BE990-8-AU - Research Methods in Financial Econometrics
Topic RT2: Detecting Financial Bubbles

Professor Robert Taylor
Room 3.17, Essex Business School
robert.taylor@essex.ac.uk

Autumn Term, 2025-26

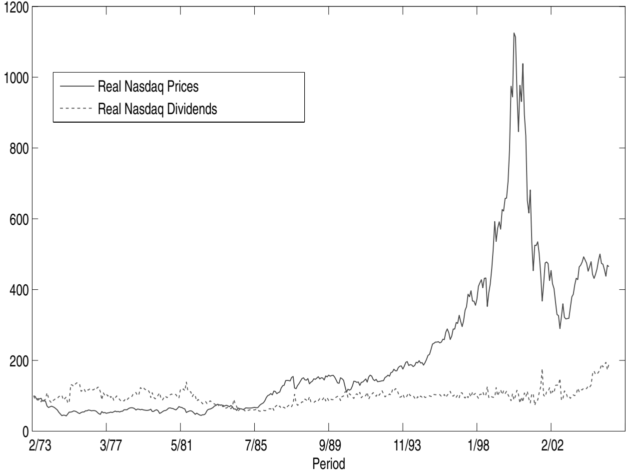
1. Introduction

- "How do we know when irrational exuberance has unduly escalated asset values?" (Alan Greenspan, 1996)
- "Experience can be a powerful teacher. The rise and fall of internet stocks, which created and then destroyed $8 trillion of shareholder wealth, has led a new generation of economists to acknowledge that bubbles can occur." (Alan Krueger, 2005)

- During the 1990s, led by DotCom stocks and the internet sector, the U.S. stock market experienced spectacular rises in all major indices, especially the Nasdaq index. As a result, there was much popular talk among economists about the effects of the internet and computing technology on productivity and the emergence of a "new economy" associated with these changes. What caused the unusual surge and subsequent fall in prices, whether there were bubbles, and whether the bubbles were rational or behavioural have been among the most actively debated issues in macroeconomics and finance in recent years.

- As discussed in Phillips, Wu and Yu (2011), many attribute the episode to financial bubbles: examples include Greenspan (1996), Thaler (1999), Shiller (2000), Cooper *et al.* (2001), Ritter and Welch (2002), Ofek and Richardson (2002), Lamont and Thaler (2003), and Cunado *et al.* (2005).
- Greenspan's (1996) remarks, including the phrase "irrational exuberance" to characterise herd stock market behaviour, have been influential in thinking about financial markets and herd behaviour.
- Immediately after Greenspan coined the phrase, stock markets fell sharply worldwide although this did not halt the general upward march of the United States market. Indeed, over the 1990s, the Nasdaq index rose to the historical high of 5,048.62 points on March 10, 2000 from 329.80 on October 31, 1990; see Figure 1.

Figure 1: Real NASDAQ price and dividend series, 02/1973–06/2005



- In response, academic researchers have started to develop methods to examine empirically the Nasdaq market performance in relation to the market perceptions of exuberance by Greenspan and others.
- To allow data testing we must first define financial exuberance in a time series modelling context. This has been done in terms of explosive autoregressive [AR] behaviour within short periods of the series. Researchers have then introduced new econometric methodology based around sub-sample implementations of the familiar Dickey-Fuller (1979) [DF] unit root statistic to assess the empirical evidence of exuberant behaviour in the Nasdaq stock market index.

- This formulation is compatible with several different explanations of this period of market activity, including the rational bubble literature, herd behaviour, and exuberant and rational responses to economic fundamentals. All these propagating mechanisms can lead to explosive characteristics in the data. Hence, the empirical issue becomes one of identifying the origination, termination, and extent of the explosive behaviour. Although with traditional test procedures "there is little evidence of explosive behavior" (Campbell et al., 1997, p. 260), these new procedures have found evidence of explosive periods of price exuberance in the Nasdaq and others series.

- Among the potential explanations of explosive behaviour in economic and financial variables, the most prominent are perhaps models with rational bubbles. Accordingly, the literature has tended to relate the analysis of explosive behaviour to the rational bubble literature. Here it is well known that standard econometric tests encounter difficulties in identifying rational asset bubbles (Flood and Garber, 1980; Flood and Hodrick, 1986; Evans, 1991).
- The use of tests based on sub-samples of the data overcomes these difficulties enabling detection and dating of periods of exuberance.

- When Greenspan coined the phrase "irrational exuberance" it was perhaps as a warning that the market might be overvalued and at risk of a financial bubble. The subsequent rise and fall of internet stocks to the extent of $8 trillion of shareholder wealth renewed a long-standing interest among economists in the possibility of financial bubbles.
- Theoretical studies on rational bubbles in the stock market include Blanchard (1979), Blanchard and Watson (1982), Shiller (1984), Tirole (1982, 1985), Evans (1989), Evans and Honkapohja (1992), and Olivier (2000). Empirical studies include Shiller (1981), West (1987, 1988), Campbell and Shiller (1987, 1989), Diba and Grossman (1988), Froot and Obstfeld (1991), and Wu (1997). Flood and Hodrick (1990) and Gurkaynak (2005) survey various econometric methodologies and test results for financial bubbles.

