Testing Multiple Financial Bubbles using the Recursive Flexible Window Methodology

Swarna D. Dutt
University of West Georgia

Dipak Ghosh
Emporia State University

Abstract

Identifying and dating financial bubbles in real time is in the forefront of current empirical research. Their accuracy provides very useful "warning alerts" to central bankers and fiscal regulators with real time data. But the complexity of their nonlinear structure and the inherent sudden break mechanisms makes the econometric testing challenging, to say the least. The new recursive flexible window methodology provided by Phillips, Shi and Yu (2015) gives consistent results and delivers significant power gains when multiple bubbles occur. It successfully identifies the well-known historical episodes of exuberance and collapse. As a first step here, we extend their data set of the S&P 500 stock market from (January 1871 to December 2016. The existence of (more) bubbles in some and not all these data sets might give us some indication of where, in an economy, bubbles are more likely to occur. We believe that this information will be of interest to researchers in this field.

1. Introduction

History is replete with incidents of financial crisis, which ex-post become a wakeup call for policymakers and the people. Again and again it was stated by experts that the present crisis was preceded by "asset market bubbles" and / or "excessive credit expansion." But the fact of the matter remains that we do not have good quantitative markers which can ex-ante indicate the genesis of a momentum being built in the asset / credit markets which may lead to a catastrophe down the line. If we had quantitative observable "warning signs" many an economic debacle can be avoided. Thus after the most recent global financial crisis of 2007-2009, the main thrust in the Basel III accord was to emphasize on more close and determinate market surveillance, so that bankers and policy makers could be forewarned of a possible impending implosion.

But as Cooper (2008) said "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 prick an asset price bubble, we are told it is impossible to see the bubble, while it is in its inflationary phase." Thus we have to accept that there is no practical way to identify the genesis of a crisis. Thus the task at hand is to try to identify possible quantitative markers from the data, that something is "awry" and that a speculative bubble is probably taking shape. It will worsen if measures to "quell" it is not taken, now. That is where the Phillips, Shi and Yu (PSY henceforth, 2014b) research comes into effect. This paper offers the first powerful and credible "quantitative metric" to detect exuberance in financial data, right where it is originating. Once detected, the counteractive policies can be promulgated and implemented.

Looking at long term S&P 500 data from 1871-2010 (about 140 years), the authors propose a recursive algorithm, which can actually diagnose and identify ex-ante the signs of "turbulence within the force" if you will. This procedure helps us pinpoint the start of the problem, and can thus help us monitor the markets. Since we know that history has proven that it has a bad habit of repeating itself, this early warning diagnostic tool will come in handy, in helping make / alter policies to avert the impending crisis. The best part of this test is that it can be implemented on current data in real time and try to detect the "fault lines."

In the economics literature we have multiple tests to detect ex-post the crisis, and then explain it. Gurkaynak (2008) is a good review of this documentation. But there was no test to ex-ante identify the origination of a bubble which is in the making. There were no econometric detectors of a future market crisis. Phillips, Wu and Yu (PWY henceforth, 2011) presented a recursive method to detect exuberance in asset prices during an inflationary phase. The advantage here being that the early detection (ex-ante acknowledgement) can help banks / regulators / policy makers to address the problem in its nascent state. PWY was very effective in the early detection of bubble markers, provided there was a single bubble / turbulence in the data sample. They proved the effectiveness of the test using NASDAQ (PWY, 2011) and the US housing bubble (PY, 2011). This was an incredible contribution to the economics research literature.

