JOURNAL OF TIME SERIES ANALYSIS
J. Time Ser. Anal. **39**: 863–891 (2018)
Published online 19 July 2018 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12409

SPECIAL ISSUE

REAL-TIME MONITORING FOR EXPLOSIVE FINANCIAL BUBBLES

SAM ASTILL,[a*] DAVID I. HARVEY,[b] STEPHEN J. LEYBOURNE,[b] ROBERT SOLLIS[c] AND A. M. ROBERT TAYLOR[a]

[a] *Essex Business School, University of Essex, Colchester, UK*
[b] *Granger Centre for Time Series Econometrics, School of Economics, University of Nottingham, Nottingham, UK*
[c] *Newcastle University Business School, Newcastle upon Tyne, UK*

We propose new methods for the real-time detection of explosive bubbles in financial time series. Most extant methods are constructed for a fixed sample of data and, as such, are appropriate only when applied as one-shot tests. Sequential application of these tests, declaring the presence of a bubble as soon as one of these statistics exceeds the one-shot critical value, would yield a detection procedure with an unknown false-positive rate likely to be far in excess of the nominal level. Our approach sequentially applies the one-shot tests of Astill *et al.* (2017), comparing sub-sample statistics calculated in real time during the monitoring period with the corresponding sub-sample statistics obtained from a prior training period. We propose two procedures: one based on comparing the real-time monitoring period statistics with the maximum statistic over the training period, and another that compares the number of consecutive exceedances of a threshold value in the monitoring and training periods, the threshold value obtained from the training period. Both allow the practitioner to determine the false-positive rate for any given monitoring horizon, or to ensure that this rate does not exceed a specified level by setting a maximum monitoring horizon. Monte Carlo simulations suggest that the finite-sample false-positive rates lie close to their theoretical counterparts, even in the presence of time-varying volatility and serial correlation in the shocks. The procedures are shown to perform well in the presence of a bubble in the monitoring period, offering the possibility of rapid detection of an emerging bubble in a real-time setting. An empirical application to monthly stock market index data is considered.

Received 03 April 2018; Accepted 30 May 2018

Keywords: Rational bubble; explosive autoregression; real-time monitoring procedure; subsampling.

JEL. C22; C12; G14.

MOS subject classification: 62G09; 62L12; 62M10; 62P20.

1. INTRODUCTION

The presence of historical asset price bubbles, in which asset prices rise well above their fundamental value at a particular point in time, is widely documented. Well-known historical episodes include the South Sea bubble of 1720, the Dot-Com bubble that originated in the mid-1990s, and the U.S. housing market bubble of the late 1990s and early 2000s, while the Bitcoin price can be seen as a very recent example. In all instances, asset prices, having risen to unsustainable levels, were subject to large crashes, causing significant economic damage. Given the damage caused by the collapse of asset price bubbles, it is of vital importance for policy makers to be able to identify asset price bubbles as they occur to attempt to limit their economic damage.

In light of this, a number of tests for asset price bubbles have been proposed in the economic and financial literature. The seminal paper of Diba and Grossman (1988) proposed testing for asset price bubbles using standard left-tailed augmented Dickey–Fuller (ADF) test statistics applied to both the levels and first differences of a series. More recently, the detection of asset price bubbles using right-tailed ADF tests applied to the levels of a series has been discussed in depth. The first contribution in the literature was made by Phillips *et al.* (2011), who developed a test of the null of no explosive behaviour against the alternative of explosivity based on a sequence of *forward*

* Correspondence to: Sam Astill, Essex Business School, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, UK.
E-mail: sastill@essex.ac.uk.

recursive right-tailed ADF statistics applied to both the price and dividend series of a particular asset, with a bubble signalled if explosivity is found in the price series but not in the corresponding dividend series.

While early contributions, such as those of Diba and Grossman (1988) and Phillips *et al.* (2011), were designed to detect a historical asset price bubble in a series, the policy relevance of detecting a historical bubble episode is perhaps limited given that the subsequent collapse of such bubbles will already have occurred. Arguably of considerably more empirical interest is the detection of ongoing asset price bubbles. As such, recent developments in the literature have focused on detecting end-of-sample asset price bubbles prior to their collapse. Phillips *et al.* (2015) proposed tests for an end-of-sample bubble based on a sequence of *backward* recursive ADF statistics applied to the price and dividend levels of a series, and showed that performing a recursion in this manner yields a test with better power to detect end-of-sample bubbles than the tests of Phillips *et al.* (2011). More recently, Astill *et al.* (2017) (AHLT hereafter) proposed a test for end-of-sample asset price bubbles in which a test statistic is applied to the first differences of a small number of end-of-sample observations. Critical values for the test are estimated using the sub-sampling method of Andrews (2003) and Andrews and Kim (2006), whereby a large number of statistics analogous to the statistic of interest are calculated over a training period within which the null hypothesis of no explosivity is assumed to hold. AHLT show that this method displays greater power than the tests of Phillips *et al.* (2015) for the sort of short-lived end-of-sample bubble episodes that are arguably of most interest to practitioners.

A major limitation of the tests described above is that they are designed for use as one-shot tests applied at a given nominal significance level. In practice, it would arguably be more useful for practitioners to be able to sequentially apply a test for asset price bubbles as new data points are obtained as part of an ongoing real-time monitoring exercise. While sequential application of the tests of Phillips *et al.* (2015) or AHLT using the critical values appropriate for their use as one-shot tests could be considered for such a monitoring exercise, these would not be size controlled because the overall false-positive rate (FPR), defined as the probability of at least one test in the sequence rejecting when the null was true and, hence, no bubble was present, of such a monitoring procedure would be unknown. Sequentially performing these tests in this manner would lead to an FPR that would likely be well above the nominal level at which the individual tests are performed, and would increase, other things being equal, as the monitoring horizon grew, because of the usual multiple testing problem.

In response to the multiple testing issues discussed above, Homm and Breitung (2012) introduced a cumulative sum of squares (CUSUM)-based monitoring procedure, which, under certain conditions, controls the FPR when monitoring multiple periods into the future. A limitation of their CUSUM procedure, however, is that using critical values based on asymptotic theory leads to an overly conservative test. Homm and Breitung (2012) therefore recommend using finite-sample critical values simulated from a Gaussian random walk. Although in large samples the FPR of this procedure is not dependent on the Gaussianity assumption, non-normality could have an impact with small sample sizes. Moreover, their recommended approach is based on the assumptions that the driving shocks are unconditionally homoskedastic and are serially uncorrelated. The former is especially relevant when testing for the presence of bubbles; see Harvey *et al.* (2016, p.549) who argue that '…volatility changes in innovations to price series processes could be induced by the presence of a speculative bubble, but equally it could be the case that changes in volatility occur without an explosive bubble period occurring'. They show that the ADF-based bubble detection tests of Phillips *et al.* (2011) can display severe over-rejections of the null when time-varying volatility, rather than an explosive bubble, is present in the data. As we will show in Section 4, this is also the case for the CUSUM procedure whose empirical FPR can be severely inflated in the presence of time-varying volatility. This is especially problematic for bubble detection procedures because time-varying volatility appears to be a common trait exhibited by financial time series data.

We propose a solution to the real-time inference problems outlined above using the monitoring procedure methodology recently proposed by Harvey *et al.* (2018) (HLST hereafter) for the purposes of predictive regime detection. Specifically, HLST propose a monitoring procedure for predictive behaviour that involves the sequential application of one-shot t-tests for the null of no predictive behaviour in a small number of observations applied at multiple sequential points in time over a given monitoring horizon. In the scenario considered in HLST, the null of no predictive behaviour in the monitoring period is rejected if the number of contiguous rejections signalled by

DOI: 10.1111/jtsa.12409

individual t-tests in the monitoring period exceeds some threshold value when performed at a given nominal significance level. Critical values for the individual t-tests are obtained using the sub-sampling method of Andrews (2003) and Andrews and Kim (2006), where it is assumed a training period of observations in which the null of no predictive behaviour holds is available to the practitioner. The methods of HLST allow the theoretical FPR of the procedures to be determined for any given monitoring horizon, or, equally, can be used to ensure that the FPR does not exceed a specified level by setting a maximum monitoring horizon.

Using the approach developed in HLST, we develop a monitoring procedure for the detection of asset price bubbles in which an explosive bubble is detected in the monitoring period if the number of contiguous rejections signalled by the AHLT test performed at some predetermined significance level exceeds some threshold value. We also propose a modification to the methodology of HLST in which a bubble is identified if any given test statistic calculated in the monitoring period exceeds the largest analogous sub-sample statistic calculated in the training period, thereby obviating the need to calculate a training period critical value. In line with the HLST methodology, both procedures permit calculation of the theoretical FPR at any given point in the monitoring period. The theoretical FPR for the procedure does not require us to assume that the driving shocks are homoskedastic or serially uncorrelated. We also propose a union-of-rejections approach in which an asset price bubble is signalled if either of our proposed monitoring procedures signals the presence of a bubble. Simulations show that both approaches have finite-sample empirical FPR properties that closely mimic the asymptotic results under the null of no explosivity, including cases where the series under investigation is driven by shocks that may exhibit time-varying volatility and/or serial correlation. Under the alternative hypothesis where a bubble occurs in the monitoring period, both procedures are shown to have an appealing true-positive rate (TPR), defined as the probability of correctly detecting a bubble having monitored up to a given point in the monitoring period. The procedures therefore offer the possibility of rapid detection of an emerging bubble in a real-time setting.

The remainder of this article is structured as follows. Section 2 outlines the data generating process (DGP) assumed for end-of-sample asset price bubbles. Here we also outline the sub-sample-based test of AHLT. Section 3 discusses how the AHLT test can be adapted to construct monitoring procedures for detecting asset price bubbles with a known FPR at any given point. Section 4 presents the results of finite sample simulations in which we examine the empirical FPR and TPR of our proposed monitoring procedures. Section 5 presents results from an empirical application to monthly stock market index data. Section 6 concludes this article.

2. THE MODEL AND THE AHLT TEST

It is well known in the rational bubble literature that, where bubbles are present, they should manifest explosive characteristics in prices (see Diba and Grossman (1988)). This statistical property has motivated the use of an autoregressive model that in some periods admits a unit root while in other periods exhibits explosive autoregressive behaviour; see, *inter alia*, Diba and Grossman (1988); Phillips *et al.* (2011), Homm and Breitung (2012), Phillips *et al.* (2015), Harvey *et al.* (2016), Harvey *et al.* (2017), and AHLT. Following these authors, we will consider the time series process $\{y_t\}$ generated according to the following DGP:

$$y_t = \mu + u_t \tag{1}$$

$$u_t = \begin{cases} u_{t-1} + \varepsilon_t, & t = 1, \dots, \lfloor \tau_1 T \rfloor, \\ (1+\delta)u_{t-1} + \varepsilon_t, & t = \lfloor \tau_1 T \rfloor + 1, \dots, \lfloor \tau_2 T \rfloor, \\ u_{\lfloor \tau_2 T \rfloor} + \kappa \mathbf{1}(\delta > 0)(u_{\lfloor \tau_1 T \rfloor} - u_{\lfloor \tau_2 T \rfloor}) + \varepsilon_t, & t = \lfloor \tau_2 T \rfloor + 1, \\ u_{t-1} + \varepsilon_t, & t = \lfloor \tau_2 T \rfloor + 2, \dots, T \end{cases}, \tag{2}$$

where $\mathbf{1}(.)$ denotes the indicator function, and $\lfloor \cdot \rfloor$ denotes the integer part of its argument. The driving shocks, ε_t, in (2), are assumed to be such that $\varepsilon_t = \sigma_t z_t$, where z_t is mean-zero stationary and ergodic and where, following Harvey *et al.* (2016), the volatility term σ_t satisfies $\sigma_t = \omega(t/T)$, where $\omega(\cdot) \in \mathcal{D}$ is non-stochastic and strictly positive. For $t \leq 0$, $\sigma_t \leq \breve{\sigma} < \infty$. The DGP is assumed to be initialized at $u_0 = c$, where c is some positive constant.

In the context of (1) and (2), if $\delta = 0$, then y_t admits a unit autoregressive root throughout the sample period. This forms our null hypothesis, denoted H_0. In contrast, if $\delta > 0$, y_t admits a unit autoregressive root until time $\lfloor \tau_1 T \rfloor$, after which y_t displays explosive autoregressive behaviour until time $\lfloor \tau_2 T \rfloor$. In the case where the bubble episode terminates before the end of the sample, that is, where $\tau_2 < 1$, the parameter $\kappa \in \{0, 1\}$ determines the mechanism by which the bubble terminates. Where $\kappa = 0$, the period of explosive behaviour is followed by an immediate return to autoregressive unit root behaviour. A single period correction (crash) to the pre-explosive level of the series before the return to an autoregressive unit root is obtained if $\kappa = 1$. In either of these scenarios, $\delta > 0$ forms the alternative hypothesis, which we denote as H_1^κ. For $\tau_2 = 1$, only the first two equations apply in (2) and the explosive phase is ongoing at the end of the sample. In this case, there is no distinction between $\kappa = 0$ and $\kappa = 1$, but for convenience we will still denote the alternative here as H_1^κ when $\delta > 0$.

The conditions placed on ε_t above allow for both conditional and unconditional heteroskedasticity and for stationary serial correlation in the driving shocks. The conditions placed on σ_t imply that the unconditional volatility of ε_t is bounded and displays a countable number of jumps. This allows for processes displaying (possibly) multiple one-time volatility shifts (which need not be located at the same point in the sample as the putative regime associated with bubble behaviour), polynomially (possibly piecewise) trending volatility, and smooth transition variance breaks, among others. The conventional homoskedasticity assumption, that $\sigma_t = \sigma$ for all t, is also permitted, since here $\omega(s) = \sigma$ for all s.