- In the rational bubble literature, bubbles manifest explosive characteristics in prices. This statistical property motivates a definition of exuberance propagated by the explosive autoregressive [AR] process

$$x_t = \mu + \phi x_{t-1} + \varepsilon_t,$$

  where, for certain subperiods of the data, $\phi > 1$.
- Rational bubbles can be illustrated using the present value theory of finance whereby fundamental asset prices are determined by the sum of the present discounted values of the expected future dividend sequence.

- Begin with the standard no arbitrage condition,

$$P_t = \frac{1}{1+R} E_t(P_{t+1} + D_{t+1})$$

where P_t is the real stock price (ex-dividend) at time t, D_t is the real dividend received from the asset for ownership between $t-1$ and t, and R is the discount (or risk-free) rate ($R > 0$) assumed time-invariant. E_t denotes expectations taken at time t.

- Campbell and Shiller (1989) show a log-linear approximation and recursive substitution yields

$$p_t = p_t^f + b_t$$

where $p_t = \log(P_t)$. By convention, p_t^f, which is exclusively determined by expected dividends, is called the fundamental component of the stock price, and b_t is called the rational bubble component and follows a first-order explosive AR.

- The properties of p_t are therefore determined by p_t^f and b_t. In the absence of bubbles, i.e., $b_t = 0$, for all t, we will have $p_t = p_t^f$, and so p_t is determined solely by p_t^f and, hence, by $d_t = \log(D_t)$. In this case if p_t and d_t are both (unit root) integrated processes of order one, denoted I(1), then they can be shown to be co-integrated with the co-integrating vector $[1, -1]$.

- If a bubble is present in the stock price, i.e. $b_t \neq 0$, this formulation requires that any rational investor, who is willing to buy that stock, must expect the bubble to grow at rate R. If this is the case and if b_t is strictly positive, this sets the stage for speculative investor behaviour: rational investors are willing to buy an "overpriced" stock in the belief that through price increases, they will be sufficiently compensated for the extra payment b_t. If investors expect prices to increase at rate R and buy shares, the stock price will indeed rise and complete the loop of a self-fulfilling prophecy.

- If bubbles are present, i.e., $b_t \neq 0$, then p_t like b_t will be explosive, irrespective of whether d_t is I(1), or a stationary process, denoted I(0). Here $\Delta p_t = p_t - p_{t-1}$ is also explosive and therefore cannot be I(0). This implication motivated Diba and Grossman (1988) to look for bubble behaviour by applying standard unit root tests to Δp_t. Finding an empirical rejection of the null of a unit root in Δp_t, Diba and Grossman (1988) concluded that p_t was not explosive and so there was no bubble in the stock market.
- Where d_t is I(1), Diba and Grossman (1988) looked for evidence of the absence of bubbles by testing for a co-integrating relation between p_t and d_t. In the presence of bubbles, p_t is always explosive and hence cannot co-move or be co-integrated with d_t if d_t is itself not explosive. Therefore, an empirical finding of co-integration between p_t and d_t may be taken as evidence against bubbles.

- Evans (1991) questioned the validity of the empirical tests employed by Diba and Grossman (1988) by arguing that none of these tests have much (if any) power to detect periodically collapsing bubbles. He demonstrated by simulation that the low power of standard unit root and co-integration tests in this context is due to the fact that a periodically collapsing bubble process can behave much like an I(1) process or even like a stationary linear autoregressive process provided that the probability of collapse of the bubble is not negligible. As a result, Evans (1991, p. 927) claimed that "periodically collapsing bubbles are not detectable by using standard tests."

- The foregoing suggests that a direct way to test for bubbles is to examine evidence for any explosive behaviour in p_t and d_t. Of course, explosive characteristics in p_t could in principle arise from d_t and the two processes would then be explosively co-integrated. However, if d_t is demonstrated to be nonexplosive, then the explosive behaviour in p_t will provide sufficient evidence for the presence of bubbles because the observed behaviour may only arise through the presence of b_t.
- It seems likely that in practice explosive behaviour in p_t may only be temporary or short-lived, as in the case of stock market bubbles that collapse after a certain period of time. This can be taken into account empirically by looking at *sub-samples* of the data.

- Motivated by this, Phillips, Wu and Yu (2011, In. Econ. Rev.) (PWY), focus on testing for explosive AR behaviour using the largest of a set of forward recursive (these are particular sequences of sub-samples) right-tailed DF tests applied to the price and dividend series in levels only. If the test finds explosive AR behaviour for the prices but not for the dividends, this indicates that an explosive rational bubble exists.
- PWY apply their tests to the Nasdaq Composite stock price index and dividend index between February 1973 and June 2005; the test detects the presence of an explosive rational bubble beginning in mid-1995.
- These tests have become popular in testing for speculative bubbles in asset prices. Gilbert (2010), using commodities futures prices (2006-2008), finds evidence of bubbles in copper, nickel and crude oil markets.