But then came the question of "economic reality" which showed that there usually were multiple recurring financial crises, over long periods. Ahamed (2009) gave us evidence of 60 different financial crises, in the 17th century alone. Thus the next step in the evolution of these detection tests was to create the one that could identify multiple bubbles in the same sample period. A test to clearly make periodic collapsing and recovering economic data was simply not there. This recursive identification is extremely complex compared to identifying a single bubble. The main problem is computationally handling the non-linear structure of multiple breaks / bubbles in the data. With the presence of multiple break points in the data, the discriminatory power of the detectors go down dramatically and hence the upswings and downswings are not decipherable in the same data stream. Thus the challenge is twofold:

1) Come up with a statistical metric which can detect multiple factual fractures in the non-linear data stream
2) Be powerful and effective enough so as not to have a low false negative detection tolerance (to avoid unnecessary policies) and also a high positive detection tolerance (so as to ensure good and early effective policy application.)

This paper presents a recursive econometric technique to detect / test / date financial bubbles in the same sample data, and separate them when multiple bubbles are present. Here the authors extend on their (PWY, 2011) methodology, which is based on a sequence of forward recursive right tailed ADF unit root tests, using the Sup ADF (designated SADF) measure. This process allows for a dating strategy to identify the origination and termination dates of a specific bubble. This is achieved by using "backward regression techniques." In case of a single bubble, the PWY test is consistent, as shown in PY 2009. This detection algorithm is better able to date the ups and downs of financial data, as opposed to the CHOW tests, CUSUM tests etc. as evidenced by Homm and Breitung (2012). Its added strength is that it can detect exuberance in the data arising from different sources, as would happen in real life.

But PWY is consistent in detecting single bubbles or exuberance points in the sample data. What if there are multiple bubbles, originating and decaying in sequence over time. PWY is not proven to be consistent in such cases. It cannot be confidently used in examining long term market data where exuberance and collapse are evident ex-post.

This paper presents an extension of the SADF tests, in form of a generalized SADF called the GSADF method. It includes a recursive backward regression technique, to time identify the origin and collapse of bubbles. It is a right tailed ADF test, but has a flexible window width to separate one bubble from the next, to the next sequentially, since their lengths are bound to be different. In structure and logic, it is analogous to the left-sided recursive unit root test of Leybourne, Kim and Taylor (2007), this being a right-sided double recursive unit root test.

It's an ex-ante procedure to detect different start and end points of bubbles in real time data, i.e., identify and separate multiple bubble episodes over the same sample set. This test has been proven to consistently give good results, when multiple bubbles are present. Thus it can credibly be applied to analyzing long term historical data. Along with the ex-ante dating algorithm and the GSADF test, the authors develop a modified PWY algorithm, which reinitializes the test sequentially, after the detection of each bubble. This sequential test works in deciphering multiple bubbles from explosion to collapse, and separate them over time.

It is applied to the S&P 500 stock market data from January 1871- December 2010. It has been able to identify all the historically documented bubble episodes, like the 1929 crash, 1954 boom, 1987 black Monday and the latest dot-com bubble. Section 2 describes the reduced form model, the new rolling window recursive test and its limit theory. Section 3 elaborates the data stamping strategies, to separate single, double and multiple bubbles in the same sample period. Section 4 is simulation results of the size, power and performance of the dating strategy tests. In section 5, they apply the PWY test, the sequential PWY test and the CUSUM test to the same S&P data. Section 6 concludes.

PSY (2015) develops the limit theories and consistency properties in case of single and multiple bubbles. PSY (2015, b) is a supplement describing the robustness checks of this testing procedure.

2. Rolling window test for bubbles

It originates with the standard asset pricing model:

$$P_t = \sum_{i=0}^{\infty} (\frac{1}{1+r_f}) i \, E_t(D_{t+i} + U_{t+i}) + B_t \quad (1)$$

where
P_t = after dividend price of an asset
D_t = payoff (dividend) from the asset
r_f = risk free interest rate
U_t = unobservable fundamentals
B_t = bubble component

Here $P_t^f = P_t - B_t$ (market fundamentals) and B_t satisfies the sub martingale property

$$E_t = (B_{t+1}) = (1 + r_f)B_t \quad (2)$$

This equation sets up the alternative scenarios for the presence / absence of bubbles in the data. For example: If there are no bubbles, the B_t =0, then the degree of non-stationarity [I(0) or I(1)] of asset prices is controlled by asset payoffs or dividends (D_t) and the unobservable economic / market fundamentals. The advantage of the reduced form model is that it pretty much encompasses all standard formulations as intrinsic bubbles (Froot and Obstfeld, 1991), herd behavior (Abreu and Brunnermeier, 2003), time varying discounting (Phillips and Yu, 2011)[1].

