as bubbles may be caused by very rapid changes in the data. Harvey et al. (2014) note, for instance, the potential in the test for crash identification in the presence of nonstationary volatility, which may be addressed by using finite sample critical values based on the wild bootstrap.

Figure 8 plots the recursive ADF statistic against the corresponding 95% critical value sequence. We see that the PWY strategy (based on the SADF test) identifies only two episodes—*the post long-depression period* and *the dot-com bubble*. The estimated origination and termination dates for these two episodes are precisely the same as those from the PSY strategy.

Empirical results from the sequential PWY procedure are shown in Figure 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 Subsection 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, leading to a break in the calculated sequence. Hence, the sequential PWY strategy, just like PWY, has some inherent disadvantages in detecting multiple bubbles. In practice, one could potentially predivide the sample period into subsamples for testing, but, as shown in the example of Subsection 4.2.1, the subsample approach may well be sensitive to the preselection of the sample periods. The sequential ADF plot has two breaks in the figure, each corresponding to the re-initialization of the test procedure following a collapse. Findings from the sequential PWY test are identical to those from the PWY procedure, indicating two bubble episodes (*the post long-depression period* and *the dot-com bubble*).

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

four and five years). These adjustments produced only minor discrepancies, and the empirical results are qualitatively unchanged.

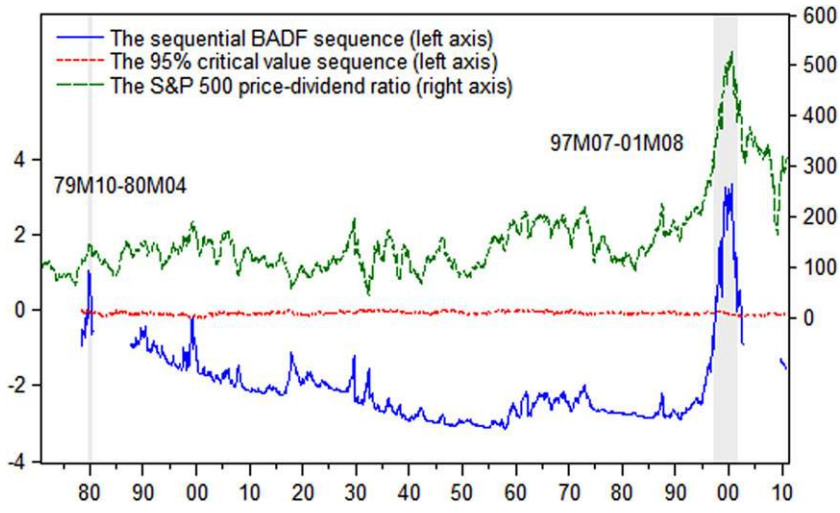

FIGURE 9

DATE-STAMPING BUBBLE PERIODS IN THE S&P 500 PRICE–DIVIDEND RATIO: THE SEQUENTIAL PWY STRATEGY

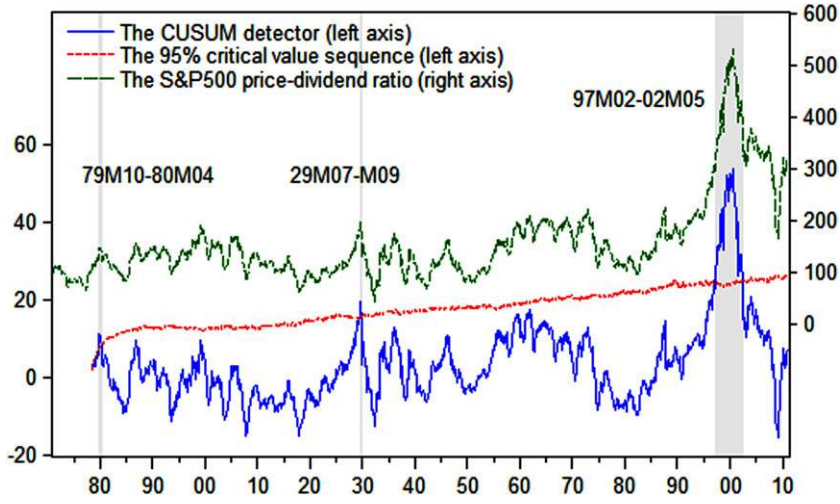

FIGURE 10

DATE-STAMPING BUBBLE PERIODS IN THE S&P 500 PRICE–DIVIDEND RATIO: THE CUSUM MONITORING PROCEDURE

time trend). To be consistent with the SADF and GSADF dating strategies, we chose a training sample of 90 months. Figure 10 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 evident in Figure 10, the CUSUM test identifies three bubble episodes: *the post long-depression period* (1879M10–1880M04), *the great crash episode* (1929M07–M09), and *the dot-com bubble* (1995M11–2001M08). Although the estimated origination and termination dates of the first and third episodes are exactly the same as those from the other three procedures, the great crash episode identified by the CUSUM procedure has a much shorter duration than that identified by the PSY strategy. In addition, the CUSUM procedure gives no warning to any other

events that are identified by the PSY dating strategy, including the 1987 Black Monday episode. So CUSUM monitoring may be regarded as a relatively conservative surveillance device.[30]

6. CONCLUSION

This article 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 in a double recursion 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 episodes of exuberance and collapse 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 shown to be 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 quite different behaviors 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 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 financial bubbles over this long period, whereas all other procedures seem more conservative and locate fewer episodes of exuberance.