In developing a real-time monitoring exercise, our interest lies in the early detection of an explosive regime. That is, we wish to rapidly detect departures from the null hypothesis H_0 and make claim to have entered an explosive regime alternative, H_1^κ. We will consider $y_1, \ldots, y_{T^*}$, $T^* = \lfloor \lambda T \rfloor \leq \lfloor \tau_1 T \rfloor$ for some $\lambda \in (0, 1)$, as a training period (or training sample). The assumption that $T^* \leq \lfloor \tau_1 T \rfloor$ implies that no explosive behaviour is present in the training period.[1] We will subsequently consider monitoring for an explosive regime from some time period $T^\dagger$ onwards (with $T^\dagger > T^*$), employing the training period data in a calibration role.

Our approach is based on the test suggested by AHLT. This is a heteroskedasticity-robust sub-sample test statistic for upward explosive behaviour, which in AHLT was proposed as a one-shot test for an end-of-sample financial bubble. In generic notation, this sub-sample statistic is constructed from a user-chosen finite-length window of m first differences $\Delta y_{e-m+1}, \Delta y_{e-m+2}, \ldots, \Delta y_e$ (where e indicates the most recent value of y_t used in the statistic's construction) and is given by

$$S_{e,m} := \frac{\sum_{t=e-m+1}^{e} (t - e + m) \Delta y_t}{\sqrt{\sum_{t=e-m+1}^{e} \{(t - e + m) \Delta y_t\}^2}}.$$

The one-shot test in AHLT is based on the use of the sub-sampling method for estimating critical values developed in a general context in Andrews (2003) and applied to the case of tests for end-of-sample breakdown of co-integration in Andrews and Kim (2006). This approach involves calculating analogous sub-sample statistics for all possible date windows within the training period (over which H_0 is assumed to hold): that is, calculating $S_{e,m}$ for $e = m + 1, \ldots, T^*$, and then calculating an upper-tail empirical critical value from these statistics for a significance level π, say, which we denote by cv_π.[2] It follows from Andrews (2003) and Andrews and Kim (2006) that cv_π is a consistent estimate for the true π significance level critical value as $T \to \infty$. The AHLT test statistic is then $S_{T^*+m,m}$, that is, $S_{e,m}$ applied to the first available window of m periods that does not include data from the training period. The one-shot test simply compares $S_{T^*+m,m}$ with the critical value cv_π , and under H_0 has an FPR

[1] We will investigate the impact that violations of this maintained assumption, such that explosive autoregressive behaviour is present in the training period, have on our proposed monitoring procedures in Section 4.3. In practice, we recommend using a historical bubble detection test, such as the wild bootstrap implementation of the Phillips *et al.* (2011) test proposed in Harvey *et al.* (2016), to test the null hypothesis that no bubbles are present in the chosen training period. Any bubble small enough not to be detected by these tests is unlikely to have a large impact on the detection properties of our monitoring procedure. Where such tests detect a bubble, the training period could simply be redefined to exclude the detected bubble periods.

[2] Note that cv_π can be defined such that $cv_\pi = S_{(\lfloor (1-\pi)(T^*-m) \rfloor)}$, where $S_{(j)}$, $j = 1, \ldots, T^* - m$ are the ascending order statistics of $S_{e,m}$, $e = m + 1, \ldots, T^*$.

DOI: 10.1111/jtsa.12409

of π for large T without requiring knowledge of the joint null distribution of the $S_{e,m}$ statistics. Note also that under H_1^κ, the test will reject with probability equal to π if the explosive regime has not commenced within the testing window, that is, $T^* + m \leq \lfloor \tau_1 T \rfloor$.

3. REAL-TIME MONITORING PROCEDURES

Our goal in this article is to develop a real-time monitoring procedure for detecting the emergence of an explosive bubble, and hence we move beyond the one-shot testing framework to consider a *sequence* of $S_{e,m}$ statistics. Suppose we wish to begin monitoring at the present time period, say $t = T^\dagger$. We would then set the training sample end-date to be $T^* = T^\dagger - m$, allowing the calculation of the first monitoring statistic $S_{e,m}$ with $e = T^\dagger = T^* + m$, which uses data from $T^* + 1$ to $T^\dagger = T^* + m$. In the next period ($t = T^\dagger + 1$), the second monitoring statistic $S_{e,m}$ with $e = T^\dagger + 1 = T^* + m + 1$ can be calculated (which uses data from $T^* + 2$ to $T^\dagger + 1 = T^* + m + 1$). Now suppose that the monitoring continues in this manner; then at any given point in the monitoring period $t = T'$, the sequence of monitoring statistics $S_{e,m}$ for $e = T^* + m, \ldots, T'$ will have been calculated. Of course, if one were to conduct the one-shot test at a marginal π significance level repeatedly through the monitoring period, that is, a detection procedure based on sequentially comparing $S_{e,m}$ with cv_π, $e = T^* + m, \ldots, T'$, then the FPR of the detection procedure at time period $t = T'$ (that is, the probability of falsely detecting a bubble having monitored up to time period $t = T'$) would exceed π due to the multiple testing involved, increasing monotonically with T'. Moreover, for any given T', the precise asymptotic FPR associated with such a procedure cannot be ascertained without knowing the joint null distribution of the $S_{e,m}$ statistics. Instead, we adapt alternative procedures recently developed by HLST (in the context of predictive regression testing) to allow real-time monitoring for a bubble while being able to determine the FPR for a given value of T', or, equivalently, to determine the appropriate time period $t = T'$ at which the FPR reaches a predetermined desired level.

To begin, we again consider the training period statistics $S_{e,m}$, $e = m + 1, \ldots, T^*$; but now rather than obtaining an upper-tail empirical critical value cv_π, we instead consider the *maximum* $S_{e,m}$ statistic, that is, $S^*_{\max} := \max_{e \in [m+1, T^*]} S_{e,m}$. In the spirit of Section 3 of HLST, we can then devise a real-time monitoring procedure based on comparing the $S_{e,m}$ statistics, calculated in real time over the monitoring period, with the training period maximum statistic $S^*_{\max}$. That is, in the first monitoring time period $T^\dagger = T^* + m$, we calculate $S_{T^*+m,m}$ and conclude that a bubble is detected, that is, H_0 is rejected, if $S_{T^*+m,m} > S^*_{\max}$, at which point we would terminate the monitoring procedure. If H_0 is not rejected, in the next time period we continue monitoring and calculate $S_{T^*+m+1,m}$, rejecting H_0 and terminating the procedure if $S_{T^*+m+1,m} > S^*_{\max}$. Real-time monitoring continues in this manner, with a bubble detected at time T' if $S_{T',m} > S^*_{\max}$. So, in general, the monitoring procedure is terminated at the first point where $S_{e,m}$, $e = T^* + m, \ldots$ exceeds $S^*_{\max}$. Of course, continued monitoring in this way indefinitely will eventually lead to a rejection of H_0 even when the null is true. Hence for such a procedure to be statistically rigorous, it is critical to understand the FPR of such a procedure at each monitoring point.

Consider an arbitrary point in time during the monitoring period $t = T'$. Defining $S'_{\max} := \max_{e \in [T^*+m, T']} S_{e,m}$, the procedure described above can equivalently be expressed as

$$\text{Reject } H_0 \text{ at time } t = T' \text{ if } S'_{\max} > S^*_{\max}$$

with the monitoring terminating at time $t = T'$ if a rejection occurs. HLST present uniformity arguments relating to the location of the maximum value of $S_{e,m}$ to show that, under H_0, the limiting probability that the maximum $S_{e,m}$ statistic lies in the monitoring period, as opposed to the training period, is simply the limit ratio of the number of test statistics calculated in the monitoring period $(T' - T^* - m + 1)$ to the number of test statistics calculated across the two periods combined $((T' - T^* - m + 1) + (T^* - m) = T' - 2m + 1)$. So, on defining

$$\alpha = \lim_{T^*, T' \to \infty} \frac{T' - T^* - m + 1}{T' - 2m + 1} \tag{3}$$

DOI: 10.1111/jtsa.12409

it follows that, under H_0

$$\lim_{T^*,T'\to\infty} \Pr\left(S'_{\max} > S^*_{\max}\right) = \alpha. \tag{4}$$

The asymptotic FPR for the monitoring procedure run up to time $t = T'$ is therefore given by α, and in practice for a given T', T^*, and m, we would approximate α using

$$\alpha \approx \frac{T' - T^* - m + 1}{T' - 2m + 1}. \tag{5}$$

Hence, if a bubble is detected at some time $t = T'$, the corresponding FPR can be immediately computed. In what follows, we will denote this real-time bubble detection procedure by MAX_m.

As HLST show, the function (5) is monotonically increasing in T', and hence the FPR increases as monitoring continues. In practice, it may be desirable to specify a monitoring end-point which ensures that the FPR does not exceed a certain predetermined desired level. On rearranging (5), we can obtain T' as a function of α:

$$T' \approx \frac{T^* + m - 1 - \alpha(2m - 1)}{1 - \alpha}. \tag{6}$$

Hence, for a given choice of α, we can identify the monitoring time period at which the FPR will reach the level α. If the intention is to ensure that the FPR does not exceed this predetermined level, (6) can be used to calculate the appropriate monitoring end-point.

In addition to the MAX_m procedure, we also consider an alternative real-time monitoring procedure related to the method proposed in Section 4 of HLST. Following the approach of HLST, we consider comparison of the $S_{e,m}$ statistics over both the training and monitoring periods with the critical value cv_π, which is obtained from the training period as in AHLT (see Section 2). First, let $R_e := \mathbf{1}(S_{e,m} > cv_\pi)$ record whether or not a statistic for a given value of e exceeds cv_π, and define the following measure over $e = L$ to $e = U$ with $U \geq L$:

$$R(L, U) := (U - L + 1) \prod_{e=L}^{U} R_e.$$

If $S_{e,m}$ exceeds cv_π for all $e = L, \ldots, U$, then $R(L, U) = U - L + 1$ represents the length of a sequence of contiguous exceedances; otherwise, $R(L, U) = 0$. Next, we define the longest contiguous sequence of exceedances in the training period as

$$m^* := \max_{L,U\in[m+1,T^*]} R(L, U).$$

Then, in the real-time monitoring period, at a given monitoring time $t = T'$, we define the longest contiguous sequence of exceedances in the monitoring period thus far (that is, up to $t = T'$) as

$$m' := \max_{L,U\in[T^*+m,T']} R(L, U).$$

The real-time detection procedure is then given by

$$\text{Reject } H_0 \text{ at time } t = T' \text{ if } m' > m^*$$

with the monitoring terminating at time $t = T'$ if a rejection occurs. Hence a bubble is detected if there exists a longer continuous sequence of exceedances in the monitoring period than is obtained in the training period.

DOI: 10.1111/jtsa.12409

Using the uniformity arguments outlined in HLST pertaining to the location of the longest contiguous sequence of exceedances, we can show that, for α defined by (3), the analogous result to (4) is that, under H_0

$$\lim_{T^*,T'\to\infty} \Pr\left(m' > m^*\right) = \alpha.$$

The results in (5) and (6) pertaining to the relationships between T' and α also hold here, allowing practical control of the procedure's FPR. In what follows, we will denote this second procedure by SEQ_m.[3]

Note that for SEQ_m, the first time period at which it would be possible to reject H_0 is $t = T^\dagger + m^*$, because this is the first occasion in the monitoring period where $R(L, U)$ can exceed m^*. However, for MAX_m, the first time period at which it would be possible to reject H_0 is $t = T^\dagger$, that is, the period at which monitoring begins, which is m^* periods earlier than for SEQ_m, giving the potential for MAX_m to deliver an earlier detection outcome under H_1^κ if the bubble originates very early in the monitoring period.

One interpretation of the MAX_m procedure is that it is a special case of SEQ_m where we set $cv_\pi := \max_{e\in[m+1,T^*]} S_{e,m}$ (the largest order statistic in the training period). Then m^* is by definition zero (as no $S_{e,m}$ exceeds cv_π in the training period), and we detect an explosive regime in the monitoring period if we obtain an $S_{e,m}$ statistic that exceeds $\max_{e\in[m+1,T^*]} S_{e,m}$, which can be seen as a monitoring period 'contiguous exceedance' of 1, which is greater than $m^* = 0$ (see the decision rule for SEQ_m).

As an interesting side issue, suppose we have obtained no rejection of H_0 up to some time period $T^{**} > T^* + m$. We might then consider 'resetting' the monitoring procedure by updating the training period from $y_1, \ldots, y_{T^*}$ to $y_1, \ldots, y_{T^{**}}$. To evaluate the effect of this, consider the MAX_m procedure. For any $T' > T^{**}+m$, our new decision rule would be to reject H_0 if $S'_{\max} > S^{**}_{\max}$, where $S^{**}_{\max} := \max_{e\in[m+1,T^{**}]} S_{e,m}$. However, since we have found no rejections of H_0 up to time T^{**}, that is, $S_{e,m} < S^*_{\max}$ for all $e \in [T^* + m, T^{**}]$, it follows that $\max_{e\in[m+1,T^*]\cup[T^*+m,T^{**}]} S_{e,m} = S^*_{\max}$, and hence asymptotically (given finite m), $S^{**}_{\max} = S^*_{\max}$. Hence, trivially, in the limit we can write

$$\Pr(S'_{\max} > S^{**}_{\max}) = \Pr(S'_{\max} > S^*_{\max})$$

so that the rejection probability at time $t = T'$ associated with the original training period and the updated training period procedures are identical. The practical implication of this result is that, asymptotically, both the FPR under the null and the TPR under the alternative are unaffected by updating the training period, and hence there is no virtue in updating the training period in any attempt to improve the FPR or TPR of the monitoring procedure. Similar arguments can be made for the SEQ_m procedure in the limit also.

Finally, our discussion in this section assumes, for simplicity, that there is no separation between the data period used for the training period and the data used for monitoring, with the former spanning $t = 1, \ldots, T^*$ and the latter starting at $t = T^\dagger - m + 1 = T^* + 1$. More generally, the last time period included in the training sample could be $T^* - k$ for some $k > 0$, thereby allowing for a separation between the training period and the start of the monitoring period. This might be relevant in cases where a historical bubble episode was thought to have occurred towards the end of the training period (see footnote 1). In this case, the expressions for α and T' in (5) and (6) become

$$\alpha \approx \frac{T' - T^* - m + 1}{T' - 2m + 1 - k}$$

$$T' \approx \frac{T^* + m - 1 - \alpha(2m - 1 + k)}{1 - \alpha}.$$

Note that, relative to $k = 0$, the FPR α is now increased for a given T'.