A possible outcome would be like this: If D_t is an I(1) process, the U_t has to be either I(0) or I(1) and asset prices can at the most be a I(1) process. But based on eq. (2), if there are bubbles, then asset prices will be explosive. Thus when the fundamentals are I(1) and D_t is first difference stationary, we can infer bubbles if asset prices show evidence of explosive behavior. Eq (1) is one way to include a bubble variable in the standard asset pricing model, but the jury is still out on this. [2, 3]

According to Phillips and Magdalinos (2007), explosive behavior in asset prices is a primary indicator of market exuberance, which can be identified in empirical tests using the "recursive testing procedure" like the right side unit root test of PWY. This recursive procedure starts with a martingale null (with drift to capture long historical trends in asset data.) The model specification is:

$$y_t = dT^{-n} + \theta_{yt-1} + \epsilon_t \quad (3)$$

where $€_t$ is iid $(0, б^2)$, $\Theta = 1$, and d is a constant, T is the sample size, and the parameter η controls the magnitude of the intercept and the drift, as $T \rightarrow \infty$. Solving eq. 3, gives us the deterministic trend, dt/T^n. Here there are three possibilities:

1) If n>0, the drift will be small compared to the linear trend.
2) If n>1/2, the drift is small relative to the martingale
3) If n=1/2, the output behaves like a Brownian motion, which is evident in many financial time series data.

The researcher needs to be careful and exercise caution because the emphasis here is on the alternative hypothesis, because departures from market fundamentals are the markers of interest. But as with all types of model specifications, we know that they are sensitive to intercepts, trends and trend breaks etc. as described in PSY (2014a). Eq. 3 is tested for exuberance using the rolling window ADF approach or the recursive approach of the authors. The basic logic is that if the rolling window regression starts from the r_1^{th} fraction and ends with the r_2^{th} fraction (from sample size T), then $r_2 = r_1 + r_w$, where r_w is the size of the window. This model is:

$$\Delta y_t = \propto_{r1,r2} + \beta_{r1,r2}\, y_{t-1} + \Sigma_{i=1}^{k} \gamma_{r1,r2}^{i}\, \Delta_{yt-1} + \epsilon_t \quad (4)$$

where k is the lag length, and ϵ_t is iid, with $(0, б^2_{r1,r2})$. The basic form is reformulated to

include the presence of "multiple bubbles" to separate the market switching time periods from explosion to contraction, and again explosion sequentially. They use the Sup ADF test called SADF. It is a recursive / repeated estimation procedure with window size r_w., where r_w goes from r_0 (smallest sample window fraction) to r_1 (largest sample window fraction), and sample end point $r_2 = r_w$, going from 0 to 1. The SADF statistic i: [4]
$SADF (r_0)$ n= $\sup_{r2\in[r0,1]} ADF^{r2}_{0}$

Rolling window GSADF test

The ADF regression is run on eq. 4, recursively, but continuously on sub-samples of the data based on window width chosen according to r_0, r_1, $r_2 \ldots\ldots r_w$. The subsamples chosen here are more extensive than the SADF test. The difference here is that we allow the window width to change within the feasible range where $r_w = r_2 - r_1$. The GSADF statistic is:
$GSADF (r_0) = \sup r_2 \in[r_0, 1] \{ADF^{r2}_{r1}\}$

$r_1\in[0, r_2\text{-}r_0]$ (5)

The GSADF statistic as given in eq. (5) [5]. Here we see that the limit distribution of the GSADF holds (is identical), but with the intercept and the assumption of a random walk structure, we have no drift or small drift. The GSADF's asymptotic distribution depends on the "smallest window width size r_0." Care needs to be exercised on choosing the width of r_0. It depends on the number of observations in the sample.