APPENDIX: LIMIT THEORY FOR 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* $u_t = \psi(L)\varepsilon_t = \Sigma_{j=0}^{\infty}\psi_j\varepsilon_{t-j}$, *where* $\Sigma_{j=0}^{\infty} j\left|\psi_j\right| < \infty$ *and* $\{\varepsilon_t\}$ *is an i.i.d. sequence with mean zero, variance* σ^2, *and* $\mathbb{E}\left|\varepsilon_t\right|^{4+\delta} < \infty$ *for some* $\delta > 0$.

LEMMA A.1. *Suppose* u_t *satisfies error condition EC. Define* $M_T(r) = 1/T\sum_{s=1}^{[Tr]} u_s$ *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:*

[30] The conservative nature of the test arises from the fact the residual variance estimate $\hat{\sigma}_r$ (based on the data $\{y_1, ..., y_{\lfloor Tr\rfloor}\}$) 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.

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

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

(5) $T^{-3/2}\sum_{t=\lfloor Tr_1\rfloor}^{\lfloor Tr_2\rfloor}\varepsilon_t t \xrightarrow{L} \sigma[r_2W(r_2) - r_1W(r_1) - \int_{r_1}^{r_2}W(s)ds]$.

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

(7) $T^{-1/2}(\eta_{[Tr]} - \eta_0) \xrightarrow{p} 0$.

(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) $T^{-5/2}\sum_{t=\lfloor Tr_1\rfloor}^{\lfloor Tr_2\rfloor}\xi_{t-1}t \xrightarrow{L} \psi(1)\sigma\int_{r_1}^{r_2}W(s)sds$.

(11) $T^{-2}\sum_{t=\lfloor Tr_1\rfloor}^{\lfloor Tr_2\rfloor}\xi_{t-1}^2 \xrightarrow{L} \sigma^2\psi(1)^2\int_{r_1}^{r_2}W(s)^2ds$.

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

$$(a)\ T^{-1}\sum_{t=\lfloor Tr_1\rfloor}^{\lfloor Tr_2\rfloor} y_{t-1}\varepsilon_t \xrightarrow{L} \tfrac{1}{2}\sigma^2\psi(1)\left[W(r_2)^2 - W(r_1)^2 - r_w\right].$$

$$(b)\ T^{-3/2}\sum_{t=\lfloor Tr_1\rfloor}^{\lfloor Tr_2\rfloor} y_{t-1} \xrightarrow{L} \psi(1)\sigma\int_{r_1}^{r_2}W(s)\,ds.$$

$$(c)\ T^{-2}\sum_{t=\lfloor Tr_1\rfloor}^{\lfloor Tr_2\rfloor} y_{t-1}^2 \xrightarrow{L} \sigma^2\psi(1)^2\int_{r_1}^{r_2}W(s)^2\,ds.$$

$$(d)\ T^{-3/2}\sum_{t=\lfloor Tr_1\rfloor}^{\lfloor Tr_2\rfloor} y_{t-1}u_{t-j} \xrightarrow{p} 0, j = 0, 1, \cdots.$$

PROOF OF THEOREM 1. The fitted regression model (4) is

$$\Delta y_t = \sum_{j=1}^{k}\hat{\phi}_{r_1,r_2}^{j}\Delta y_{t-j} + \hat{\alpha}_{r_1,r_2} + \hat{\beta}_{r_1,r_2}y_{t-1} + \hat{\varepsilon}_t = X_t'\hat{\theta}_{r_1,r_2} + \hat{\varepsilon}_t. \tag{A.1}$$

Under the null hypothesis $\triangle y_t = d_T + u_t$ where $d_T = dT^{-\eta}$ and $u_t = \varepsilon_t \sim iid(0, \sigma^2)$, the error in the least squares estimator $\hat{\theta}_{r_1,r_2}$ from the true value $\theta_{r_1,r_2} := (\phi'_{r_1,r_2}, \alpha_{r_1,r_2}, \beta_{r_1,r_2})' = (0'_k, d_T, 0)'$ is given by

$$\hat{\theta}_{r_1,r_2} - \theta_{r_1,r_2} = \left[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t'\right]^{-1} \left[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t \varepsilon_t\right], \tag{A.2}$$

where $X_t = [d_T + u_{t-1}, d_T + u_{t-2}, \ldots d_T + u_{t-p+1}, 1, y_{t-1}]'$. Upon appropriate standardization and using (d) of Lemma A.2,[31] it is clear that the limit of the signal matrix $\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t'$ is block diagonal between the leading $k \times k$ block and the lower diagonal 2×2 block. Hence, the asymptotic behavior of the ADF statistic involving $\hat{\beta}_{r_1,r_2}$ depends only on the lower diagonal 2×2 block of $\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t'$ and the final 2×1 block of $\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t \varepsilon_t$. These components are