[3] Notice that the dependence of the procedure on the choice of significance level π at which the individual $S_{e,m}$ tests are conducted is implicit. The value of π influences the lengths of the contiguous exceedances: the larger the value of π, the smaller the value of cv_π, and the longer we would expect the sequences of contiguous exceedances to be. Other things being equal, this will have the effect of increasing both m' and m^*.

DOI: 10.1111/jtsa.12409

4. FINITE-SAMPLE SIMULATIONS

We now examine the finite-sample properties of our proposed MAX_m and SEQ_m monitoring procedures. The results we present suggest that the TPR of the MAX_m procedure is generally higher when a bubble is beginning to emerge, whereas it is higher for the SEQ_m procedure as we move further into the bubble phase. Therefore, in addition to the MAX_m and SEQ_m procedures, we also report the properties of a 'union-of-rejections procedure', denoted U_m, in which a bubble episode is signalled if either the MAX_m or SEQ_m procedures reject H_0, potentially allowing us to exploit the favourable TPR properties of the MAX_m and SEQ_m procedures for detecting shorter or longer duration bubbles respectively. The U_m procedure will clearly, however, not have its FPR controlled in the same way as the two constituent procedures. If the MAX_m and SEQ_m procedures are both performed with an FPR of α, then the FPR of the U_m monitoring procedure will be no smaller than α. However, as MAX_m and SEQ_m make use of the same underlying test statistics, they are not independent of each other and we will show that the increase in the empirical FPR from using a union-of-rejections approach is relatively modest.

We also compare the performance of our monitoring procedures to the CUSUM monitoring procedure of Homm and Breitung (2012). Assuming a training period of length T^*, they propose using the following monitoring statistic:

$$S^t_{T^*} := \frac{1}{\hat{\sigma}_t}\left(y_t - y_{T^*}\right), \text{ with } \hat{\sigma}_t := \sqrt{(t-1)^{-1}\sum_{j=1}^{t}(\Delta y_j)^2},$$

where $t > T^*$ is the monitoring observation. Homm and Breitung (2012) show that, under the assumptions of serially uncorrelated and homoskedastic ε_t, if $S^t_{T^*}$ is computed multiple times at dates $T^*+1, \ldots, E$, then under H_0 for any $k > 1$

$$\lim_{T^*\to\infty} \Pr\left(|S^t_{T^*}| > \left(c_t \times \sqrt{t}\right) \text{ for some } t \in \{T^*+1, \ldots, kT^*\}\right) < \exp\left(-b_\alpha/2\right), \tag{7}$$

where $c_t := \sqrt{b_\alpha + log(t/T^*)}$. The monitoring procedure proposed in Homm and Breitung (2012) rejects H_0 if $S^t_{T^*} > c_t\sqrt{(t)}$ for some $t > T^*$. For a test performed at the nominal asymptotic significance level $\alpha = 0.05$, for instance, the value of b_α used to compute c_t is equal to 4.6.

Here and throughout this section, we generate data according to (1)–(2) with $T = 300$ and $T^* + m = 220$, and examine the behaviour of our proposed monitoring procedures using window widths of $m = 5, 10, 15$. We set $\mu = 0$ (all procedures are invariant to μ under H_0 and H_1^κ) and set $u_0 = 100$ so that under H_1^κ, the bubbles generated are generally upwardly explosive (all procedures are invariant to u_0 under H_0).

4.1. Empirical FPR

We begin by analysing the empirical FPR of the monitoring procedures for data generated under the null hypothesis, H_0. We will consider the cases where ε_t is a Gaussian white noise process in Section 4.1.1, displays time-varying volatility in Section 4.1.2, and is serially correlated in Section 4.1.3. We assume a common monitoring start date for all procedures of $T^* + m = 220$ and treat the sample $t = 1, \ldots, T^* + m - 1$ as the training period for the CUSUM procedure, with the training period for the MAX_m, SEQ_m, and U_m procedures given by the sample $t = 1, \ldots, T^*$. Homm and Breitung (2012) note that choosing the critical value, b_α, for the CUSUM monitoring procedure according to (7) can lead to a very conservative test and recommend using finite-sample critical values based on simulated data generated under H_0 with Gaussian white noise innovations. As such, in the simulations that follow, we select finite-sample critical values for the CUSUM monitoring procedure such that the empirical FPR of the CUSUM procedure is equal to the theoretical FPR of MAX_m and SEQ_m when the latter procedures have a theoretical FPR of 0.10 when the data are generated under the null hypothesis with $NIID(0, 1)$ innovations.

DOI: 10.1111/jtsa.12409

4.1.1. Gaussian White Noise Innovations

Figure 1(a) reports the empirical FPR of the three proposed procedures as a function of the monitoring period when $\varepsilon_t \sim NIID(0,1)$ throughout the sample, with each point on the figure representing the empirical FPR of a particular monitoring procedure if run from time $t = T^* + m$ to $t = T'$. Also plotted on the figure is the theoretical FPR of MAX_m and SEQ_m run up to time $t = T'$, calculated from (3). As expected, the empirical FPRs of the monitoring procedures all closely track the theoretical FPR, with the empirical FPR of the MAX_m procedure tracking the theoretical FPR the closest. The empirical FPR of SEQ_m is slightly lower than the level implied by theory. Interestingly, the empirical FPR of U_m is not much higher than the theoretical FPR of the two monitoring procedures used in its construction, which is likely due to the high degree of correlation between MAX_m and SEQ_m. The empirical FPR of the CUSUM procedure is lower than that of our proposed procedures for short monitoring horizons and greater than that of all but U_m for longer monitoring periods. By construction, as detailed above, the empirical FPR of the CUSUM procedure is set equal to the theoretical FPR of MAX_m and SEQ_m when the latter is equal to 0.10.

4.1.2. Time-Varying Volatility

We next examine the empirical FPR of our proposed monitoring procedures and of the CUSUM procedure in the case where a structural change in the volatility of the shocks, ε_t, occurs either in the training period or in the monitoring period. While the procedures developed in this article are (asymptotically) robust to heteroskedasticity of the form specified in Section 2, the CUSUM-based monitoring procedure of Homm and Breitung (2012) is based on the assumption that ε_t is homoskedastic, as noted above.

We consider first the case where ε_t follows a time-varying GARCH(1,1) process. Specifically, we generate data under the null hypothesis with the shocks generated as $\varepsilon_t = h_t^{1/2} v_t$, where $v_t \sim NIID(0,1)$ and $h_t = 1.00 + 0.05\varepsilon_{t-1}^2 + \beta_t h_{t-1}$. We examine the empirical FPR of the various monitoring procedures in the case where a switch to a higher volatility regime occurs in the monitoring period, with the time-varying parameter β_t satisfying[4]

$$\beta_t = \begin{cases} 0.64, & t = 1, \ldots, T^* + m - 1, \\ 0.95, & t = T^* + m, \ldots, T. \end{cases}$$

Figure 1(b) reports the empirical FPR of all procedures when ε_t follows this time-varying GARCH(1,1) process. The robustness of our proposed monitoring procedures to this pattern of time-varying volatility is clearly demonstrated in the results, with the empirical FPRs of MAX_m, SEQ_m, and U_m being almost identical to those observed for the corresponding homoskedastic Gaussian case in Figure 1(a). In contrast, the empirical FPR for the CUSUM procedure is much higher at all monitoring points than was seen in the homoskedastic case.

We next investigate the empirical FPR of the various monitoring procedures when a one-time shift in unconditional volatility occurs in either the monitoring period or the training period. To that end, we generated data under the null hypothesis with the shocks such that $\varepsilon_t \sim NIID(0, \sigma_t^2)$, where

$$\sigma_t = \begin{cases} \sigma_1, & t = 1, \ldots, t_v, \\ \sigma_2, & t = t_v + 1, \ldots, T. \end{cases}$$

We first examine the case where a shift in volatility occurs at the start of the monitoring period by setting $t_v = 219$. Figure 2 reports the empirical FPR of the various monitoring procedures for the cases of an upward shift in volatility at the start of the monitoring period ($\sigma_1 = 1, \sigma_2 = 3$) and of a downward shift in volatility at the start of the monitoring period ($\sigma_1 = 3, \sigma_2 = 1$). The results, again, show the robustness of MAX_m, SEQ_m, and U_m to shifts in volatility, in that the empirical FPR of each of these procedures is almost identical to the corresponding

[4] To control for dependence on initialization effects, v_t was generated for $t = -299, \ldots, T$ and the conditional variance h_t was initialized at its unconditional value (when $\beta_t = 0.64$) at time $t = -299$, with $\beta_t = 0.64$ for $t = -299, \ldots, 0$. The first 300 observations on ε_t were then discarded.

DOI: 10.1111/jtsa.12409

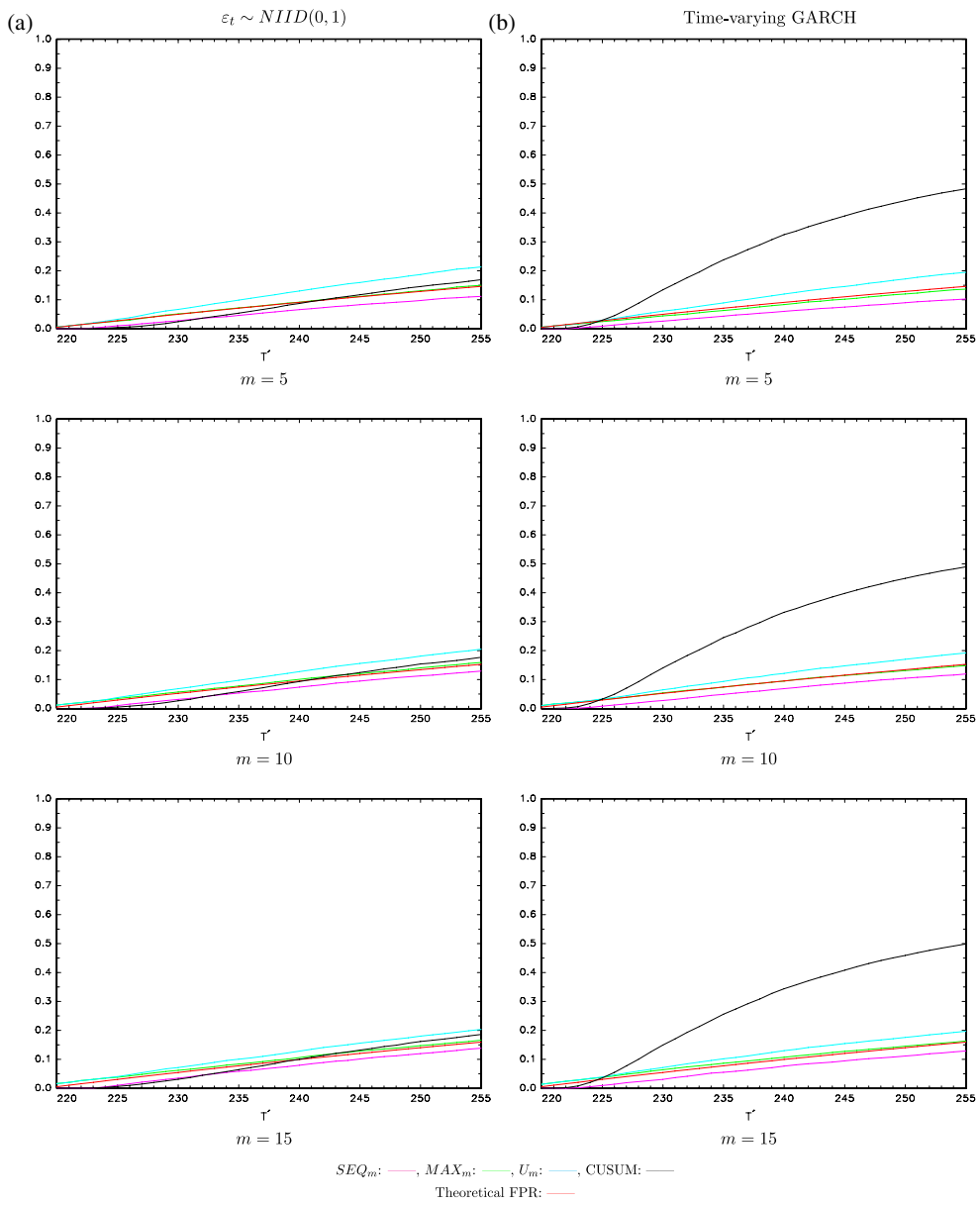


Figure 1. False positive rate—Gaussian white noise and time-varying GARCH. (a) $\epsilon \sim NIID(0,1)$. (b) Time-varying GARCH [Color figure can be viewed at wileyonlinelibrary.com]

empirical FPR seen for the homoskedastic case. The empirical FPR of the CUSUM procedure is, in contrast, markedly different to the homoskedastic case. In particular, it is larger (smaller) than in the homoskedastic case for an upward (downward) volatility shift at the beginning of the monitoring period.