- Homm and Breitung (2012) apply the PWY test and a related Chow-type test to stock, commodity and house price data, finding a body of evidence for bubbles.
- Using the PWY test, Bettendorf and Chen (2013) find evidence of explosive bubbles in the sterling-US dollar nominal exchange rate. This appears to be driven by explosive behaviour in the relevant price index ratio for traded goods.

- A number of further bubble detection procedures have been developed since the original PWY procedure. Most notably, Phillips, Shi and Yu (2014) (PSY) develop two further procedures based on backward recursions through the data, and on a double recursion. The first is designed to detect end of sample bubbles, while the second is used to detect multiple possible bubbles in a series, rather than a single possible bubble episode.
- Astill, Harvey, Leybourne and Taylor (2017) develop a procedure for detecting an end of sample bubble using a sub-sampling method introduced in Andrews (2003). They show this to display greater power than the PSY Dickey-Fuller based tests to detect end-of-sample explosive episodes.

2. The Stochastic Bubble Model

- DGP:

$$y_t = \mu + u_t, \quad t = 1, ..., T \tag{1}$$

$$u_t = \begin{cases} u_{t-1} + \varepsilon_t, & t = 1, ..., \lfloor \tau_{1,0} T \rfloor, \\ (1 + \delta_1) u_{t-1} + \varepsilon_t, & t = \lfloor \tau_{1,0} T \rfloor + 1, ..., \lfloor \tau_{2,0} T \rfloor, \\ (1 - \delta_2) u_{t-1} + \varepsilon_t, & t = \lfloor \tau_{2,0} T \rfloor + 1, ..., \lfloor \tau_{3,0} T \rfloor, \\ u_{t-1} + \varepsilon_t, & t = \lfloor \tau_{3,0} T \rfloor + 1, ..., T \end{cases}$$

  where $\delta_1 \geq 0$ and $\delta_2 \geq 0$. We assume that $\varepsilon_t \sim I(0)$.

- When $\delta_1 > 0$, y_t follows a unit root up to time $\lfloor \tau_{1,0} T \rfloor$, after which it displays explosive AR behaviour over $t = \lfloor \tau_{1,0} T \rfloor + 1, ..., \lfloor \tau_{2,0} T \rfloor$. When applied to asset prices, and assuming unit root behaviour in the corresponding dividend series, this can be interpreted as a bubble regime.

- At the end of the bubble period: if $\delta_2 = 0$, y_t reverts to unit root dynamics directly, while if $\delta_2 > 0$, this happens after an interim stationary regime over $t = \lfloor \tau_{2,0} T \rfloor + 1, ..., \lfloor \tau_{3,0} T \rfloor$. The latter provides a model of a crash regime, where the mean-reverting stationary behaviour acts to "offset" the explosive period to some extent.
- The magnitude of δ_2 and the duration of the collapse regime provide a flexible way of controlling the rapidity and extent to which price corrections occur when an asset price bubble terminates.
- The DGP also admits a bubble (or collapse) regime continuing to the end of the sample period, on letting $\tau_{2,0} = 1$ (or $\tau_{3,0} = 1$). When $\delta_1 = 0$, no explosive regime is present, and we also assume $\delta_2 = 0$, so that collapse regimes do not occur without a prior bubble.

- The null hypothesis, H_0, is that no bubble is present in the series such that y_t follows a unit root process throughout the sample period, i.e. $H_0 : \delta_1 = 0$ (and hence $\delta_2 = 0$). The alternative hypothesis is given by $H_1 : \delta_1 > 0$, and corresponds to the case where a bubble is present in the series, which either runs to the end of the sample (if $\tau_{2,0} = 1$), or terminates in-sample, either with or without a subsequent collapse regime depending on whether $\delta_2 = 0$ or $\delta_2 > 0$.

3. The PWY Test

- To test H_0 against H_1, PWY propose a test based on the largest of the sequence of recursive right-tailed DF tests. For serially uncorrelated ε_t, the PWY statistic is

$$PWY = \max_{\tau \in [\tau_0, 1]} DF_{0,\tau}$$

where $DF_{0,\tau}$ is the standard DF statistic, ie the t-ratio for $\hat{\phi} = 0$ in the fitted OLS regression

$$\Delta y_t = \hat{\alpha} + \hat{\phi} y_{t-1} + \hat{\varepsilon}_t \tag{2}$$

calculated over the sub-sample $t = 1, ..., \lfloor \tau T \rfloor$, i.e.

$$DF_{0,\tau} = \frac{\hat{\phi}}{\sqrt{\hat{\sigma}^2 / \sum_{t=2}^{\lfloor \tau T \rfloor} \left(y_{t-1} - \bar{y}_\tau \right)^2}}$$

where $\bar{y}_\tau = \left(\lfloor \tau T \rfloor - 1 \right)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} y_{t-1}$ and $\hat{\sigma}^2 = \left(\lfloor \tau T \rfloor - 3 \right)^{-1} \sum_{t=2}^{\lfloor \tau T \rfloor} \hat{\varepsilon}_t^2$.