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

where Σ' denotes summation over $t = \lfloor Tr_1 \rfloor, \lfloor Tr_1 \rfloor + 1, \cdots, \lfloor Tr_2 \rfloor$. Based on (3) in Lemma A.1 and (a) in Lemma A.2, the appropriate scaling matrix is $\Upsilon_T = diag(\sqrt{T}, T)$. Premultiplying (A.2) by Υ_T leads to

$$\Upsilon_T \begin{bmatrix} \hat{\alpha}_{r_1,r_2} - \alpha_{r_1,r_2} \\ \hat{\beta}_{r_1,r_2} - \beta_{r_1,r_2} \end{bmatrix} \sim \left\{ \Upsilon_T^{-1} \left[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t'\right]_{(2)\times(2)} \Upsilon_T^{-1} \right\}^{-1} \left\{ \Upsilon_T^{-1} \left[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t \varepsilon_t\right]_{(2)\times 1} \right\}, \tag{A.3}$$

where the notation $(2) \times (2)$ and $(2) \times 1$ means the lower 2×2 block and lower 2×1 block, respectively. The matrix $\Upsilon_T^{-1}[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t X_t']_{(2)\times(2)} \Upsilon_T^{-1}$ has partitioned form

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

and the vector $\Upsilon_T^{-1}[\sum_{t=\lfloor Tr_1 \rfloor}^{\lfloor Tr_2 \rfloor} X_t \varepsilon_t]_{(-2)\times 1}$ has components

$$\begin{bmatrix} T^{-1/2}\Sigma' \varepsilon_t \\ T^{-1}\Sigma' y_{t-1}\varepsilon_t \end{bmatrix} \xrightarrow{L} \begin{bmatrix} \sigma [W(r_2) - W(r_1)] \\ \frac{1}{2}\sigma^2 \left[W(r_2)^2 - W(r_1)^2 - r_w\right] \end{bmatrix}.$$

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

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

where

$$A_{r_1,r_2} = \sigma \int_{r_1}^{r_2} W(s)\, ds,\ B_{r_1,r_2} = \sigma^2 \int_{r_1}^{r_2} W(s)^2\, ds,$$
$$C_{r_1,r_2} = \sigma [W(r_2) - W(r_1)],\ D_{r_1,r_2} = \frac{1}{2}\sigma^2 \left[W(r_2)^2 - W(r_1)^2 - r_w\right].$$

[31] Notice that $\psi(1)$ in Lemma A.2 equals unity under the null hypothesis of (3).

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

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

The t-statistic $t_{r_1,r_2} = \frac{\hat{\beta}_{r_1,r_2}}{s_{\hat{\beta}_{r_1,r_2}}}$ of $\hat{\beta}_{r_1,r_2}$ uses the standard error $s_{\hat{\beta}_{r_1,r_2}}$ defined by

$$s^2_{\hat{\beta}_{r_1,r_2}} = \hat{\sigma}^2_{r_1,r_2}\left(\begin{bmatrix} \Sigma' 1 & \Sigma' y_{t-1} \\ \Sigma' y_{t-1} & \Sigma' y_{t-1}^2 \end{bmatrix}^{-1}\right)_{22} = \frac{\hat{\sigma}^2_{r_1,r_2}}{\Sigma' y_{t-1}^2 - (\Sigma' y_{t-1})^2 / \Sigma' 1}, \tag{A.4}$$

where $\hat{\sigma}^2_{r_1,r_2} = (1/\lfloor Tr_w \rfloor)\Sigma' \hat{\varepsilon}_t^2$. Under the null model where $\alpha_{r_1,r_2} = dT^{-\eta}$, $\beta_{r_1,r_2} = 0$, and $\left\{\phi^j_{r_1,r_2} = 0, \text{ for } j = 1, ..., k\right\}$, we have $\hat{\sigma}^2_{r_1,r_2} \xrightarrow{p} \sigma^2$ as $T \to \infty$, so that

$$T^2 s^2_{\hat{\beta}_{r_1,r_2}} \xrightarrow{L} \frac{\sigma^2}{B_{r_1,r_2} - A^2_{r_1,r_2}/r_w} = \frac{1}{\int_{r_1}^{r_2} W(s)^2\, ds - \left(\int_{r_1}^{r_2} W(s)\, ds\right)^2 / r_w}.$$