Figure 3 reports corresponding results for the case where the shift in volatility occurs in the training period, with $t_v = 110$. Results are again reported for an upward shift in volatility ($\sigma_1 = 1, \sigma_2 = 3$) and a downward shift in volatility ($\sigma_1 = 3, \sigma_2 = 1$). Again, the robustness of our monitoring procedures to time-varying volatility is clearly demonstrated, with the empirical FPRs being almost identical to the homoskedastic case once again. The empirical

DOI: 10.1111/jtsa.12409

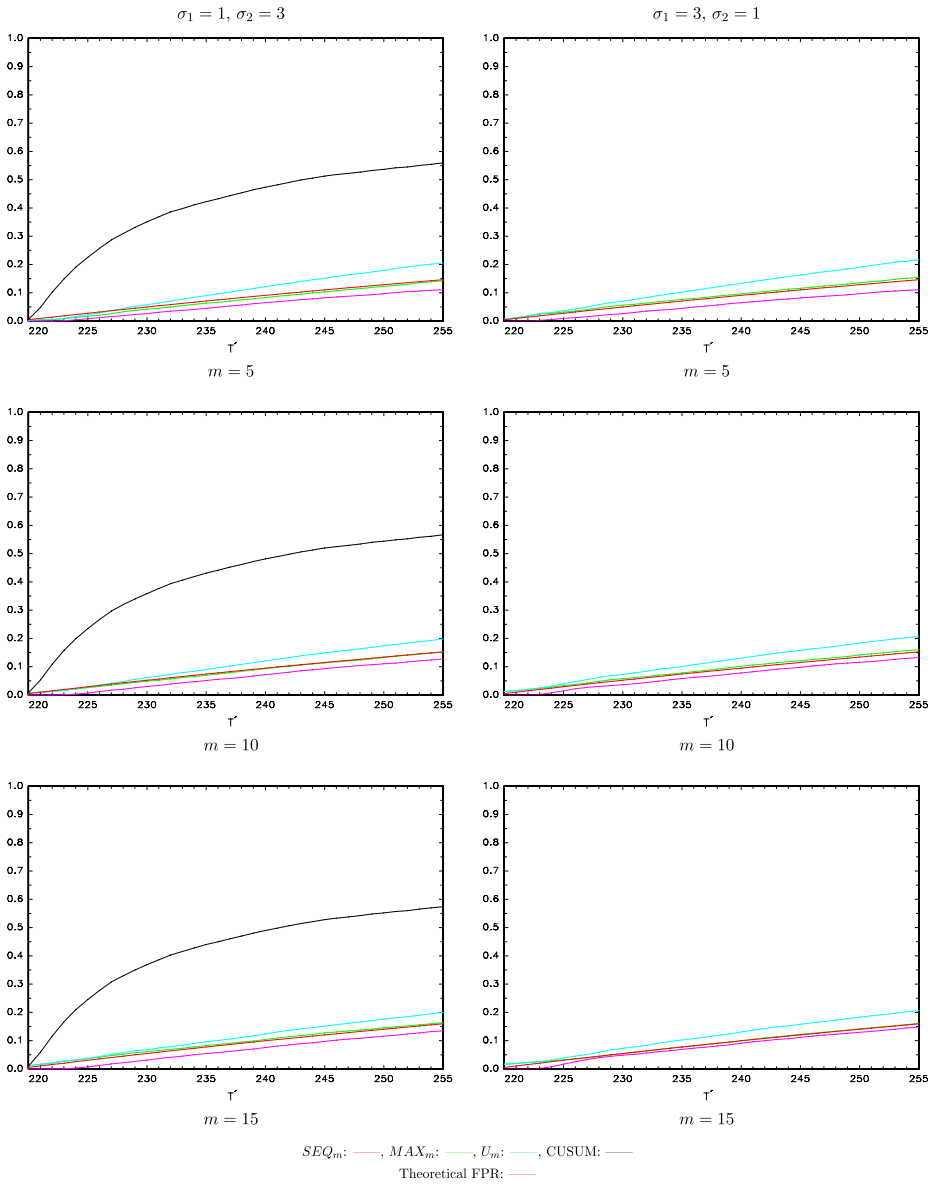


Figure 2. False positive rate—volatility shift at start of monitoring period; volatility shifts from σ_1 to σ_2 at $t_v = 219$ [Color figure can be viewed at wileyonlinelibrary.com]

FPR of the CUSUM is, again, impacted by the volatility shift, being higher (lower) than the homoskedastic case for an upward (downward) shift in volatility, although a comparison with Figure 2 shows that the impact of a volatility shift in the training period is less drastic than a shift in volatility at the start of the monitoring period.

4.1.3. Serial Correlation

The theoretical FPR of the monitoring procedures we have developed in this article is unaffected by serial correlation in the shocks, ε_t, in contrast to the CUSUM-based monitoring procedure of Homm and Breitung (2012). To investigate the impact of serial correlation on the empirical FPR of the various procedures, we generated ε_t according to the MA(1) process $\varepsilon_t = v_t - \theta v_{t-1}$, with $v_t \sim NIID(0, 1)$ and initialized at $v_0 = 0$. Results for $\theta = \pm 0.5$

DOI: 10.1111/jtsa.12409

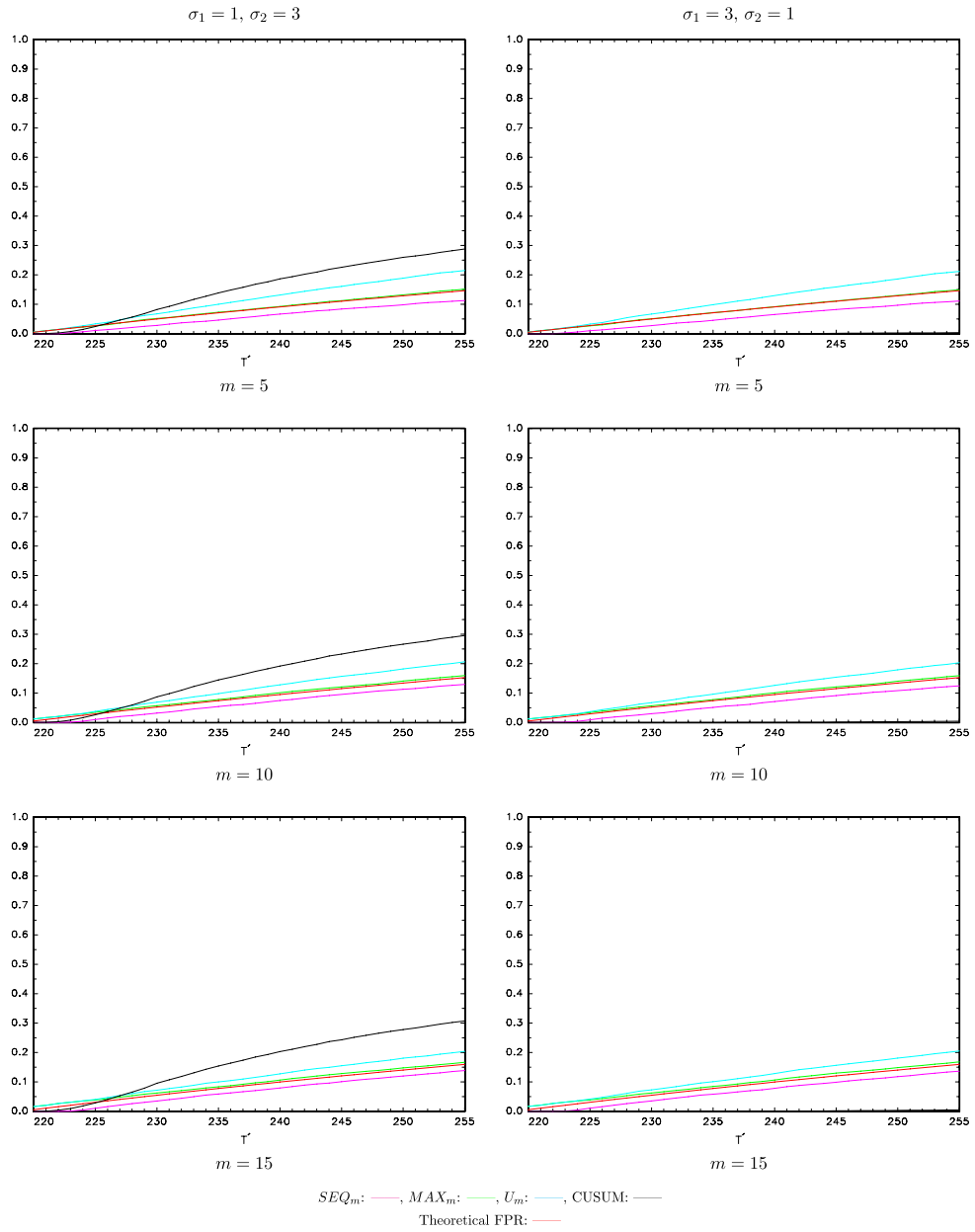


Figure 3. False positive rate—volatility shift in training period; volatility shifts from σ_1 to σ_2 at $t_v = 120$ [Color figure can be viewed at wileyonlinelibrary.com]

are reported in Figure 4. It can be seen from these results that the empirical FPRs of our proposed procedures are robust to serial correlation, with the empirical FPR of MAX_m, SEQ_m, and U_m being near identical to those seen in Figure 1(a) for the case where $\varepsilon_t \sim NIID(0, 1)$. In contrast, the empirical FPR of the CUSUM procedure is severely impacted by the presence of serial correlation, being inflated relative to the baseline case of no serial correlation when $\theta = -0.5$, and reduced to almost zero for all monitoring periods when $\theta = 0.5$. Depending on the pattern of serial correlation in the data, the CUSUM procedure is therefore more likely to spuriously indicate the presence of a bubble when $\theta < 0$, and less likely to detect a bubble when it is present when $\theta > 0$.

DOI: 10.1111/jtsa.12409

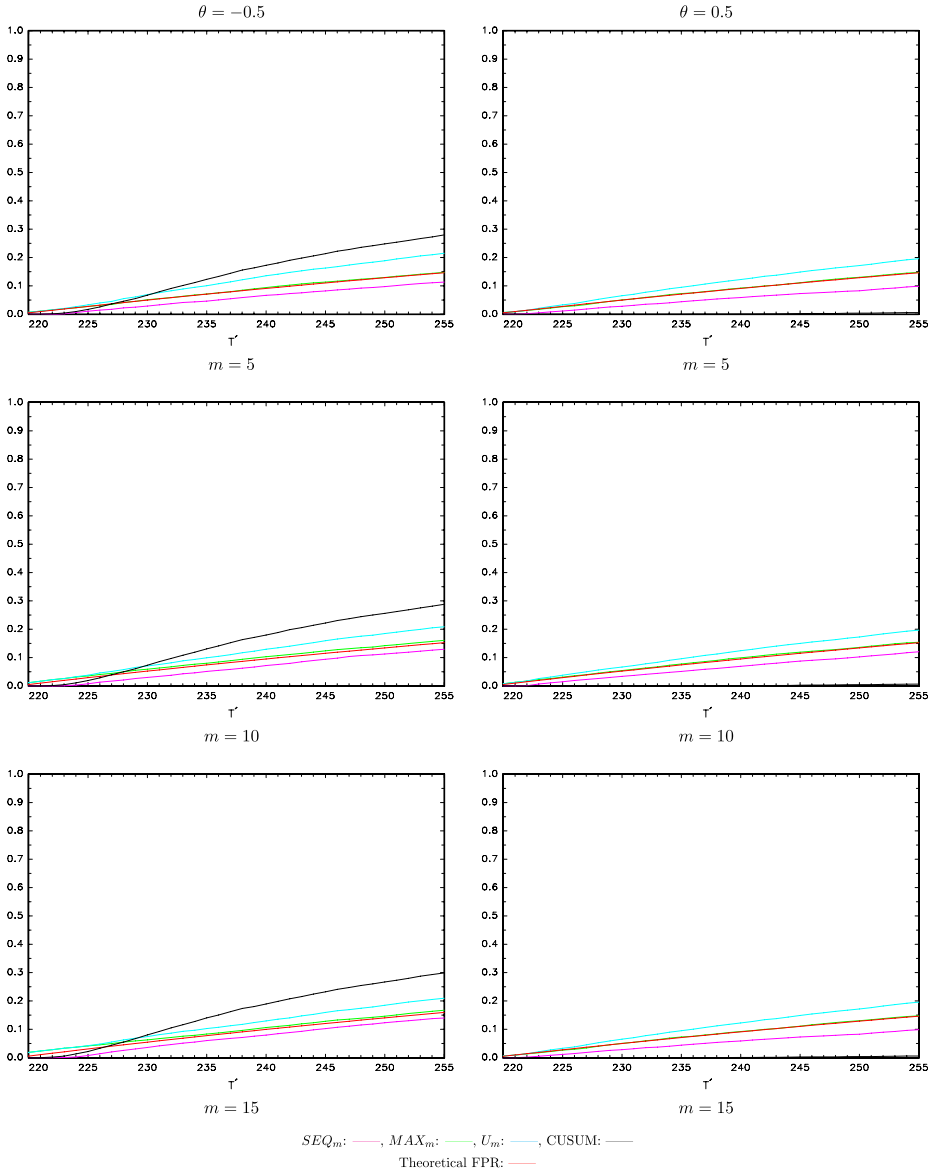


Figure 4. False positive rate—MA innovations; $\varepsilon_t = v_t - \theta v_{t-1}, v_t \sim NIID(0,1)$ [Color figure can be viewed at wileyonlinelibrary.com]

The results in this section have highlighted the lack of robustness of the empirical FPR of the CUSUM procedure to both time-varying volatility and serial correlation. As a result, we will not consider it further in this article.

4.2. Empirical TPR

We now proceed to examine the TPR (the probability of correctly detecting a bubble having monitored up to a given time period $t = T'$) of our proposed monitoring procedures to detect a bubble episode emerging in the monitoring period. We generate $\varepsilon_t \sim NIID(0,1)$ throughout, and initially examine a bubble of length 10 observations, generating data according to (1)–(2) with $\delta > 0$, $\lfloor \tau_1 T \rfloor = 230$ and $\lfloor \tau_2 T \rfloor = 240$.