Case (1): If T is small, r_0 has to be made large enough to ensure the inclusion of an adequate number of observations.
Case (2): If T is large, r_0 should be set small, so as to be able to include different "explosive" burst in the data. The authors run simulations and derive the critical values (CV's). The conclusions are:

1) As r_0 decreases, CV's of the test statistic increases

2) For r_0 given, the CV's are constant in finite samples.

3) GSADF statistic CV's are larger than the SADF statistic, which is larger than the ADF statistic, and its concentration also increases, increasing confidence in the test outcomes. The backward SADF statistic is the sup value of the ADF sequence run over this interval.

$BSADF\ r_2 (r_0) = \sup_{r1\in[0,\ r2\text{-}r0]} \{ADF^{\ r2}_{r1}\}$

We see here that the ADF test is a special case of the backward sup ADF (BSADF) when r_1=0. The empirical steps are:

1) Determine $ADFr_2$ and the sup ADF within the feasible range of r_2 (from r_0 to r_1.) The origination of the bubble is dated. This procedure imposes the condition that the bubble marker is the existence of a critical value greater than L_T = Log (T). This separates the short and temporary market blips (which happen all the time in real life) from actual exuberance. Dating is

done using the formula:[6,7,8]

$$\hat{r_e} = inf_{r2\in[r0,1]}\{r_2 : ADF_{r2} > cv_{r2}^{\beta T} \} \ (6)$$

And

$$\hat{r_f} = inf_{r2\in[\hat{r_e}+\frac{\log(T)}{T,1}]}\left\{r_2 : ADF_{r2} < cv_{r2\}}^{\beta T}\right\} \ (7)$$

where $cv^{\boldsymbol{\beta}T}_{r2}$ is the $100(1-\boldsymbol{\beta}_T)$ % critical value of the ADF statistic based on $[T_{r2}]$ observations. Here $\boldsymbol{\beta}_T \to 0$, as $T \to \infty$.

3. Data stamping strategies

The idea is to identify bubbles in real time data and then look for the "markers" identifying those bubbles / episodes of market exuberance. The problem is that the standard ADF test can identify extreme observations, as $r = [T_r]$, but cannot separate between a bubble phase observation from one which is part of a natural growth trajectory. Market growth is not an indication of bubbles. Thus ADF tests may result in finding "pseudo bubble detection." Making this distinction is the major contribution of this test. The authors run backward sup ADF or backward SADF tests, to improve the chances of deciphering a bubble from a growth trajectory. The recursive test means running SADF backwards on the sample, increasing the sample sequence using a fixed sample r_2, but varying the initial point from 0 to (r_2-r_0). This gives the SADF statistic:

$\{ADF^{r2}_{r1}\} \in |0, r_2 -0|$

Bubbles are inferred from the backward SADF statistic or the BSADF $r_2(r_0)$. The origin of the bubbles, the date and timing is the first observation whose BSADF statistic exceeds the critical value of the BSADF. The bubble ending date / time frame is the first observation whose BSADF is below the BSADF critical value. The intermediary time frame is the duration of the bubble. The origination / termination dates are calculated thus:

$$\hat{r_e} = inf_{r2\in[r0,1]}\{r_2 : BSADF_{r2}(r_0) > scv_{r2}^{\beta T} \ (8)$$

$$\hat{r_f} = r_{2\in}[inf_{\hat{r_e}+\left[\frac{\partial \log(T)}{T,1}\right]}\{r_2 : BSADF_{r2}(r_0) > scv_{r2}^{\beta T} \ (9)$$

where $scv^{\boldsymbol{\beta}T}_{r2}$ is the $100(1-\boldsymbol{\beta}_T)$% critical value of the sup ADF statistic, based on $[T_{r2}]$ observations. $\boldsymbol{\beta}_T$ goes to zero, as the sample size approaches infinity. The distinction between the SADF and the GSADF (backward sup ADF) tests, both run over $r_2 \in [r_0,1]$ is given by the statistic as:

$SADF (r_0) = sup_{r2\in[r0,1]} \{ADF_{r2}\}$

And

$GSADF (r_0) = sup_{r2\in[r0,1]} \{BSADF_{r2}(r_0)\}$

4. Simulations

Simulations were performed to examine the credibility of the PWY, sequential PWY, CUSUM and the GSADF test, in terms of size and power, but most importantly their capability to identify multiple bubble episodes. The basic data generating process is given by:

$$y_t = dT^{-n} + \theta_{yt-1} + \epsilon_t \tag{10}$$

with δ =n =1. They examine two different models, namely Evans (1991) collapsing bubble and the PWY model. Simulations using the same data set, same number of observations / replications show that the size distortion of SADF > GSADF. The next question is the effect of the lag selection length. Both SADF and the GSADF have size distortion weakness. But its magnitude is small, when we use a fixed lag length in recursive tests. But GSADF has smaller distortion than SADF and thus has a leg up on the latter in lowering the probability of "false detection." The authors recommend the fixed lag length use with the GSADF test for multiple bubbles. They find that the SADF test has an inherent weakness, evidenced again and again. It could not identify bubbles when the full sample was used, but was able to do so when the sample was truncated. But the recursive application of the GSADF test was able to identify multiple bubbles, without having to arbitrarily truncate / segment / re-select sample starting points. This is a major advantage of GSADF over the SADF procedure. Moreover, the results show that the bubble identification power of the GSADF test increases as the sample size increases.

5. Empirical Applications

We started by extending the Phillips, Shi and Yu (2015) dataset till December 2016, i.e., it extends from January 1871 to December 2016, for a total of 1752 observations. This updated data set is available on Robert Shiller's website. The data used is the real S&P 500 stock price index, and the real S&P 500 stock price index dividend. We then conduct the SADF and the GSADF tests on the stock price dividend ratio (which reflects asset prices in relation to economic / market fundamentals) according to the basic model in eq. (1). The results are given in table 1. Also given are the critical values of the two tests obtained from 2000 replications of 1752 observations.

Table 1

	Test Statistic	Finite Sample Critical Values		
		90%	95%	99%
SADF	7.6167682	1.31375	1.59540	2.05789
GSADF	7.8896859	2.1814	2.4321	2.9317

Both tests find evidence of bubbles or explosive sub-periods over the long-term data (test statics exceed the critical values). We then conduct a bubble monitoring exercise for the S&P 500 stock market using the SADF test and its critical value in Figure 1. The darker line is the critical values and the lighter line represents the SADF test statistics. The existence of a bubble, test statistic greater than the critical value, is evident in the late 19th century, and around the later 1990s (the technology bubble). It also appears to exist around 1929 (the stock market crash). The housing bubble of 2006-07 does not seem to show up in the data.

6. Conclusion

The new test, the GSADF procedure is a recursive test, able to detect multiple bubbles. It's a rolling window, right sided ADF unit root test, with a double sup-window selection criterion. The SADF test is good, but it cannot credibly detect multiple bubbles over the same sample data set. The GSADF test overcomes this weakness and has significant discriminatory power in detecting multiple bubbles. It makes it very relevant in studying the "time trajectory" of long historical data sets. We have an indication for multiple bubbles in the extended data set.