Then the t-statistic $t_{r_1,r_2} = \frac{\hat{\beta}_{r_1,r_2}}{s_{\hat{\beta}_{r_1,r_2}}}$ satisfies

$$\begin{aligned} t_{r_1,r_2} &= \frac{T\hat{\beta}_{r_1,r_2}}{\left(T^2 s^2_{\hat{\beta}_{r_1,r_2}}\right)^{1/2}} \xrightarrow{L} \left(\frac{A_{r_1,r_2}C_{r_1,r_2} - r_w D_{r_1,r_2}}{A^2_{r_1,r_2} - r_w B_{r_1,r_2}}\right)\left(\frac{B_{r_1,r_2} - A^2_{r_1,r_2}/r_w}{\sigma^2}\right)^{1/2} \\ &= \frac{r_w D_{r_1,r_2} - A_{r_1,r_2}C_{r_1,r_2}}{\left(r_w B_{r_1,r_2} - A^2_{r_1,r_2}\right)^{1/2} \sigma r_w^{1/2}} \\ &= \frac{\frac{1}{2} r_w \left\{W(r_2)^2 - W(r_1)^2 - r_w\right\} - \left(\int_{r_1}^{r_2} W(s)\, ds\right)\left\{W(r_2) - W(r_1)\right\}}{r_w^{1/2}\left\{r_w \int_{r_1}^{r_2} W(s)^2\, ds - \left[\int_{r_1}^{r_2} W(s)\, ds\right]^2\right\}^{1/2}}. \end{aligned} \tag{A.5}$$

The asymptotic distribution of the GSADF statistic is obtained by means of the continuous mapping theorem giving

$$\sup_{\substack{r_2 \in [r_0,1] \\ r_1 \in [0, r_2 - r_0]}} \left\{ \frac{\frac{1}{2} r_w \left[W(r_2)^2 - W(r_1)^2 - r_w\right] - \int_{r_1}^{r_2} W(s)\, ds \left[W(r_2) - W(r_1)\right]}{r_w^{1/2}\left\{r_w \int_{r_1}^{r_2} W(s)^2\, ds - \left[\int_{r_1}^{r_2} W(s)\, ds\right]^2\right\}^{1/2}} \right\}, \tag{A.6}$$

and, thus, the stated result. However, derivation of (A.6) is not an immediate application of the continuous mapping theorem, which would require tightness of the random function sequence $\left\{t_{r_1,r_2}\right\}$ as well as the finite-dimensional limit theory given above in the one-dimensional case (A.5). Instead, as in the proof of theorem 1 of Zivot and Andrews (1992), a rigorous proof of (A.6) is more easily accomplished by the application of the continuous mapping theorem to a functional of the sample partial sum process $X_T^0(r) = \sqrt{T} M_T(r) = T^{-1/2} \sum_{s=1}^{\lfloor Tr \rfloor} u_s = T^{-1/2} y^0_{\lfloor Tr \rfloor}$ and the error variance estimate $\hat{\sigma}^2_{r_1,r_2}$.

We proceed to construct this functional and derive the limit result (A.6). First, note that under the null we have $y_t = \sum_{s=1}^t u_s + O_p(d\frac{t}{T^\eta}) =: y_t^0 + O_p(T^{1-\eta})$ so that $X_T(r) := T^{-1/2} y_{\lfloor Tr \rfloor}$

$= T^{-1/2} y^0_{\lfloor Tr \rfloor} + O_p(T^{1/2-\eta}) = X^0_T(r) + o_p(1)$ uniformly in r. Since $\psi(1) = 1$ under the null and $X^0_T(r) \Rightarrow \sigma W(r)$ by Lemma A.1, we have

$$T^{-2}\left\{\Sigma' y_{t-1}^2 - \left(\Sigma' y_{t-1}\right)^2 / \Sigma' 1\right\} = \int_{r_1}^{r_2} X_T^0(r)^2\, dr - \left(\int_{r_1}^{r_2} X_T^0(r)\, dr\right)^2 / r_w + o_p(1) \tag{A.7}$$

$$\Rightarrow \sigma^2 \left\{\int_{r_1}^{r_2} W(r)^2\, dr - \left(\int_{r_1}^{r_2} W(r)\, dr\right)^2 / r_w\right\}. \tag{A.8}$$

Define the two functionals $h_{1r}(X^0_T) := \int_0^r X^0_T(s)ds$ and $h_{2r}(X^0_T) := \int_0^r X^0_T(s)^2 ds$ of $X^0_T(r) \in D[0,1]$, the Skorohod space equipped with the uniform topology. Both h_{1r} and h_{2r} are continuous functionals by standard arguments. It is convenient to write

$$\int_{r_1}^{r_2} X_T^0(r)\, dr = h_{1r_2}\left(X_T^0\right) - h_{1r_1}\left(X_T^0\right), \quad \int_{r_1}^{r_2} X_T^0(r)^2\, dr = h_{2r_2}\left(X_T^0\right) - h_{2r_1}\left(X_T^0\right),$$

and then (A.7) and (A.8) can be written in functional form as

$$\begin{aligned} & T^{-2}\left\{\Sigma' y_{t-1}^2 - \left(\Sigma' y_{t-1}\right)^2 / \Sigma' 1\right\} \\ &= \left\{h_{2r_2}\left(X_T^0\right) - h_{2r_1}\left(X_T^0\right)\right\} - \left\{h_{1r_2}\left(X_T^0\right) - h_{1r_1}\left(X_T^0\right)\right\}^2 / r_w + o_p(1) \\ &\Rightarrow \sigma^2 \left\{h_{2r_2}(W) - h_{2r_1}(W)\right\} - \left\{h_{1r_2}(W) - h_{1r_1}(W)\right\}^2 / r_w. \end{aligned} \tag{A.9}$$