- The *PWY* statistic is therefore the largest of a sequence of forward recursive DF statistics with minimum sample length $\lfloor \tau_0 T \rfloor$. In what follows we follow PWY and set $\tau_0 = 0.1$.
- The test rejects for large positive values of the *PWY* statistic. The null distribution of the *PWY* statistic is not standard normal, however. Critical values obtained by simulation methods are provided in PWY.
- The *PWY* test is designed to detect a single bubble occurring at some point in the sample.
- Where ε_t is serially correlated, sufficient lagged dependent variables must be included in (2).

4. The PSY Tests

- Rather than being based on the largest of the sequence of forward recursive sub-sample DF statistics, the first test proposed in PSY takes the largest of the sequence of backward recursive sub-sample DF statistics; i.e.,

$$BSADF = \max_{\tau \in [0, 1-\tau_0]} DF_{\tau,1}$$

  where $DF_{\tau,1}$ is now the standard DF statistic calculated as before but over the sub-sample $t = \lfloor \tau T \rfloor, ..., T$. Critical values in PSY. The test rejects for large positive values of $BSADF$.
- This test is considerably more powerful than the PWY test to detect a single explosive episode at the end of the sample.

- The second procedure proposed in PSY is designed for the case where the series may contain more than one bubble episode. It is based on a doubly recursive sequence of sub-sample DF tests. This is the same as running a set of rolling sub-sample windows through the sample for windows of different width. This design is ideally suited to picking up multiple pockets of explosive behaviour in the sample.
- The double recursive test rejects for large positive values of

$$GSADF = \max_{r_2 \in [r_0, 1], r_1 \in [0, r_2 - r_0]} DF_{r_1, r_2}$$

where DF_{r_1,r_2} denotes the standard DF test calculated using the observations $t = \lfloor r_1 T \rfloor, ..., \lfloor r_2 T \rfloor$. Critical values in PSY.

- A number of interesting issues surround the PWY and PSY tests.
- A rejection suggests explosive behaviour somewhere in the sample. But the tests do not tell us where the bubbles occur. PWY and PSY develop auxiliary procedures, conducted where the bubble detection tests reject, to *date stamp* the beginning and end of the bubble regime(s).
- Like standard full sample DF tests, the PWY and PSY tests assume the errors, ε_t, are homoskedastic. It is not hard to see that if ε_t displayed a period of very high volatility in a sub-sample of the data that this would be observationally similar to an explosive bubble. Harvey, Leybourne, Sollis and Taylor (2016) show that the PWY test severely spuriously over-rejects the null hypothesis in such cases. The problem can be solved by using bootstrap methods. We will return to this issue in Topic RT4.

- The tests in PWY and PSY have tended to be used retrospectively, eg to see if there is statistical evidence of a DotCom bubble etc. But in reality we may be much more interested in *real-time monitoring* of the data with a view to picking up a bubble NOW as early as possible while it is on-going. As we will see, the *BSADF* test can be used for this purpose, although it does not deliver the best possible chance of detecting end-of-sample bubbles as they happen.

5. The AHLT Tests

- For simplicity of exposition consider the simpler version of the bubble DGP defined earlier:

$$y_t = \mu + u_t, \quad t = 1, ..., T + m \tag{3}$$

$$u_t = \begin{cases} u_{t-1} + \varepsilon_t, & t = 1, ..., T, \\ \phi u_{t-1} + \varepsilon_t, & t = T + 1, ..., T + m \end{cases} \tag{4}$$

  with $\phi \geq 1$ and where the total sample length is equal to $N = T + m$.

- The series $\{y_t\}$ follows a unit root process for the first T observations and is then subject to (potential) explosive AR behaviour over the final m observations.

- The length of the bubble, m, is going to be small relative to the remaining sample size T.

- Astill, Harvey, Leybourne and Taylor (2017) [AHLT] propose a testing methodology based on the end-of-sample instability testing approach of Andrews (2003). Their procedures do not require the length of the bubble to be increasing with the sample size, as is required for the large sample validity of the PWY and PSY tests.
- The approach involves calculating the test statistic over a window of m end-of-sample observations rather than over the sample as a whole. A critical value for the test is obtained by (overlapping) sub-sampling methods.
- Specifically, $T - m + 1$ test statistics, analogous to the test statistic of interest, are calculated using a rolling window of m observations, from $t = 1, ..., m$ through to $t = T - m + 1, ..., T$ (called the *training sample*), yielding an empirical CDF for the null distribution of the test statistic from which a critical value can be taken.

- AHLT consider a number of tests within this Andrews-type framework, where the intention is to distinguish between $H_0 : \phi = 1$ and $H_1 : \phi > 1$ in the end-of-sample window.
- To do this, they compare a test statistic designed to detect explosivity based on $t = T + 1, ..., T + m$, with a critical value obtained from this same statistic applied to the $T - m + 1$ sub-samples prior to the end-of-sample window.