Figure 1

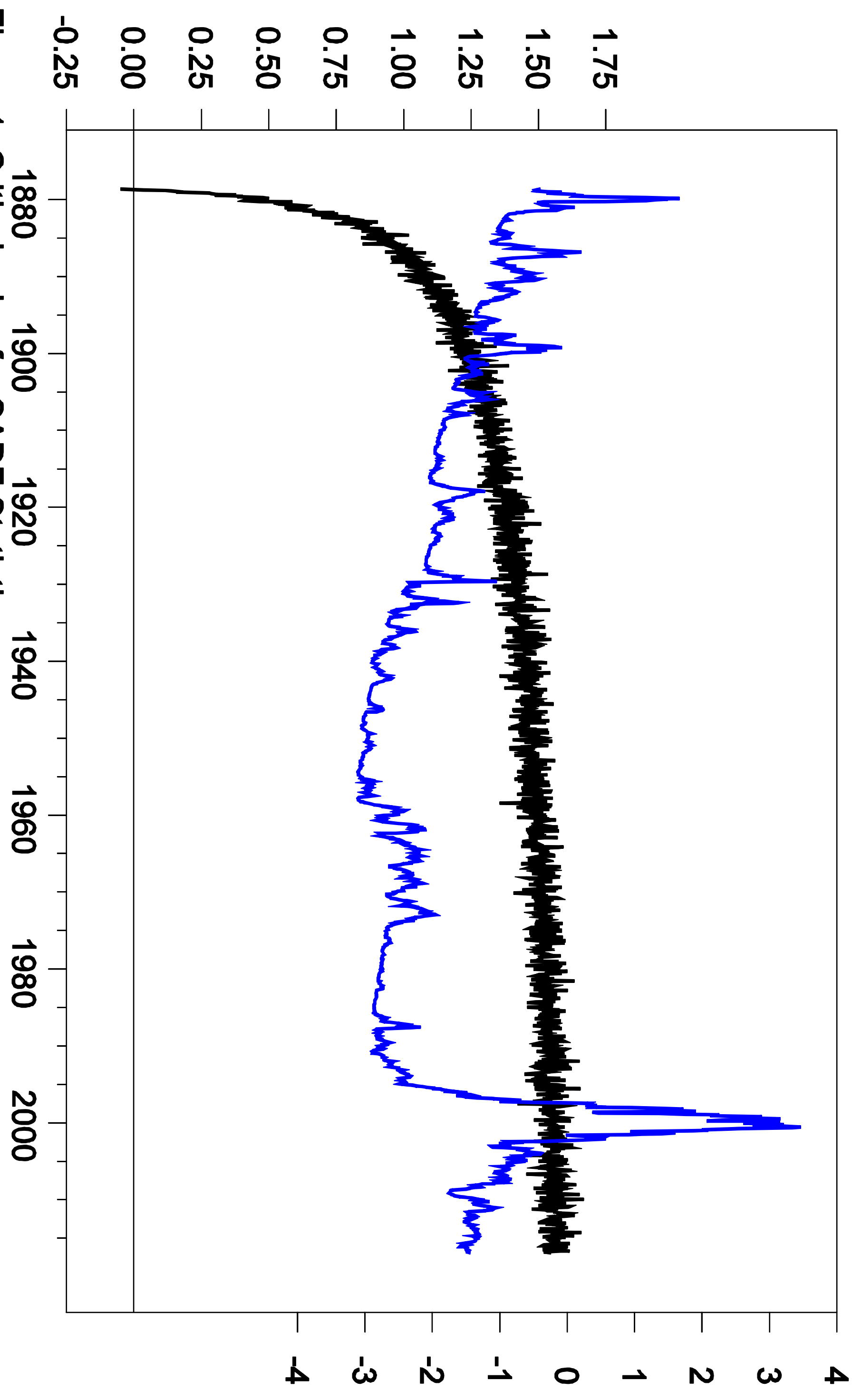

Figure 1. Critical value for SADF Statistics

Notes

1. See Shi (2011) for an overview of this literature.
2. Cochrane (2005) debates the rationale of including "bubble components" in an asset pricing model, while Cooper (2008) expresses bewilderment at the literatures attempt to rationalize the well accepted NASDAQ bubble, as an accurate reflection of the changing market times and environment.
3. Interestingly, the experts agree more on the presence of market exuberance leading to panics, either rationally or irrationally. It's based on changing economic fundamentals, arising from behavior alterations of market players, or due to changing discount rates over time etc.
4. Then there is the Markov-switching test of Hall, et.al (1999), to detect explosive behavior in the data sample, but it is open to suspicion since Shi (2013) found it to be susceptible to "false detection of explosiveness." Also, according to Funke et.al. (1994) and van Norden and Vigfusson (1998), general filtering algorithms cannot differentiate between spurious explosiveness (the marker being high variance) as opposed to generic explosive behavior. The general approach of SADF is also used by Busetti and Taylor (2004) and Kim (2000) among others, to study "market bubbles" but the simulation study done by Homm and Breitung (2012) finds the PWY (SADF) test to be the most powerful metric in detecting multiple bubbles.
5. Eq. (5), Theorem 1, from PSY (2014b)
6. The data process before the origination of the bubble is assumed to be a random walk for convenience, and it is the usual practice, but not necessary for the asymptotic properties to hold.
7. The authors have proven the consistency of $(\hat{r}_e, \hat{r}_f)$ in Phillips and Yu (2009).
8. This sequential procedure (for proper and credible application) requires a long set of observations, the longer the better, in order to re-initialize the test process after a bubble.

References

Abreu, D., and M. K. Brunnermeier, M. K. (2003). Bubbles and crashes. *Econometrica, 71*, 173-204.

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

Busetti, F., and A. M. R. Taylor. (2004). Tests of stationarity against a change in persistence. *Journal of Econometrics, 123*, 33-66.

Cochrane, J. H. (2005). Asset Pricing. Princeton: Princeton University Press.