Since h_{1r} and h_{2r} are continuous, so is the functional

$$g_{1,r_1,r_2}\left(X_t^0\right) := \left\{h_{2r_2}\left(X_T^0\right) - h_{2r_1}\left(X_T^0\right)\right\} - \left\{h_{1r_2}\left(X_T^0\right) - h_{1r_1}\left(X_T^0\right)\right\}^2 / r_w.$$

Now observe that

$$\begin{aligned} g_{1,r_1,r_2}(W) &= \int_{r_1}^{r_2} W(r)^2\, dr - \left(\int_{r_1}^{r_2} W(r)\, dr\right)^2 / r_w \\ &= \int_{r_1}^{r_2} W(r)^2\, dr - r_w \left(r_w^{-1} \int_{r_1}^{r_2} W(r)\, dr\right)^2 \\ &=: \int_{r_1}^{r_2} \underline{W}_{r_1,r_2}(r)^2\, dr, \end{aligned}$$

with $\underline{W}_{r_1,r_2}(r) = W(r) - r_w^{-1} \int_{r_1}^{r_2} W(r)dr$. Just as in Phillips and Hansen (1990, Lemma A2), we have that $\int_{r_1}^{r_2} \underline{W}_{r_1,r_2}(r)^2 dr > 0$ *a.s.* since $r_w = r_2 - r_1 \geq r_0 > 0$. It follows that the functional

$$g^1_{r_1,r_2}(W) := 1/g_{1,r_1,r_2}(W) \tag{A.10}$$

is well defined for all (r_1, r_2) such that $r_w = r_2 - r_1 \geq r_0 > 0$ and so the functional

$$g^1_{r_1,r_2}\left(X_t^0\right) := \frac{1}{g_{1,r_1,r_2}\left(X_t^0\right)}$$

is continuous with limit $g^1_{r_1,r_2}(X^0_t) \Rightarrow g^1_{r_1,r_2}(\sigma W) = \sigma^2 g^1_{r_1,r_2}(W)$.

Next, from (A.3), the numerator component $(T\hat{\beta}_{r_1,r_2})$ of the t ratio is given by

$$\begin{aligned}
&\frac{\left[-T^{-3/2}\Sigma' y_{t-1}\ T^{-1}\Sigma' 1\right]\begin{bmatrix} T^{-1/2}\Sigma'\varepsilon_t \\ T^{-1}\Sigma' y_{t-1}\varepsilon_t \end{bmatrix}}{\left(T^{-2}\Sigma' y_{t-1}^2\right)\left(T^{-1}\Sigma' 1\right)-\left(T^{-3/2}\Sigma' y_{t-1}\right)^2} \\
&= \frac{\left[-\int_{r_1}^{r_2} X_T^0(r)\,dr\ r_w\right]\begin{bmatrix} X_T^0(r_2)-X_T^0(r_1) \\ \int_{r_1}^{r_2} X_T^0(r)\,dX_T^0(r) \end{bmatrix}}{\left(\int_{r_1}^{r_2} X_T^0(r)^2\,dr\right) r_w-\left(\int_{r_1}^{r_2} X_T^0(r)\,dr\right)^2}+o_p(1) \\
&= \frac{\left[-\int_{r_1}^{r_2} X_T^0(r)\,dr\ r_w\right]\begin{bmatrix} X_T^0(r_2)-X_T^0(r_1) \\ \frac{1}{2}\left[X_T^0(r_2)^2-X_T^0(r_1)^2-T^{-1}\Sigma'\varepsilon_t^2\right] \end{bmatrix}}{\left(\int_{r_1}^{r_2} X_T^0(r)^2\,dr\right) r_w-\left(\int_{r_1}^{r_2} X_T^0(r)\,dr\right)^2}+o_p(1) \\
&= \frac{\left[-\left[h_{1r_2}\left(X_T^0\right)-h_{1r_1}\left(X_T^0\right)\right]\ r_w\right]\begin{bmatrix} X_T^0(r_2)-X_T^0(r_1) \\ \frac{1}{2}\left[X_T^0(r_2)^2-X_T^0(r_1)^2-T^{-1}\Sigma'\varepsilon_t^2\right] \end{bmatrix}}{r_w g_{1,r_1,r_2}\left(X_t^0\right)}+o_p(1) \\
&= \frac{\frac{r_w}{2}\left[X_T^0(r_2)^2-X_T^0(r_1)^2-\frac{\lfloor Tr_w\rfloor}{T}\hat{\sigma}_{r_1,r_2}^2\right]-\left[h_{1r_2}\left(X_T^0\right)-h_{1r_1}\left(X_T^0\right)\right]\left[X_T^0(r_2)-X_T^0(r_1)\right]}{r_w g_{1,r_1,r_2}\left(X_t^0\right)}+o_p(1) \qquad \text{(A.11)} \\
&\Rightarrow \frac{\frac{r_w}{2}\left[W(r_2)^2-W(r_1)^2-r_w\right]-\left[h_{1r_2}(W)-h_{1r_1}(W)\right]\left[W(r_2)-W(r_1)\right]}{r_w g_{1,r_1,r_2}(W)}.
\end{aligned}$$