- A natural candidate to use in this approach is the DF statistic from the regression

$$\Delta y_t = \hat{\alpha} + \hat{\phi} y_{t-1} + \hat{\varepsilon}_t, \;\; t = j+1, ..., j+m \tag{5}$$

  The test statistic of interest is calculated on the sub-sample $t = T+1, ..., T+m$, with the critical value obtained using the test statistics calculated on the sub-samples $t = j+1, ..., j+m$, for each of $j = 1, ...T-m$ (NB only $T-m$ sub-samples, as first lost due to differencing the data).
- H_0 is rejected in favour of H_1 in the upper-tail of the statistic's distribution. Notice that lagged difference augmentation is not required since any dependence in ε_t is common to all sub-samples.
- The Andrews-type approach applied to DF will result in a correctly sized test for large T, although for small m the test may lack power because the autoregressive parameter ϕ is likely to be inaccurately estimated.

- A simple alternative statistic to employ in the Andrews-type framework can be motivated by considering the properties of the first differences of y_t. Under H_0, $\Delta y_t = \varepsilon_t$ throughout the full sample period. But under H_1, $\Delta y_t = \varepsilon_t$ up to time $t = T$, at which point the bubble regime commences and $\Delta y_t = (\phi - 1)u_{t-1} + \varepsilon_t$.
- An explosive series cannot be differenced to stationarity. As such Δy_t will be I(0) for the first T observations and explosive for the final m observations. Andrews and Kim (2006) designed a test, here denoted R, for an end-of-sample change from I(0) to I(1) behaviour. This test should also have power against an explosive alternative.

- The R statistic of Andrews and Kim (2006) is given by

$$R := \sum_{t=j+1}^{j+m} \left(\sum_{s=t}^{j+m} \Delta y_s \right)^2 . \quad (6)$$

It will have power to detect both upwardly and downwardly explosive series.

- As upwards exploding series are arguably of greater relevance than downwardly explosive series we could also utilise a one-sided variant of R given by

$$\begin{aligned} S &:= \sum_{t=j+1}^{j+m} \sum_{s=t}^{j+m} \Delta y_s \\ &\equiv \sum_{t=j+1}^{j+m} (t-j) \Delta y_t \end{aligned} \quad (7)$$

- A feature of the Andrews-type approach using either the DF, R or S statistics is that the size (probability of rejecting the null when it is true) of the procedure will be (approximately) unaffected by the presence of level shifts in the data or by the presence of (temporary) bubbles in the training sample.
- The approaches based on R and S are not robust to volatility shifts in the training sample, however. But robustness can be achieved by introducing straightforward sub-sample studentisations, viz:

$$S^* := \frac{S}{\sqrt{\sum_{t=j+1}^{j+m} (\Delta y_t)^2}}, \quad R^* := \frac{R}{\sum_{t=j+1}^{j+m} (\Delta y_t)^2}$$

- Further variants with a White-type correction in the studentisation are also considered in AHLT which are robust to volatility shifts in both the training sample and the last m observations (see section 4 of AHLT for details).

- In practice, we need to make a choice regarding the window width used when implementing the Andrews-type approach. The true bubble length m is unknown. Tests will be estimated using some window width m'.
- Using too large a window width may impact the finite sample power of the tests under H_1 as some data generated under H_0 will be included in the test statistic of interest. Using too small a window may also impact the finite sample power of the test as some data generated under H_1 may be included in the sub-sample statistics, thus distorting the critical value.

6. Monte Carlo Power Studies

6.1 Power to detect a Single Bubble

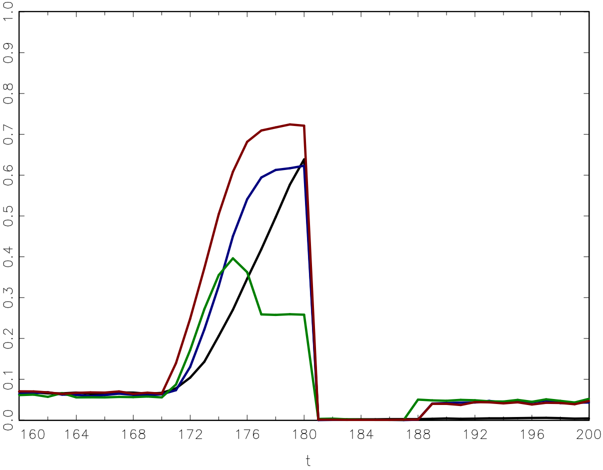
- We now examine the ability of the tests to detect a single explosive bubble, and their properties following its collapse.
- First, data were generated according to the DGP

$$y_t = \begin{cases} y_{t-1} + \varepsilon_t, & t = 1, ..., 170, \\ \phi y_{t-1} + \varepsilon_t, & t = 171, ..., 180, \\ y_{170} + \varepsilon_t & t = 181, \\ y_{t-1} + \varepsilon_t, & t = 182, ..., N \end{cases} \tag{8}$$

the rejection frequency for each test over the sub-samples $t = 1, ..., E$ with $E = \{160, 161, ..., 200\}$ was calculated.