DOI: 10.1111/jtsa.12409

Figure 5 reports the empirical TPRs when $\delta = 0.01$ for both the case where the series reverts to a unit root process following the bubble, that is, H_1^0, and the case where the bubble collapses, that is, H_1^1. The results show that, for a given value of m, MAX_m has a higher TPR than SEQ_m when monitoring close to the bubble inception date, with the TPR differential being most pronounced for larger values of m. As noted previously, the rationale behind this result is that the earliest possible date that the number of contiguous exceedances in the monitoring period for the SEQ_m procedure could exceed m^* is at time $t = T^* + m + m^*$, the point at which $m^* + 1$ test statistics have been calculated in the monitoring period. In contrast, MAX_m has the potential to reject the null as early as $t = T^* + m$. As the bubble episode continues, the difference in TPR between the two procedures becomes less

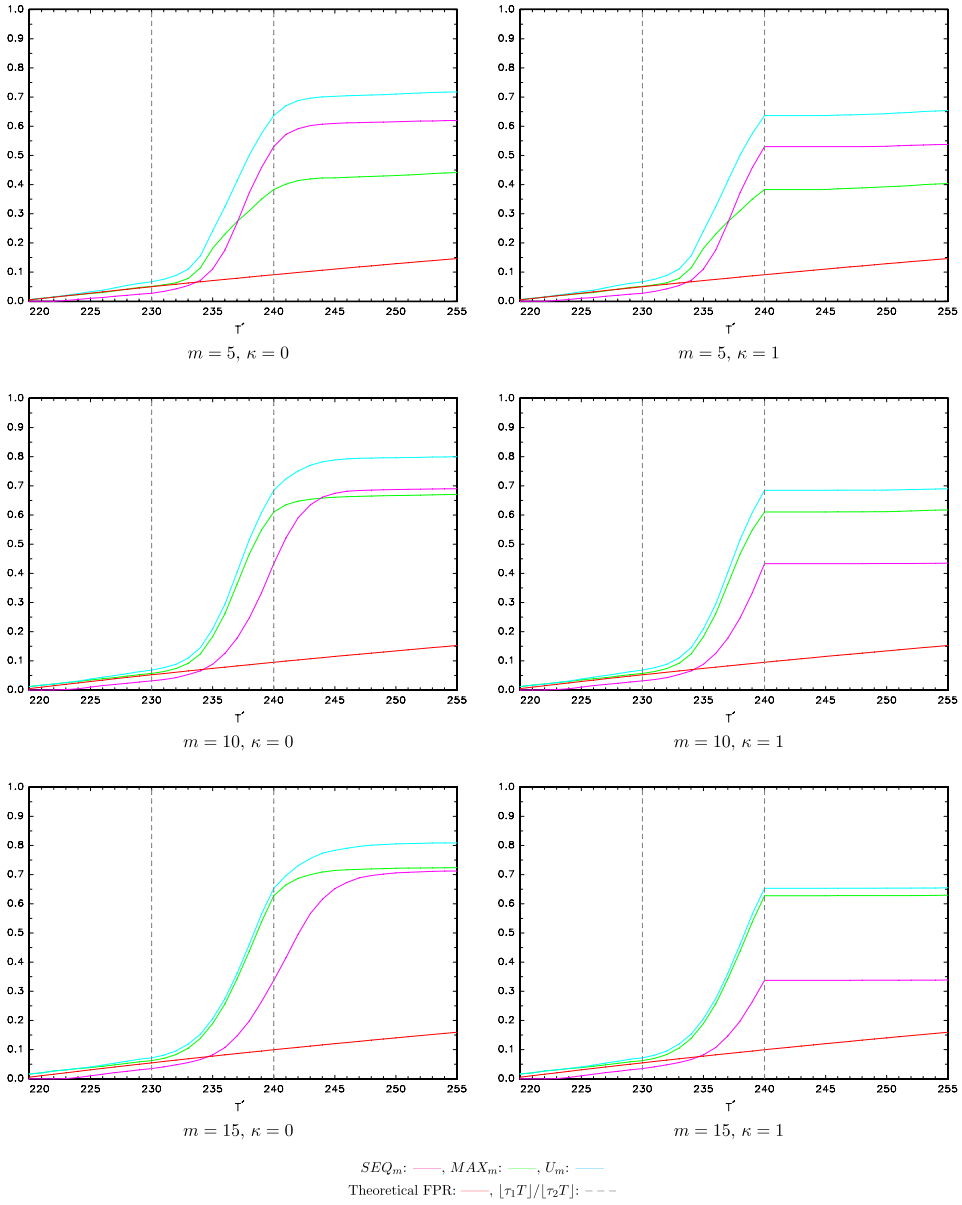


Figure 5. True positive rate $\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor = 10$, $\delta = 0.010$ [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12409

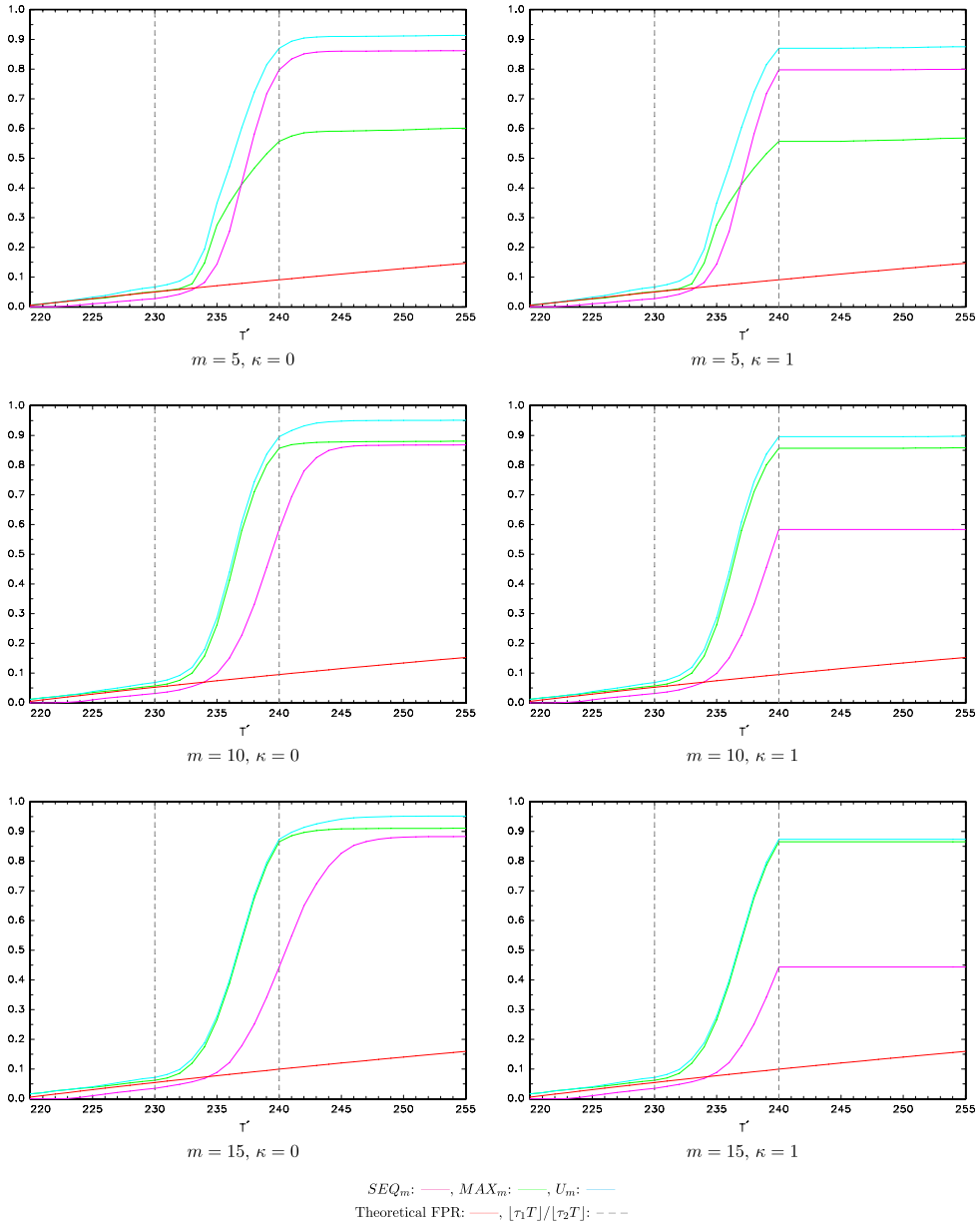


Figure 6. True positive rate $\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor = 10$, $\delta = 0.015$ [Color figure can be viewed at wileyonlinelibrary.com]

pronounced, with SEQ_m eventually displaying the higher TPR nearer to the termination of the bubble for $m = 5$. The U_m procedure has a uniformly higher TPR than both MAX_m and SEQ_m (by construction) and is well equipped to detect the bubble both when it is close to inception or termination. Under H_1^0, where the series reverts to a unit root process without collapse following the bubble, the TPR of each individual procedure to detect the bubble still increases with the monitoring horizon, even when evaluating the procedure at $T' > \lfloor \tau_2 T \rfloor$, as the $S_{e,m}$ test statistic evaluated at time $e = T'$ will still contain explosive observations when $\lfloor \tau_2 T \rfloor < T' \leq \lfloor \tau_2 T \rfloor + m - 1$. Under H_1^1, the collapse ensures that there are no further rejections signalled by any procedure after $T' = \lfloor \tau_2 T \rfloor$, as indicated by the curves in the figures flattening out to horizontal lines for $T' > \lfloor \tau_2 T \rfloor$. This is due to the fact that the constituent

$S_{e,m}$ test statistics calculated for $e > \lfloor \tau_2 T \rfloor$ will be computed using the large negative value of $\Delta y_{\lfloor \tau_2 T \rfloor + 1}$ caused by the collapse, giving them a very small probability of exceeding the relevant critical value.

With regard to the choice of m, a trade-off clearly exists. Other things being equal, the TPRs of the procedures early in the bubble regime are higher for smaller values of m compared to larger values of m, so that bubbles are detected more rapidly for smaller m. On the other hand, the eventual TPRs that the procedures attain by the end of the bubble regime are higher for larger m.

Figure 6 reports the results for a bubble again of length 10 observations with $\delta > 0$, $\lfloor \tau_1 T \rfloor = 230$ and $\lfloor \tau_2 T \rfloor = 240$, but now with the bubble driven by an explosive offset of $\delta = 0.015$. The TPR of each procedure at any given time $t = T'$ is, as we would expect, larger than when $\delta = 0.01$. We observe a broadly similar pattern with

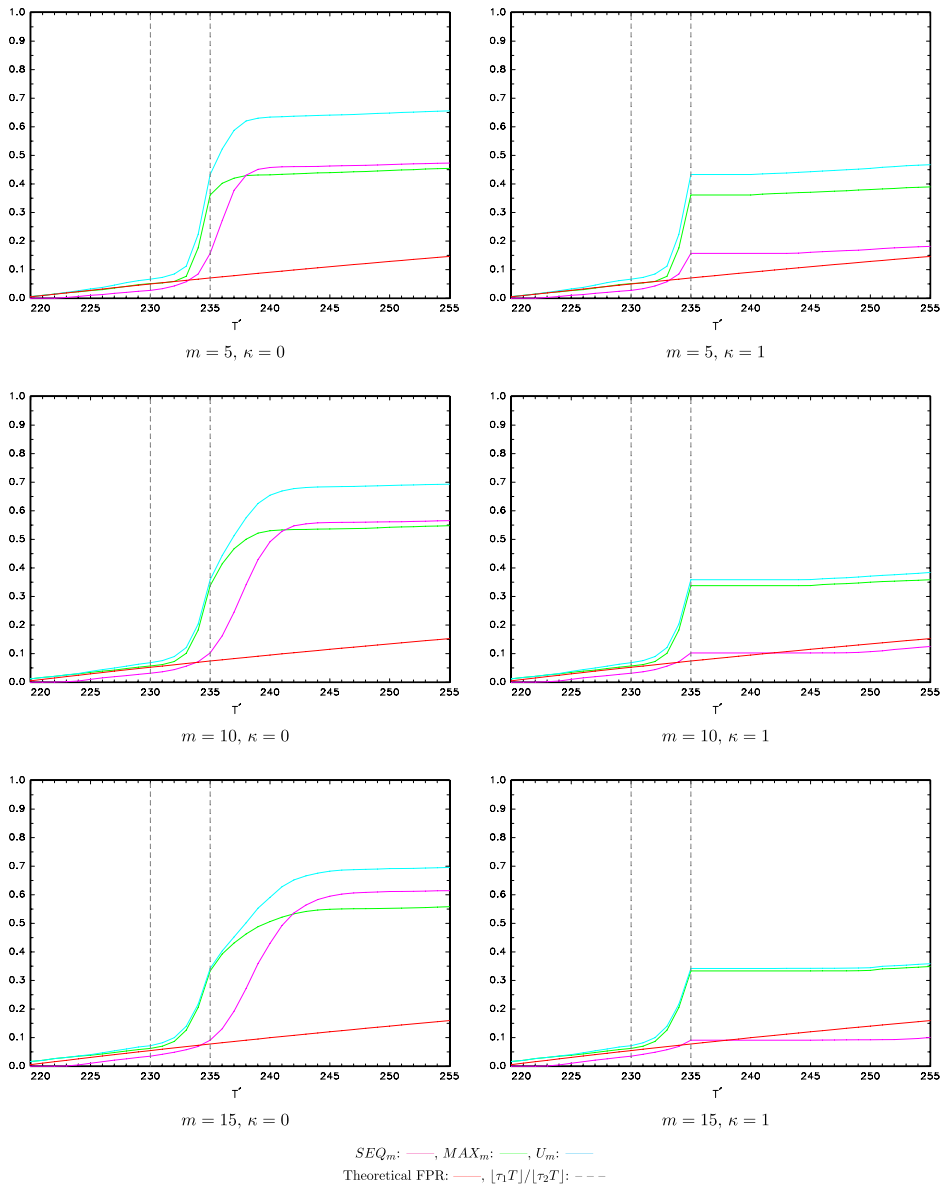


Figure 7. True positive rate $\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor = 5$, $\delta = 0.020$ [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12409

regard to the relative TPRs of the procedures, with MAX_m best suited to detecting bubbles early on, and SEQ_m only displaying a higher TPR closer to the end of the bubble when $m = 5$.