Define

$$\begin{aligned}
&g_{2,r_1,r_2}\left(X_T^0,\hat{\sigma}_{r_1,r_2}^2\right) \\
&= \frac{r_w}{2}\left[X_T^0(r_2)^2-X_T^0(r_1)^2-r_w\hat{\sigma}_{r_1,r_2}^2\right]-\left[h_{1r_2}\left(X_T^0\right)-h_{1r_1}\left(X_T^0\right)\right]\left[X_T^0(r_2)-X_T^0(r_1)\right] \qquad \text{(A.12)} \\
&\Rightarrow g_{2,r_1,r_2}\left(\sigma W,\sigma^2\right)
\end{aligned}$$

$$= \sigma^2\frac{r_w}{2}\left[W(r_2)^2-W(r_1)^2-r_w\right]-\sigma^2\left[h_{1r_2}(W)-h_{1r_1}(W)\right]\left[W(r_2)-W(r_1)\right] \qquad \text{(A.13)}$$

so that (A.11) is

$$\frac{g_{2,r_1,r_2}\left(X_T^0,\hat{\sigma}_{r_1,r_2}^2\right)}{r_w g_{1,r_1,r_2}\left(X_t^0\right)}+o_p(1)\Rightarrow\frac{g_{2,r_1,r_2}\left(\sigma W,\sigma^2\right)}{r_w g_{1,r_1,r_2}(\sigma W)}.$$

Using (A.4) and (A.9), we have

$$\begin{aligned}
T^2 s_{\hat{\beta}_{r_1,r_2}}^2 &= \frac{\hat{\sigma}_{r_1,r_2}^2}{T^{-2}\Sigma' y_{t-1}^2-\left(T^{-1}\Sigma' y_{t-1}\right)^2/\Sigma' 1} \\
&= \frac{\hat{\sigma}_{r_1,r_2}^2}{g_{1,r_1,r_2}\left(X_t^0\right)}+o_p(1). \qquad \text{(A.14)}
\end{aligned}$$

It follows from (A.11) and (A.14) that the t ratio can be written as

$$t_{r_1,r_2} = \frac{T\hat{\beta}_{r_1,r_2}}{\left(T^2 s^2_{\hat{\beta}_{r_1,r_2}}\right)^{1/2}} = \frac{g_{2,r_1,r_2}\left(X_T^0, \hat{\sigma}^2_{r_1,r_2}\right)}{r_w g_{1,r_1,r_2}\left(X_t^0\right)}\left(\frac{g_{1,r_1,r_2}\left(X_t^0\right)}{\hat{\sigma}^2_{r_1,r_2}}\right)^{1/2} + o_p(1)$$

$$= \frac{g_{2,r_1,r_2}\left(X_T^0, \hat{\sigma}^2_{r_1,r_2}\right)}{r_w\left(g_{1,r_1,r_2}\left(X_T^0\right)\right)^{1/2}\hat{\sigma}_{r_1,r_2}} + o_p(1) =: g_{r_1,r_2}\left(X_T^0, \hat{\sigma}^2_{r_1,r_2}\right) + o_p(1),$$

which defines the required functional $g_{r_1,r_2}(\cdot,\cdot)$ representing the t ratio in terms of $(X_T^0, \hat{\sigma}^2_{r_1,r_2})$. Since $\hat{\sigma}^2_{r_1,r_2} \to_p \sigma^2$, we have

$$t_{r_1,r_2} = g_{r_1,r_2}\left(X_T^0, \hat{\sigma}^2_{r_1,r_2}\right) \Rightarrow g_{r_1,r_2}\left(\sigma W, \sigma^2\right) = \frac{g_{2,r_1,r_2}(\sigma W, \sigma)}{r_w\left(g_{1,r_1,r_2}(\sigma W)\right)^{1/2}\sigma} = \frac{g_{2,r_1,r_2}(W,1)}{r_w g_{1,r_1,r_2}(W)^{1/2}}$$

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

$$= g_{r_1,r_2}(W,1).$$

In view of the continuity of $g_{2,r_1,r_2}(X_T^0, \hat{\sigma}^2_{r_1,r_2})$ and $1/g_{1,r_1,r_2}(X_T^0,)$, the functional $g_{r_1,r_2}(X_T^0, \hat{\sigma}^2_{r_1,r_2}) = g_{1,r_1,r_2}(X_T^0, \hat{\sigma}^2_{r_1,r_2})/r_w(g_{2,r_1,r_2}(X_T^0))^{1/2}\hat{\sigma}_{r_1,r_2}$ is continuous for all (r_1, r_2) such that $r_w = r_2 - r_1 \geq r_0 > 0$.