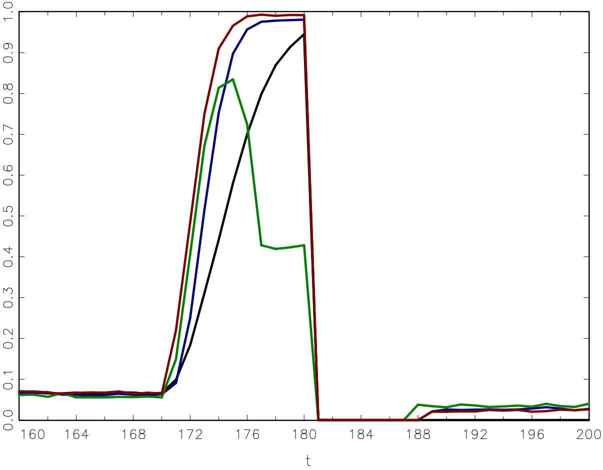
- This simulation exercise is designed to examine how the probability of each test to detect a bubble evolves as the bubble episode continues.
- Results are reported for a window width of $m' = 8$, but results are similar for other window lengths.

Rejection Frequency - Single Collapsed Bubble, $\phi = 1.01$



R_8^*: ——, DF_8: ——, S_8^*: ——, BSADF: ——

Rejection Frequency - Single Collapsed Bubble, $\phi = 1.02$



R_8^*: ——, DF_8: ——, S_8^*: ——, BSADF: ——

- As can be seen from these graphs, the AHLT tests are better able to detect a short-lived bubble.
- Moreover, the robustness of the AHLT tests to past bubble episodes can be seen by the fact that the size of these tests are little impacted by the past bubble once the test statistics of interest no longer includes the crash date data point.
- The size of the *BSADF* test drops to practically zero after the collapse date and remains there for the remainder of the sample. Indeed it is this feature that makes these statistics useful for date-stamping bubbles.
- We now examine how the size of the tests after a collapsed bubble affects their ability to detect future bubbles.

6.1 Power with Multiple Bubbles

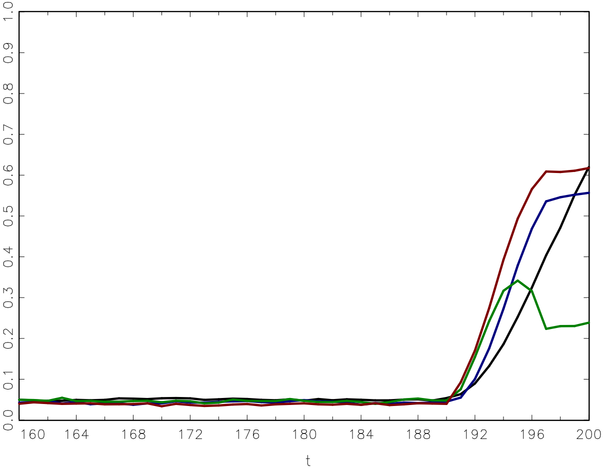
- To examine the behaviour of the tests when the sample contains multiple bubbles, data were generated according to the DGP:

$$y_t = \begin{cases} y_{t-1} + \varepsilon_t, & t = 1, ..., TB_{1,o}, \\ 1.01 y_{t-1} + \varepsilon_t, & t = TB_{1,o} + 1, ..., TB_{1,e}. \\ y_{TB_{1,o}} + \varepsilon_t & t = TB_{1,e} + 1. \\ y_{t-1} + \varepsilon_t, & t = TB_{1,e} + 2, ..., TB_{2,o}, \\ 1.01 y_{t-1} + \varepsilon_t, & t = TB_{2,o} + 1, ..., N \end{cases} \quad (9)$$

  where $N = 200$.
- This DGP allows for the series to contain two bubbles, both an end of sample bubble from $t = TB_{2,o} + 1, ..., N$ and a mid-sample bubble from $t = TB_{1,o} + 1, ..., TB_{1,e}$ that collapses at $t = TB_{1,e} + 1$.

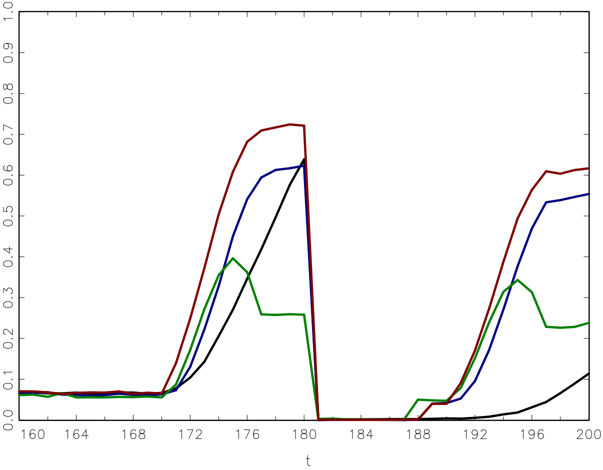
- We set $TB_{2,o} = 190$ and examine two cases:
- The first is the case where the initial bubble and the end-of-sample bubble are separated by a relatively long time period ($TB_{1,o} = 95$ and $TB_{1,e} = 105$).
- The second is the case where the initial bubble and the end-of-sample bubble are separated by a relatively short time period ($TB_{1,o} = 170$ and $TB_{1,e} = 180$).
- Again we utilise a window width for the AHLT tests of $m' = 8$.