Figure 7 reports the empirical TPRs for a bubble of length 5 observations with $\lfloor \tau_1 T \rfloor = 230$ and $\lfloor \tau_2 T \rfloor = 235$ driven by an explosive offset of $\delta = 0.02$. The difference in TPR between MAX_m and SEQ_m is much more pronounced than for a bubble of length 10. This is likely caused by the fact that there are relatively few bubble observations that can contribute to SEQ_m delivering more than m^* contiguous exceedances before the bubble terminates. The difference between MAX_m and SEQ_m is particularly highlighted for the bubble with collapse (H_1^1) cases; while MAX_m has some ability to detect a bubble before its termination, SEQ_m clearly struggles here. These

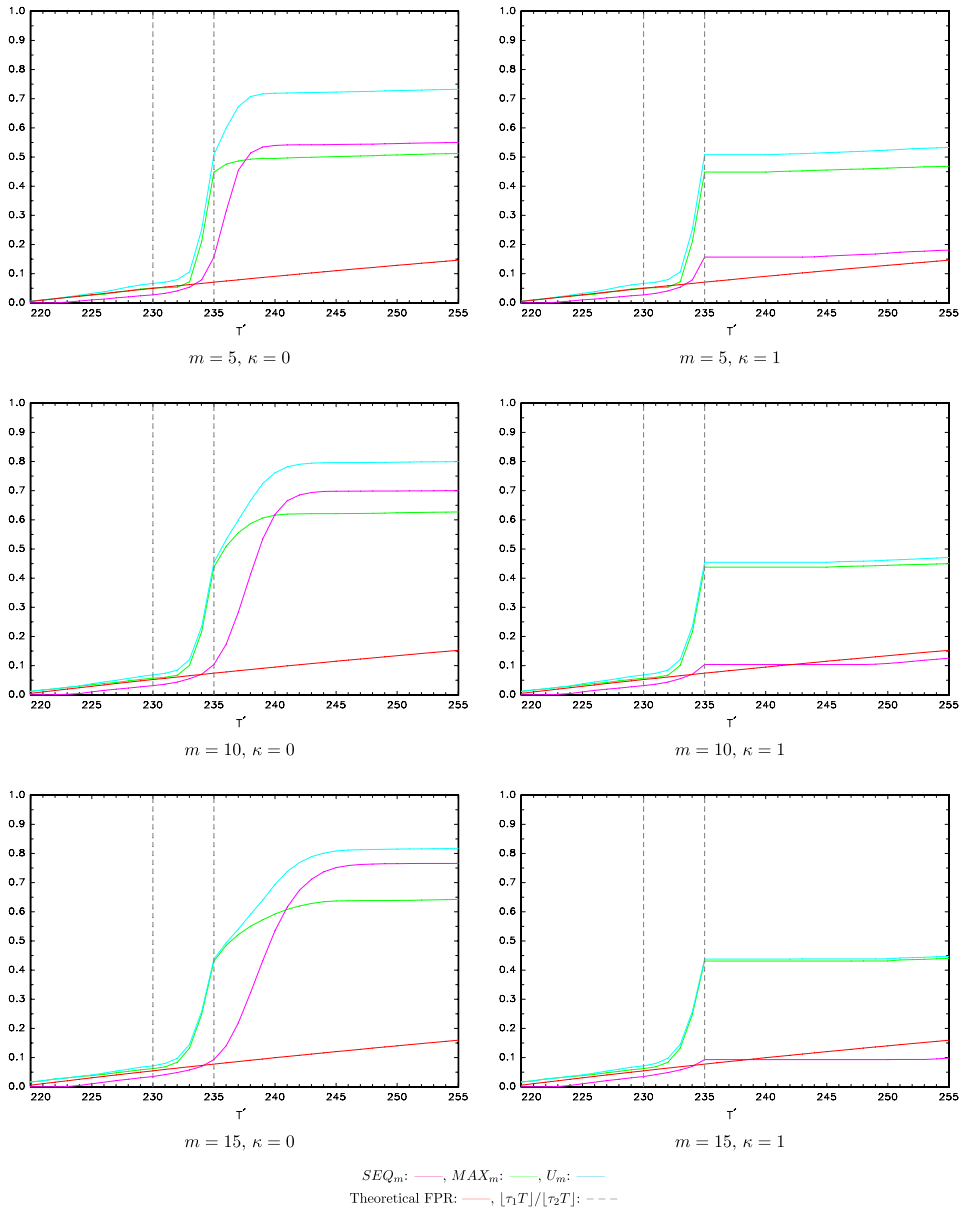


Figure 8. True positive rate $\lfloor \tau_2 T \rfloor - \lfloor \tau_2 T \rfloor = 5$, $\delta = 0.030$ [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12409

observations are reinforced by the results in Figure 8, in which the same bubble is driven by an explosive offset of $\delta = 0.03$; here, the low TPR of SEQ_m under H_1^1 is not improved by the larger magnitude of δ.

Figure 9 reports the empirical TPRs for a bubble of length 15 observations with $\lfloor \tau_1 T \rfloor = 230$ and $\lfloor \tau_2 T \rfloor = 245$ driven by an explosive offset of $\delta = 0.007$. The relative TPRs of MAX_m and SEQ_m are broadly similar to those seen for a bubble of length 10, with, once again, MAX_m having greater ability to detect a bubble early on. The real difference is that SEQ_m now has a higher TPR than MAX_m before the bubble terminates for both $m = 10$ and $m = 5$. Finally, results for a bubble of length 15 observations driven by an explosive offset of $\delta = 0.01$ are reported

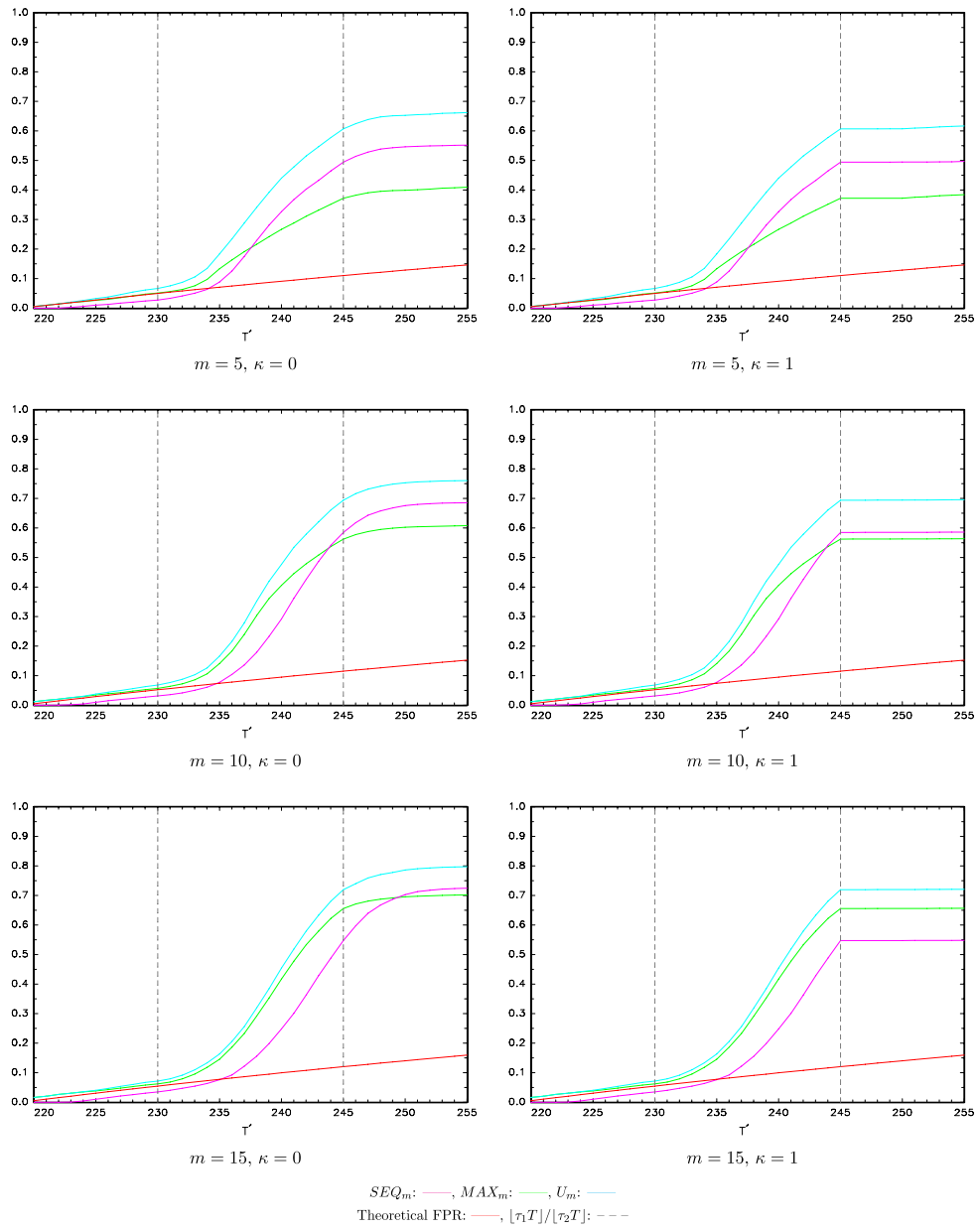


Figure 9. True positive rate $\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor = 15$, $\delta = 0.007$ [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12409

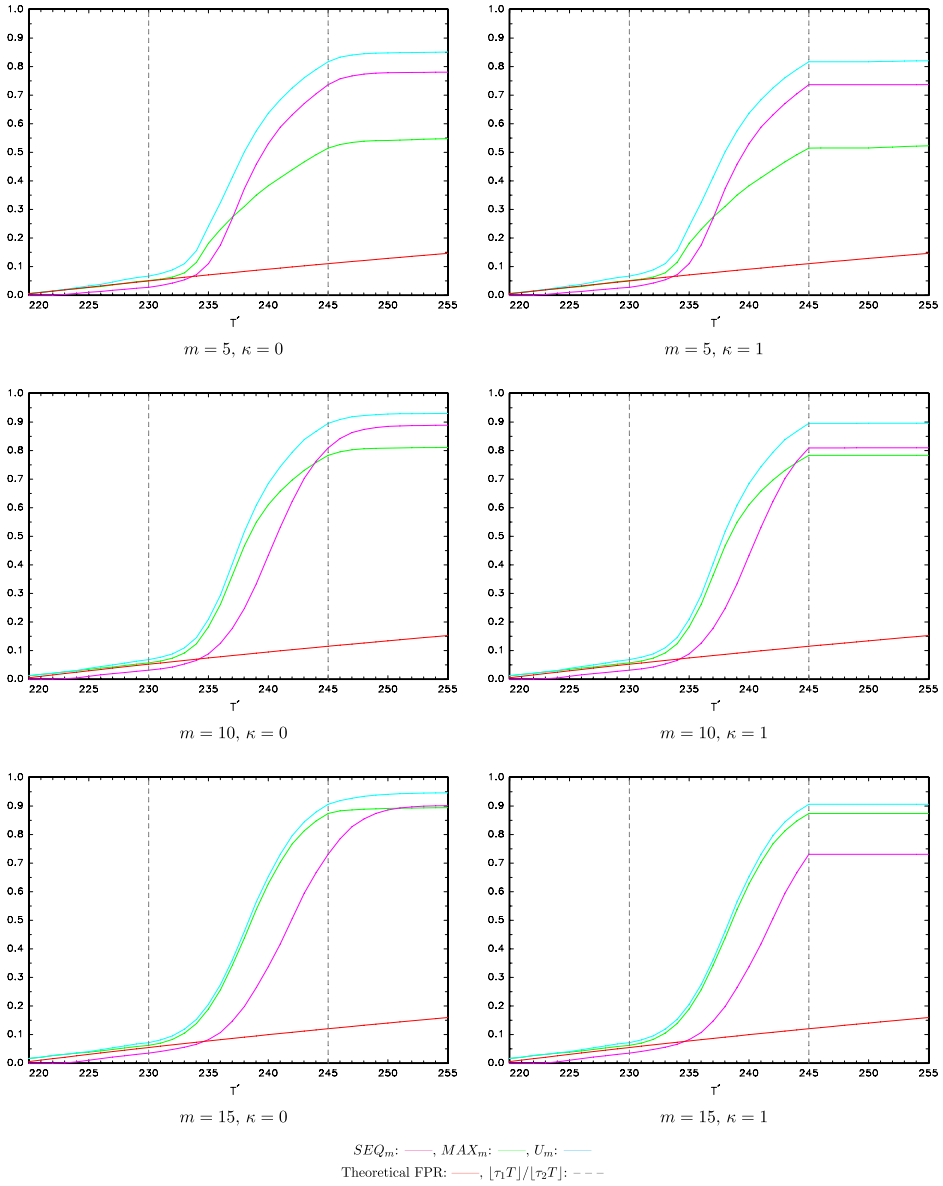


Figure 10. True positive rate $\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor = 15$, $\delta = 0.010$ [Color figure can be viewed at wileyonlinelibrary.com]

in Figure 10. These results follow the same pattern as those in Figure 9, but with the TPRs of all procedures being uniformly higher.

4.3. Training Period Bubbles

The construction of our monitoring procedures is based on the assumption that the training period data $t = 1, \ldots, T^*$ adheres to the null hypothesis of no explosivity. Clearly, this assumption could be violated in practice. While one could pretest the training period data for the presence of a bubble using, for instance, the wild bootstrap implementations of the Phillips *et al.* (2011) test developed in Harvey *et al.* (2016), such a pretest is not guaranteed

DOI: 10.1111/jtsa.12409

to detect earlier explosive episodes, particularly ones that are relatively short in duration and/or display only a small deviation from an autoregressive unit root. In light of this, we now present Monte Carlo simulation results to assess the impact that a single collapsed bubble in the training period has on the empirical FPR and TPR of our detection procedures. To that end, data were generated according to $y_t = u_t$ with

$$u_t = \begin{cases} u_{t-1} + \varepsilon_t, & t = 1, \ldots, 100, \\ 1.03u_{t-1} + \varepsilon_t, & t = 101, \ldots, 105, \\ u_{100} + \varepsilon_{106}, & t = 106, \\ u_{t-1} + \varepsilon_t, & t = 107, \ldots, \lfloor \tau_1 T \rfloor, \\ (1+\delta)u_{t-1} + \varepsilon_t, & t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor, \\ u_{\lfloor \tau_2 T \rfloor} + \kappa \mathbf{1}(\delta > 0)(u_{\lfloor \tau_1 T \rfloor} - u_{\lfloor \tau_2 T \rfloor}) + \varepsilon_t, & t = \lfloor \tau_2 T \rfloor + 1, \\ u_{t-1} + \varepsilon_t, & t = \lfloor \tau_2 T \rfloor + 2, \ldots, T \end{cases}$$

with $\varepsilon_t \sim NIID(0, 1)$. The series y_t therefore admits a single collapsed explosive episode in the training period of length five observations driven by an autoregressive parameter of 1.03. This bubble length and magnitude were chosen as the relatively high explosive autoregressive parameter of 1.03 will clearly impact upon the empirical FPR and TPR of our tests, while the duration mimics short-lived bubbles that would be difficult to detect using a pretest for explosivity performed on the training period data.