The continuous functional $g_{r_1,r_2}(\cdot,\cdot)$ maps $D[0,1]\times\mathbb{R}^+$ onto a function defined on $\Lambda_0 = \left\{(r_1,r_2) : 1 \geq r_2 \geq r_1 + r_0 \text{ and } 1 - r_0 \geq r_1 \geq 0\right\}$. Define the double sup functional $g^*(g_{r_1,r_2}) = \sup_{(r_1,r_2)\in\Lambda_0} g_{r_1,r_2}$, which maps functions defined on Λ_0 onto $\mathbb{R}$. Let g_{r_1,r_2} and $\breve{g}_{r_1,r_2}$ be two functions defined on Λ_0 such that $\sup_{(r_1,r_2)\in\Lambda_0}\left|g_{r_1,r_2} - \breve{g}_{r_1,r_2}\right| < \varepsilon$ for some given $\varepsilon > 0$. The function $g^*(g_{r_1,r_2})$ is continuous with respect to the uniform norm on its domain because

$$\left|g^*(g_{r_1,r_2}) - g^*(\breve{g}_{r_1,r_2})\right| = \left|\sup_{(r_1,r_2)\in\Lambda_0}\left[g_{r_1,r_2} - \breve{g}_{r_1,r_2}\right]\right| \leq \sup_{(r_1,r_2)\in\Lambda_0}\left|g_{r_1,r_2} - \breve{g}_{r_1,r_2}\right| < \varepsilon.$$

We therefore deduce by continuous mapping the weak convergence

$$\sup_{(r_1,r_2)\in\Lambda_0} t_{r_1,r_2} = \sup_{(r_1,r_2)\in\Lambda_0} g^*\left(g_{r_1,r_2}\left(\left(X_T^0, \hat{\sigma}^2_{r_1,r_2}\right)\right)\right) \Rightarrow \sup_{(r_1,r_2)\in\Lambda_0} g^*\left(g_{r_1,r_2}((W,1))\right)$$

$$= \sup_{\substack{r_2\in[r_0,1]\\ r_1\in[0,r_2-r_0]}} g_{r_1,r_2}(W,1),$$

$$(W,1),$$

giving (A.6) as required.

REFERENCES

Abreu, D., and M. K. Brunnermeier, "Bubbles and Crashes," *Econometrica* 71 (2003), 173–204.

Ahamed, L., *Lords of Finance: The Bankers Who Broke the World* (New York: Penguin Press, 2009).

Ang, A., and G. Bekaert, "Stock Return Predictability: Is It There?" *Review of Financial Studies* 20 (2007), 651–707.