Multiple Bubbles - Long Separation



R_8^*: ——, DF_8: ——, S_8^*: ——, BSADF: ——

Multiple Bubbles - Short Separation



R_8^*: ——, DF_8: ——, S_8^*: ——, BSADF: ——

- When the two bubbles are separated by a relatively long time period, the size of the *BSADF* test has recovered to some extent by the time of the second bubble.
- However, when the two bubbles are separated by a relatively short time period, the undersize exhibited by the *BSADF* test following a collapsed bubble severely impacts on its ability to detect the second bubble.
- Comparing the two graphs we see that the distance between the two bubble episodes has little impact on the ability of the AHLT tests to detect the second bubble.

7. Empirical Application

- We now compare the performance of the S^* test relative to the *BSADF* test by undertaking the pseudo real-time detection exercise of PSY. S^* was chosen as further simulations show it to have the best overall power of the AHLT tests.
- Using monthly data on the S&P500 price dividend ratio from 1987M01-2010M12, we perform the S^* test for samples ending at each date in the sample, replicating how an on-going (real time) bubble detection exercise would evolve.
- We then record the earliest date that the S^* test finds evidence of a bubble for a number of known bubble episodes and compare this to the first date bubble behaviour is detected by the *BSADF* test.

Past Bubble	$BSADF$	S^*_5	S^*_8	S^*_{10}
Post long-depression	1879M10	1879M10	1879M11	1879M11
The Great Crash	1928M11	1927M08	1927M08	1927M08
Postwar Boom	1955M01	1954M02	1954M05	1954M05
Black Monday	1986M06	-	1986M04	1986M04
Dot-com Bubble	1995M11	1995M05	1995M06	1995M06

- As can be seen, the S^* test would often have detected past bubbles sooner than the $BSADF$ test if used as part of an on-going monitoring exercise.
- Using a smaller window width leads to earlier detection in some instances, but using a window width of 5 the S^* test would have failed to detect the bubble leading up to Black Monday.

8. Real-time Monitoring for Bubbles

- As we have seen in sections 5, 6 and 7, real-time monitoring for a bubble could be undertaken by the repeated application (as extra sample data becomes available with the passage of time) of the end-of-sample one-shot bubble detection tests used in the simulation study in section 6; e.g. the *BSADF* test of PSY or the R (or R^*) or S (or S^*) tests of AHLT.
- However, there is an obvious problem with this approach. In particular the false positive rate [FPR] of these monitoring procedures would not be controlled and they would all eventually signal a bubble with probability one, even where no bubble was present in the monitoring period. This is an example of the *multiple testing problem* bias.

- We will now briefly discuss two real-time bubble monitoring procedures which avoid the multiple testing issue such that the practitioner can set the FPR they wish for the procedure.
- The first is a CUSUM-based method, due to Homm and Breitung (2012, Journal of Financial Econometrics).
- The second, due to Astill *et al.* (2018, Journal of Time Series Analysis), is an approach based on the one-shot tests of AHLT from section 5, but which is able to deliver a controlled FPR.

8.1 The Homm-Breitung CUSUM-based Procedure

- Under the assumption that $\varepsilon_t \sim IID(0, \sigma^2)$, and assuming a *training period* of $t = 1, ..., T$, Homm and Breitung (2012) propose testing for explosive behaviour in the monitoring period using the following CUSUM statistic:

$$S_T^t := \frac{1}{\tilde{\sigma}_t} \sum_{j=T+1}^{t} \Delta y_j$$

where $t > T$ is the monitoring observation, and where $\tilde{\sigma}_t^2$ is a consistent estimate of σ^2; in their numerical work, Homm and Breitung (2012) use the first-difference estimator, $\tilde{\sigma}_t^2 := (t-1)^{-1} \sum_{j=2}^{t} \Delta y_j^2$.

- Homm and Breitung (2012) show that if the CUSUM statistic, S_N^t, is computed sequentially at dates $t = T+1, ..., \lfloor \lambda T \rfloor$, then under the null hypothesis, H_0, of no explosive behaviour, for any $\lambda > 1$

$$T^{-1/2} S_T^{\lfloor Tr \rfloor} \Rightarrow W(r) - W(1), \quad 1 < r \leq \lambda \tag{10}$$

where $W(r)$ is a standard Wiener process, and, hence, that

$$\begin{aligned} \lim_{T\to\infty} \Pr(|S_T^t| &> c_t\sqrt{t} \text{ for some } t \in T+1, ..., \lfloor \lambda T \rfloor) \\ &\leq \exp\left(-b_\alpha/2\right) \end{aligned}$$

where $c_t := \sqrt{b_\alpha + log(t/T)}$.