Cooper, G. (2008). The Origin of Financial Crises: Central Banks, Credit Bubbles and the Efficient Market Fallacy. New York: Vintage Books.

Evans, G.W. (1991). Pitfalls in testing for explosive bubbles in asset prices. *American Economic Review, 81*, 922-930.

Froot, K. A., and M. Obstfeld. (1991). Intrinsic bubbles: The case of stock prices. *American Economic Review, 81*, 1189-1214.

Funke, M., Hall, S., and M. Sola. (1994). Rational bubbles during Poland's hyperinflation: implications and empirical evidence. *European Economic Review, 38*, 1257-1276.

Gurkaynak, R. S. (2008). Econometric tests of asset price bubbles: taking stock. *Journal of Economic Surveys, 22*, 166 -186.

Hall, S.G., Z. Psaradakis, and M. Sola. (1999). Detecting periodically collapsing bubbles: A Markov switching unit root test. *Journal of Applied Econometrics, 14*, 143-154.

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

Kim, J. Y. (2000). Detection of change in persistence of a linear time series. *Journal of Econometrics, 95*, 97-116.

Leybourne, S., Kim T-H., and A. M. R. Taylor. (2007). Detecting multiple changes in persistence. *Studies in Nonlinear Dynamics and Econometrics, 11*, 1-32.

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

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

Phillips, P.C.B., Shi, S., and J. Yu. (2014b). Testing for Multiple Bubbles: Historical Episodes of Exuberance and Collapse in the S&P 500. Working Paper, Yale University.

Phillips, P.C.B., Shi, S., and J. Yu. (2015). Testing for Multiple Bubbles: Limit Theory of Dating Algorithms. *International Economic Review, 56,* 1079–1134

Phillips, P.C.B., Shi, S., and J. Yu. (2015, b). Supplement to Two Papers on Multiple Bubbles. Manuscript available from: https://sites.google.com/site/shupingshi/TN_GSADFtest.pdf?attredirects=0&d=1.

Phillips, P.C.B., Wu, Y., and J. Yu. (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? *International Economic Review, 52,* 201-22.

Phillips, P. C. B., and J. Yu. (2009). 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.

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

Shi, S. (2011). Econometric Tests for Nonlinear Exuberance in Economics and Finance. PhD thesis, The Australian National University.

Shi, S. (2013). Specification sensitivities in the Markov-switching unit root test for bubbles. *Empirical Economics, 45*, 697-713.

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