AVERY, C. AND P. ZEMSKY, "Multidimensional Uncertainty and Herd Behavior in Financial Markets," *American Economic Review* 88 (1998), 724–48.
BERK, K. N., "Consistent Autoregressive Spectral Estimates," *Annals of Statistics* 2 (1974), 489–502.
BHARGAVA, A., "On the Theory of Testing for Unit Roots in Observed Time Series," *Review of Economic Studies* 53 (1986), 369–84.
BLANCHARD, O. J., "Speculative Bubbles, Crashes and Rational Expectations," *Economics Letters* 3 (1979), 387–89.
BUSETTI, F., AND A. M. R. TAYLOR, "Tests of Stationarity Against a Change in Persistence," *Journal of Econometrics* 123 (2004), 33–66.
CASPI, I., "Right-Tailed ADF Tests: Eviews Add-In," Mimeo, Bank of Israel, 2013.
CHU, C. J., M. STINCHCOMBE, AND H. WHITE, "Monitoring Structural Change," *Econometrica* 64 (1996), 1045–65.
COCHRANE, J. H., "Explaining the Variance of price–dividend Ratios," *The Review of Financial Studies* 5 (1992), 243–80.
———, *Asset Pricing* (Princeton: Princeton University Press, 2005).
COOPER, G., *The Origin of Financial Crises: Central Banks, Credit Bubbles and the Efficient Market Fallacy* (New York: Vintage Books, 2008).
DIBA, B. T., AND H. I. GROSSMAN, "Explosive Rational Bubbles in Stock Prices?" *The American Economic Review* 78 (1988), 520–30.
EVANS, G. W., "Pitfalls in Testing for Explosive Bubbles in Asset Prices," *The American Economic Review* 81 (1991), 922–30.
FERGUSON, N., *The Ascent of Money* (New York: Penguin Press, 2008).
FROOT, K. A., AND M. OBSTFELD, "Intrinsic Bubbles: The Case of Stock Prices," *American Economic Review* 81 (1991), 1189–214.
FUNKE, M., S. HALL, AND M. SOLA, "Rational Bubbles during Poland's Hyperinflation: Implications and Empirical Evidence," *European Economic Review* 38 (1994), 1257–76.
GURKAYNAK, R. S., "Econometric Tests of Asset Price Bubbles: Taking Stock," *Journal of Economic Surveys* 22 (2008), 166–86.
HALL, S. G., Z. PSARADAKIS, AND M. SOLA, "Detecting Periodically Collapsing Bubbles: A Markov-Switching Unit Root Test," *Journal of Applied Econometrics* 14 (1999), 143–54.
HARVEY, D. I., S. J. LEYBOURNE, R. SOLLIS, AND R. TAYLOR, "Testings for Explosive Financial Bubbles in the Presence of Non-stationary Volatility," Working Paper, 2014, available from http://www.nottingham.ac.uk/lezdih/bubble_bootstrap.pdf.
HOMM, U. AND J. BREITUNG, "Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods," *Journal of Financial Econometrics* 10 (2012), 198–231.
KIM, J. Y., "Detection of Change in Persistence of a Linear Time Series," *Journal of Econometrics* 95 (2000), 97–116.
LEE, J. AND P. C. B. PHILLIPS, "Asset Pricing with Financial Bubble Risk," Working Paper, Yale University, 2011.
LEYBOURNE, S., T. -H. KIM, AND A. M. R. TAYLOR, "Detecting Multiple Changes in Persistence," *Studies in Nonlinear Dynamics and Econometrics* 11 (2007), 1–32.
PÁSTOR, L., AND P. VERONESI, "Was There a NASDAQ Bubble in the Late 1990s?" *Journal of Financial Economics* 81 (2006), 61–100.
PHILLIPS, P. C. B, AND B. E. HANSEN, "Statistical Inference in Instrumental Variables Regression with I (1) Processes," *The Review of Economic Studies* 57 (1990), 99–125.
———, AND T. MAGDALINOS, "Limit Theory for Moderate Deviations from a Unit Root," *Journal of Econometrics* 136 (2007), 115–30.
———, AND P. PERRON, "Testing for a Unit Root in Time Series Regression," *Biometrika* 75 (1988), 335–46.
———, AND S. SHI, "Financial Bubble Implosion," Working Paper, Presented at NZESG Conference, February, 2014.
———, AND V. SOLO, "Asymptotics for Linear Processes," *Annals of Statistics* 20 (1992), 971–1001.
———, AND J. YU, "Limit Theory for Dating the Origination and Collapse of Mildly Explosive Periods in Time Series Data," Unpublished manuscript, Sim Kee Boon Institute for Financial Economics, Singapore Management University, 2009.
———, AND ———, "Dating the Timeline of Financial Bubbles During the Subprime Crisis," *Quantitative Economics* 2 (2011), 455-491.
———, S. SHI, AND J. YU "Specification Sensitivity in Right-Tailed Unit Root Testing for Explosive Behaviour," *Oxford Bulletin of Economics and Statistics* 76 (2014), 315–33.
———, ———, AND ———, "Testing for Multiple Bubbles: Limit Theory of Dating Algorithms," *International Economic Review*, 56 (2015a), 1043–1098.
———, ———, AND ———, "Supplement to Two Papers on Multiple Bubbles," Manuscript, 2015b, available from https://sites.google.com/site/shupingshi/TN_GSADFtest.pdf?attredirects=0&d=1.

———, Y. Wu, and J. Yu, "Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Asset Values?" *International Economic Review* 52 (2011), 201–26.

Said, S. E. and D. A. Dickey, "Testing for Unit Roots in ARMA Models of Unknown Order," *Biometrika* 71 (1984), 599–607.

Shi, S., "Econometric Tests for Nonlinear Exuberance in Economics and Finance," PhD Thesis, The Australian National University, 2011.

———, "Specification Sensitivities in the Markov-Switching Unit Root Test for Bubbles," *Empirical Economics* 45 (2013), 697–713.

———, and Y. Song, "Identifying Speculative Bubbles with an Infinite Hidden Markov Model," *Journal of Financial Econometrics* (2014). Advance online publication. doi: 10.1093/jjfinec/nbu025

van Norden, S. and R. Vigfusson, "Avoiding the Pitfalls: Can Regime-Switching Tests Reliably Detect Bubbles?" *Studies in Nonlinear Dynamics & Econometrics* 3 (1998), 1–22.

Xiao, Z. and P. C. B. Phillips, "An ADF Coefficient Test for a Unit Root in ARMA Models of Unknown Order with Empirical Applications to the U.S. Economy," *The Econometrics Journal* 1 (1998), 27–43.

Zivot, E. and D. W. K. Andrews,"Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business and Economic Statistics* 20 (2002), 25–44.