- The CUSUM monitoring procedure proposed in Homm and Breitung (2012) then rejects H_0 if $S_T^t > c_t\sqrt{t}$ for some $t > T$, with an explosive episode signaled at the first time point t in the monitoring period for which such an exceedance occurs. For such a (one-sided upper-tail) test at the $\alpha = 0.05$ significance level, the appropriate asymptotic setting for b_α used to compute c_t is $b_\alpha = 4.6$.
- The CUSUM procedure has the maintained hypothesis that no bubbles are present in the training sample.

8.2 The MAX_m Procedure of Astill *et al.* (2018)

- The MAX_m monitoring procedure of Astill *et al.* (2018) is based on the sequential application of AHLT's statistic with White-type studentisation:

$$S_t^{*w} := \frac{\sum_{j=t-m+1}^{t}(j-t+m)\Delta y_j}{\sqrt{\sum_{j=t-m+1}^{t}\{(j-t+m)\Delta y_j\}^2}}, \quad t > T$$

  where, as before, m is a user chosen window width.

- The MAX_m procedure signals the presence of an explosive episode if at any point t, $T < t \leq \lfloor T\lambda \rfloor$, during the monitoring period the statistic S_t^{*w} exceeds the maximum value across the corresponding sequence of statistics S_i^{*w}, $i = m+1, ..., T-m+1$, calculated over the *training period*; that is, H_0 is rejected if $\max_{k\in[T+1,\lfloor \lambda T\rfloor]} S_k^{*w} > \max_{k\in[m+1,T-m+1]} S_k^{*w}$.

- Under the null hypothesis the FPR is approximately equal to the ratio of the number of test statistics conducted in the monitoring period to the number of test statistics conducted across the monitoring and training periods combined.
- Astill *et al.* (2018) demonstrate that an approximation to the FPR of this procedure is given by

$$\alpha := \frac{\lfloor \lambda T \rfloor - T}{\lfloor \lambda T \rfloor - 2m + 1}. \tag{11}$$

The FPR of the MAX_m monitoring procedure at any point t, $T < t \leq \lfloor T\lambda \rfloor$, in the monitoring period can be computed using (11) by replacing $\lfloor \lambda T \rfloor$ with t.

- The FPR is a function of the length of the training period, T, the window width, m, used in the S_t^{*w} statistics, and of the length of the monitoring horizon, $\lfloor T\lambda \rfloor$.
- We can use the formula to calculate the maximum monitoring horizon such that the FPR is controlled at some chosen value, α^*.
- A shorter training period, ending at time $\lfloor \gamma T \rfloor - m + 1$, $\gamma < 1$, could be used instead. This will alter the FPR of the procedure.
- The Astill *et al.* (2018) monitoring procedure is robust to bubbles in the training sample and to time-varying volatility (changes in the variance of ε_t) while the CUSUM of Homm and Breitung (2012) is not. Astill *et al.* (2023) generalise the CUSUM method to allow for time-varying volatility in ε_t; this method is also robust to the presence of bubbles in the training sample.

References

Note: A large number of the references cited in the text can be found in Phillips, Wu and Yu (2011) and so are not repeated here. References from Topic RT1 are not repeated.

- Andrews, D.W.K. (2003). End-of-sample instability tests. Econometrica 71, 1661-1694.
- Andrews, D.W.K. and Kim, J.-Y. (2006). Tests for cointegration breakdown over a short time period. Journal of Business and Economic Statistics 24, 379-394.
- Astill, S., Harvey, D.I., Leybourne, S.J. and Taylor, A.M.R. (2017). Tests for an end-of-sample bubble in financial time series. Econometric Reviews 36, 651-666.
- Astill, S., Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., and Zu, Y. (2003). CUSUM-based monitoring for explosive episodes in financial data in the presence of time-varying volatility. Journal of Financial Econometrics 21, 187–227.

- Astill, S., Harvey, D.I., Leybourne, S.J., Sollis, R. and Taylor, A.M.R. (2018). Real-time monitoring for explosive financial bubbles. Journal of Time Series Analysis 39, 863–891.
- Bettendorf, T. and Chen, W. (2013). Are there bubbles in the sterling-dollar exchange rate? New evidence from sequential ADF tests. Economics Letters 120, 350–353.
- Gilbert, C.L. (2010). Speculative influences on commodity futures prices 2006–2008. Discussion Paper 197, United Nations Conference on Trade and Development (UNCTAD), Geneva.
- Harvey, D.I., Leybourne, S.J., Sollis, R. and Taylor, A.M.R. (2016). Tests for explosive financial bubbles in the presence of non-stationary volatility. Journal of Empirical Finance 38, 548-574.

- Homm, U. and Breitung, J. (2012). Testing for speculative bubbles in stock markets: a comparison of alternative methods. Journal of Financial Econometrics 10, 198-231.
- Phillips, P.C.B., Shi, S.-P. and Yu, J. (2015). Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P 500. International Economic Review 56, 1043-1078.
- Phillips, P.C.B., Wu, Y. and Yu, J. (2011). Explosive behavior in the 1990s Nasdaq: when did exuberance escalate stock values? International Economic Review 52, 201-226.