Figure 11 reports the empirical FPR of our proposed monitoring procedures when a training period bubble is present but no bubble is present in the monitoring period ($\delta = 0$). Relative to the analogous results in Figure 1(a)

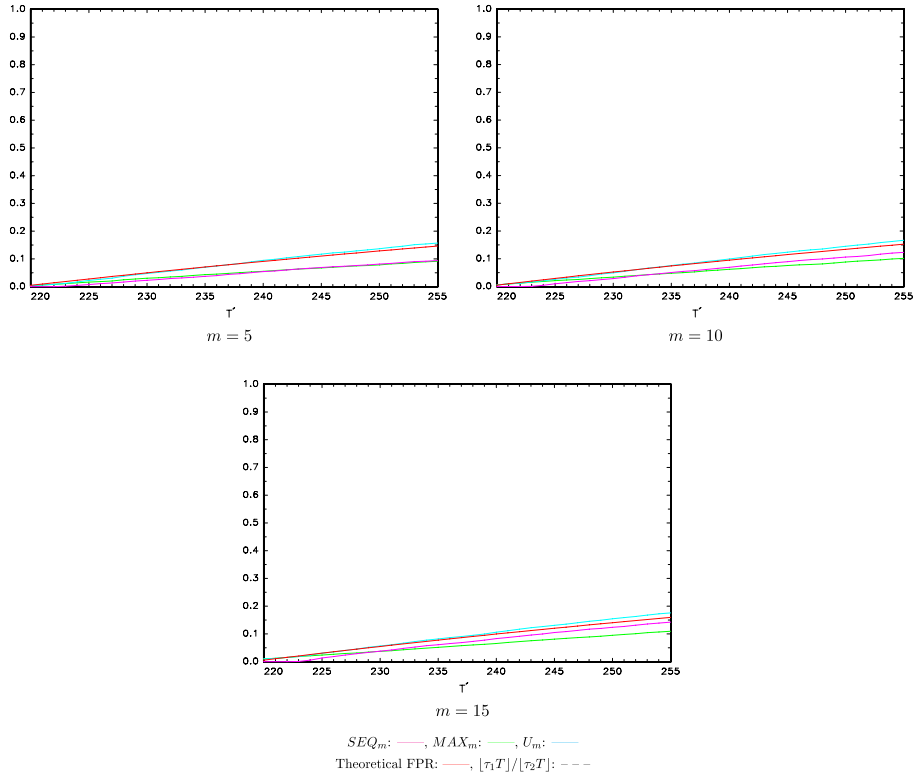


Figure 11. False positive rate—$\varepsilon_t \sim NIID(0, 1)$—training period bubble [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12409

for the case where the training period does not admit any explosive behaviour, the empirical FPR of all procedures are seen to be decreased. The presence of a training period bubble, on average, inflates the value of m^* used in the decision rule for SEQ_m and the maximum value of $S_{e,m}$ calculated in the training period used in the decision rule for MAX_m procedure, and so this effect is to be expected.

Figures 12–14 report results for the TPRs of the monitoring procedures when a training period bubble is present for monitoring period bubbles of length 10, 5, and 15 respectively. The monitoring period bubble location and the autoregressive parameter driving its magnitude are identical to those given in Figures 5, 7, and 9 to directly compare the TPR of the monitoring procedures when a training period bubble is present to the case where no training

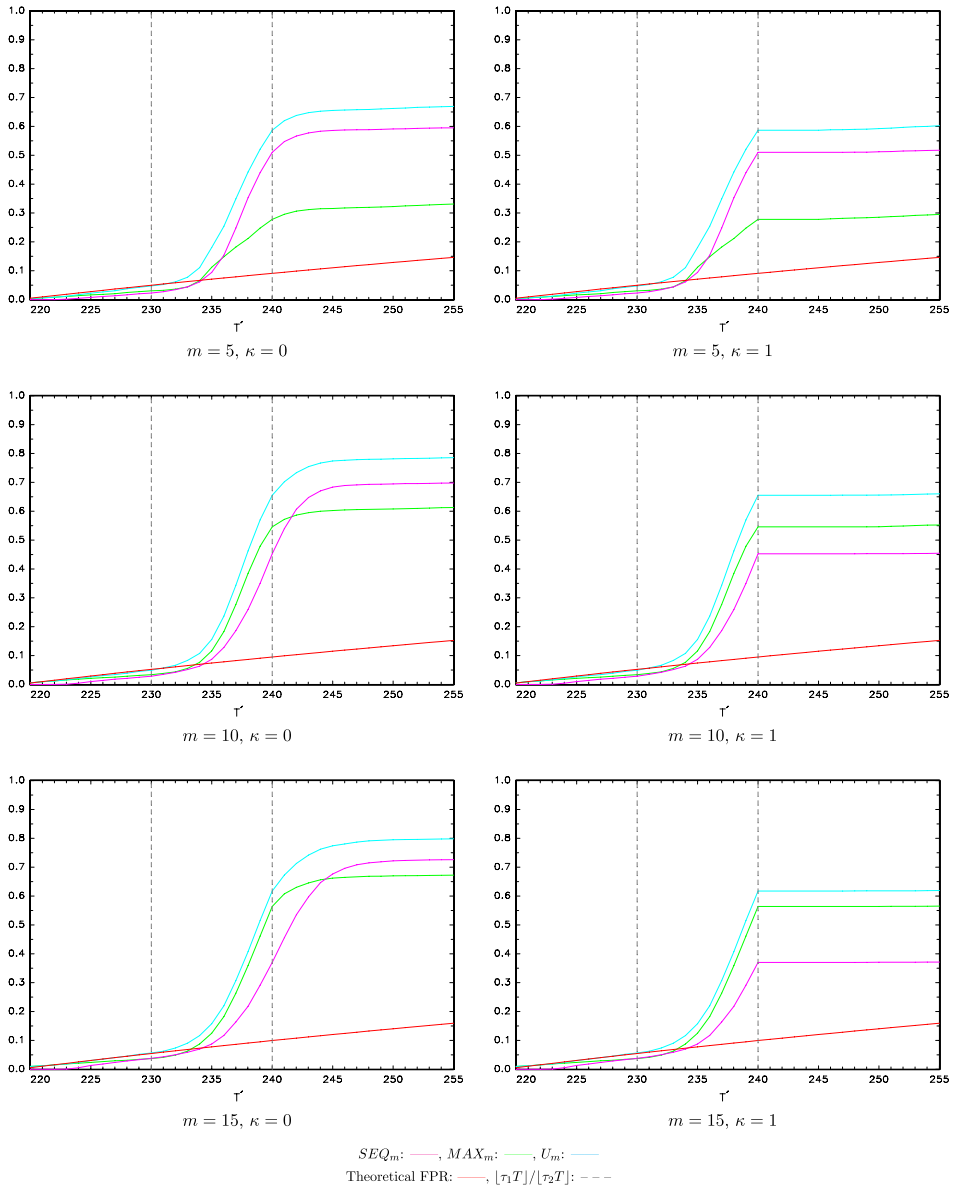


Figure 12. True positive rate $\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor = 10$, $\delta = 0.010$—training period bubble [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12409

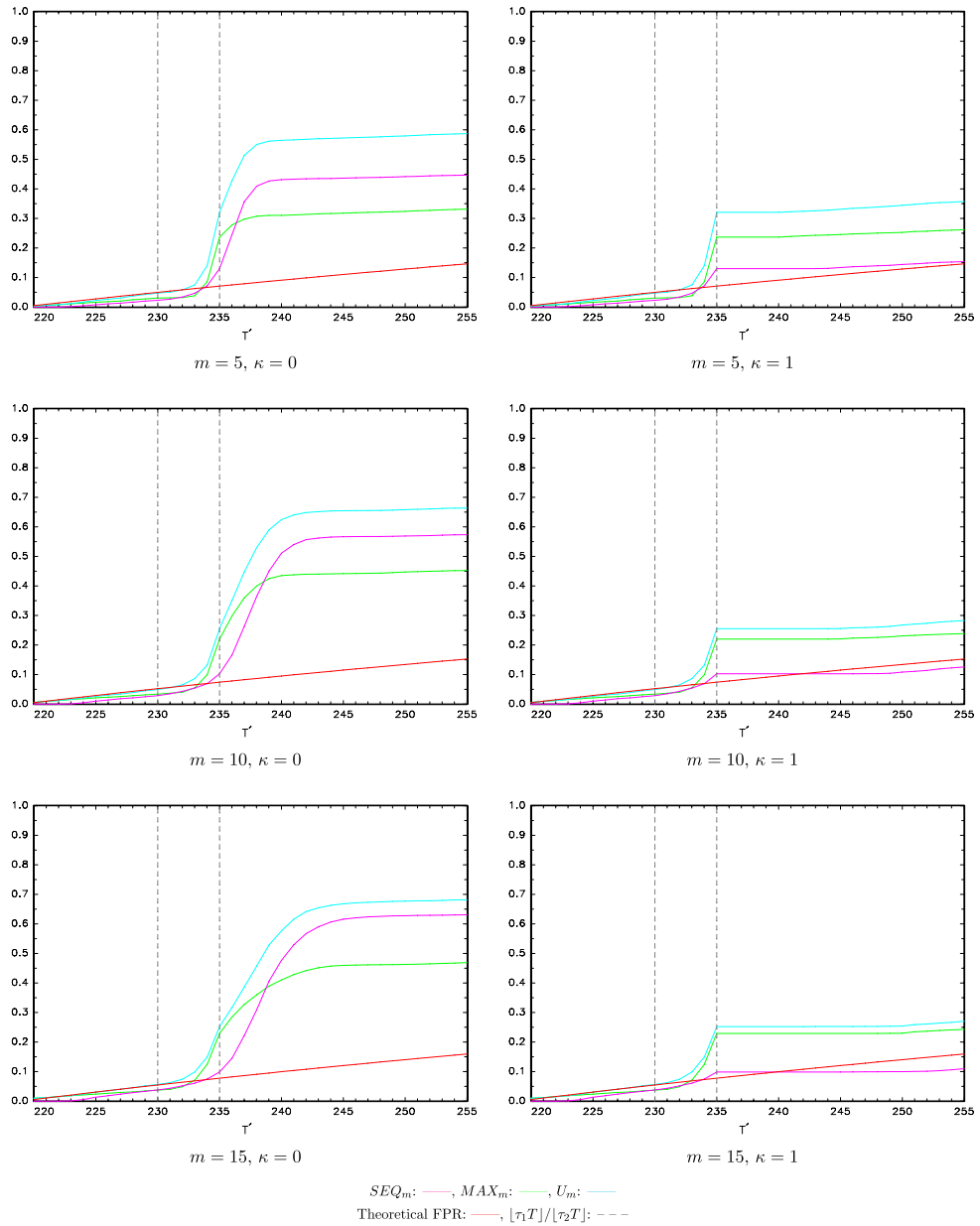


Figure 13. True positive rate $\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor = 5$, $\delta = 0.020$—training period bubble [Color figure can be viewed at wileyonlinelibrary.com]

period bubble is present. In all cases, we see that the TPR of the monitoring procedures when a training period bubble is present is lower than the case where the training period data follows a unit root process throughout, which is to be expected given the impact of the training period bubble on the empirical FPR of the procedures seen in Figure 11. While there is some reduction in TPR relative to the case where the training period is free of explosivity, this reduction is relatively modest, showing that our procedures are still useful in detecting explosive episodes in the monitoring period in cases where the training period data contains a short-lived historical bubble.

DOI: 10.1111/jtsa.12409

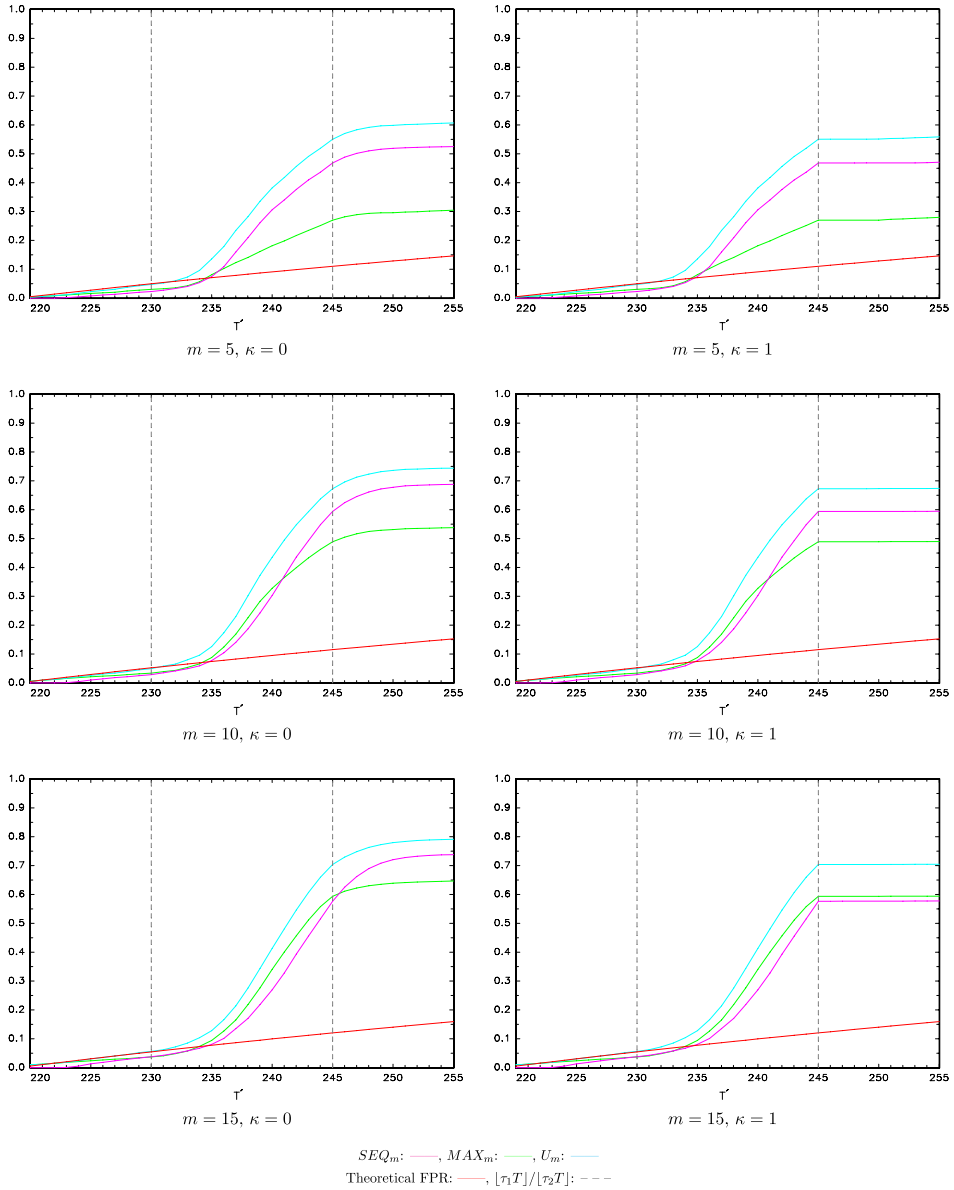


Figure 14. True positive rate $\lfloor \tau_2 T \rfloor - \lfloor \tau_1 T \rfloor = 15$, $\delta = 0.007$—training period bubble [Color figure can be viewed at wileyonlinelibrary.com]

5. EMPIRICAL APPLICATION

This section discusses an empirical application of the MAX_m and SEQ_m procedures for real-time monitoring. Each procedure is applied to five monthly data series on stock market indices collected from Thomson Reuters Datastream: the DAX 30 index (Germany), FTSE All Share index (UK), Nasdaq Composite index (USA), Nikkei 225 index (Japan), and the S&P 500 index (USA). Our full dataset covers the period January 1973 to January 2002. The training period starts in January 1973, and we assume that monitoring starts in January 1995 ($T^* + m = 265$) and continues sequentially through to January 2002, unless the monitoring terminates upon detection of a bubble. The nominal stock market index data is converted to real values using the consumer price index for each country

DOI: 10.1111/jtsa.12409

collected from the Federal Reserve Bank of St Louis FRED database. We choose to examine data over this period since, in previous research on detecting bubbles in stock market indices using a test statistic based on the recursive ADF tests (sup-ADF tests), explosive behaviour associated with the Dot-Com bubble has previously been detected during this period. For example, using sup-ADF tests, Phillips *et al.* (2011) found that the Nasdaq Composite stock market index became explosive in mid-1995. The results obtained using our methods are not directly comparable with those obtained by Phillips *et al.* (2011), because the sup-ADF tests they use are 'one-shot' tests for retrospectively detecting periods of explosive behaviour in a fixed sample of data and are not designed to be used for real-time monitoring. It is interesting to see whether our methods, which are designed for real-time monitoring, also reveal shifts to explosive behaviour during this period for the indices considered.

Figure 15 contains plots of the first differences of each of the stock market indices. Visual inspection of these plots is suggestive that the variances of these stock indices are not constant across the sample. Therefore, prior

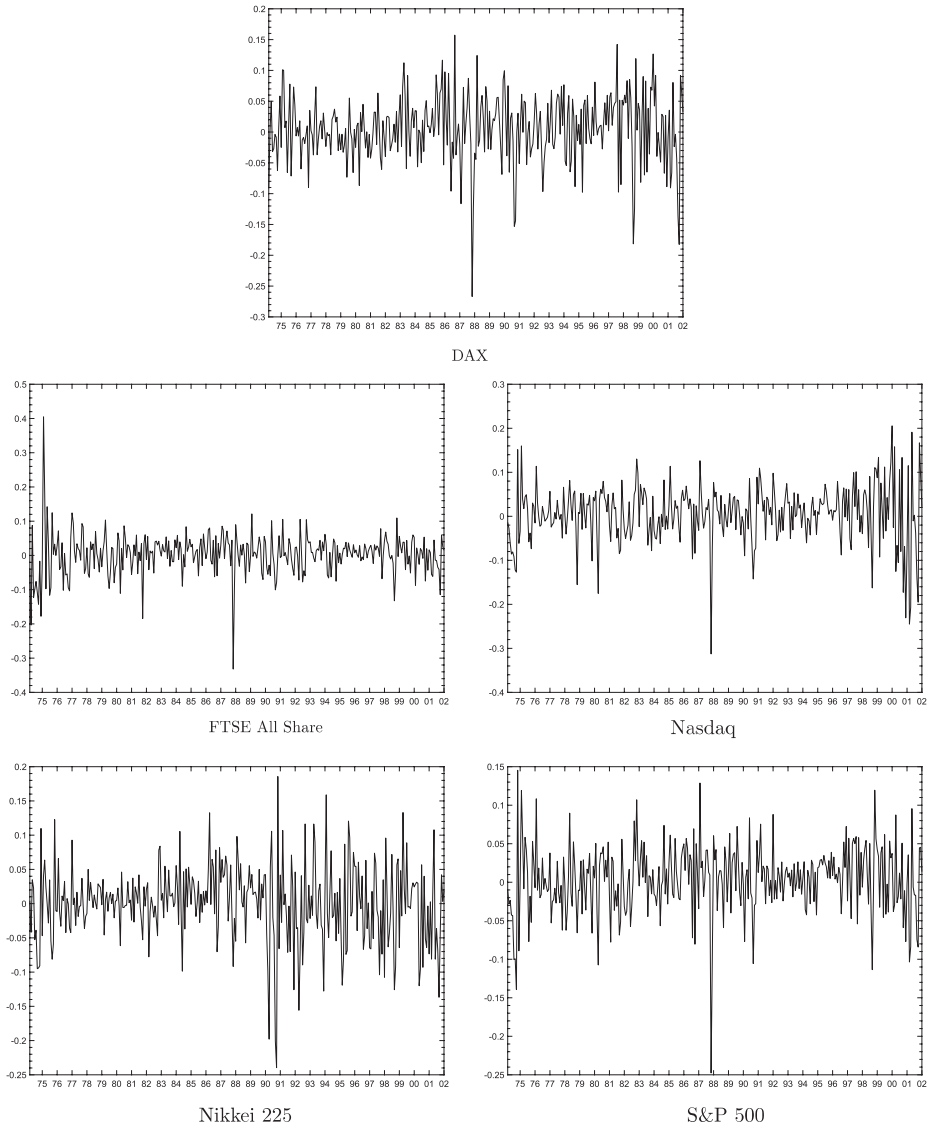


Figure 15. First difference of stock indices

DOI: 10.1111/jtsa.12409

to the application of our MAX_m and SEQ_m procedures, as a first step we apply several tests to each stock market index to assess the presence of heteroskedasticity and non-normality in the data. Table I contains the results from a Jarque–Bera test (*JB*; Jarque and Bera (1980)) and Engle's LM test for conditional heteroskedasticity (LM_c; Engle (1982)) applied to the first-difference of each series (demeaned), and four tests for stationary volatility proposed by Cavaliere and Taylor (2008, pp. 311–312). The test statistics are denoted by $\mathcal{H}_{KS}$, $\mathcal{H}_R$, $\mathcal{H}_{CVM}$, and $\mathcal{H}_{AD}$. Critical values for these four tests are given in Shorack and Wellner (2014); table 1, p. 413; table 2, p. 144; table 4, p. 147; and table 5, p. 148; respectively. When computing the stationary volatility test statistics, a Bartlett long-run variance estimator with lag truncation parameter of 4 is employed. The results in Table I show that for all the stock market indices the null hypotheses of normality and conditional homoskedasticity are rejected at conventional significance levels by *JB* and LM_c respectively. Moreover, for three of the indices (FTSE All share index, Nasdaq Composite index, Nikkei 225 index) the null hypothesis of stationary volatility is rejected by one or more of the Cavaliere and Taylor (2008) tests at conventional significance levels.

We also report an application of the wild bootstrap implementation of the Phillips *et al.* (2011) test developed in Harvey *et al.* (2016) as a pretest for the presence of bubbles in the training period for each of the datasets being considered. We apply the PWY^* and PWY_B^* bootstrap test procedures outlined in Harvey *et al.* (2016) with the lag order for the PWY^* and PWY_B^* bootstrap statistics set to zero, and the lag order for the original *PWY* statistic of Phillips, Wu and Yu (2011) chosen using the familiar Schwartz Bayes information criteria (BIC) with a maximum lag length of 6. Table II reports *p*-values for the bootstrap PWY^* and PWY_B^* tests with the lag order chosen by the BIC when calculating the sub-sample statistic using the largest possible window of observations in the *PWY* test, denoted $\hat{k}_{\text{BIC}}$. The results in Table II suggest that for all but the Nikkei 225 index there is no significant evidence of a bubble in the training period, with *p*-values for these series lying well above conventional significance levels. For the Nikkei index, the PWY_B^* test rejects the null of no explosivity at the 0.05 significance level, whereas the PWY^* test marginally fails to reject at the 0.05 level. We therefore note that there is some evidence for the presence of explosive episodes in the Nikkei 225 index training period data, which could potentially have some impact on the TPR of our test procedures to detect a bubble in the monitoring period.

Table I. Tests for non-normality and heteroskedasticity

	JB	LM_c	$\mathcal{H}_{KS}$	$\mathcal{H}_R$	$\mathcal{H}_{CVM}$	$\mathcal{H}_{AD}$
DAX 30	30.431***	336.887***	0.992	1.184	0.276	1.521
FTSE All Share	17.575***	319.255***	1.193	1.245	0.571**	3.414**
Nasdaq Composite	37.799***	333.382***	1.443**	1.646*	0.568**	4.246***
Nikkei 225	16.790***	327.509***	1.397**	1.490	0.456*	1.962*
S&P 500	35.219***	338.065***	0.594	0.892	0.085	0.601

Note: *JB* and LM_c are the Jarque–Bera test and Engle's LM test for conditional heteroskedasticity. $\mathcal{H}_{KS}$, $\mathcal{H}_R$, $\mathcal{H}_{CVM}$, and $\mathcal{H}_{AD}$ are the tests for stationary volatility proposed by Cavaliere and Taylor (2008). *, **, and *** indicate rejections at the 0.10, 0.05, and 0.01 level respectively using the relevant critical values.

Table II. Tests for historical bubbles in the training period

			p-Values	
	PWY statistic	$\hat{k}_{\text{BIC}}$	PWY^*	PWY_B^*
DAX 30	0.660	0	0.214	0.169
FTSE all share	−0.878	0	0.827	0.798
Nasdaq	−0.824	1	0.823	0.816
Nikkei 225	1.992	0	0.053	0.026
S&P 500	−0.895	0	0.864	0.837

Note: The second and third columns report the *PWY* test statistic and the lag length used when computing this test chosen using the BIC. The fourth and fifth columns report the *p*-values obtained using the PWY^* and PWY_B^* bootstrap test procedures with 9999 bootstrap replications.

We apply the MAX_m and SEQ_m procedures to each index with window lengths of $m = 5, 10, 15$. For cases where a bubble is detected, the months when the MAX_m and SEQ_m procedures first detect explosive behaviour when used sequentially from January 1995 are given in Table III, along with the associated FPRs at those months where detection occurs. A plot of each stock market index over the full potential monitoring period, along with a plot of the test statistic $S_{e,m}$ over the same period for the largest window size considered, $m = 15$, is given in Figures 16(a)–20(b). Also indicated on each plot are, for $m = 15$, the maximum value of the test statistic over the training period (black dotted line), the end of the training period (red dashed line), the start of the monitoring period (green dashed line), the critical value (blue solid line) used in SEQ_m, the date when MAX_m detects explosive behaviour (black-dashed line), the date when SEQ_m detects explosive behaviour (black dashed-dotted line), and the empirical FPR associated with monitoring out to each date in the monitoring period (magenta solid line).

The results given in Table III show that SEQ_m detects explosive behaviour (suggesting a stock market bubble) in three of the five series considered: the DAX 30 index (for $m = 5, 10, 15$), the Nasdaq Composite index (for $m = 10, 15$), and the S&P 500 index (for $m = 5, 10, 15$). The MAX_m procedure detects explosive behaviour for all five indices when $m = 15$, for two of the indices when $m = 10$ (the DAX 30 and S&P 500 index), and for one

Table III. First month where explosive behaviour is detected and empirical FPRs for real-time monitoring from 1995:1

	SEQ_m	FPR_{SEQ_m}	MAX_m	FPR_{MAX_m}
$m = 5$				
DAX 30	June 1997	0.110	N/A	N/A
FTSE all share	N/A	N/A	N/A	N/A
Nasdaq composite	N/A	N/A	N/A	N/A
Nikkei 225	N/A	N/A	N/A	N/A
S&P 500	August 1995	0.035	May 1995	0.023
$m = 10$				
DAX 30	May 1997	0.113	May 1997	0.113
FTSE all share	N/A	N/A	N/A	N/A
Nasdaq composite	January 1996	0.056	N/A	N/A
Nikkei 225	N/A	N/A	N/A	N/A
S&P 500	November 1995	0.049	September 1995	0.041
$m = 15$				
DAX 30	May 1997	0.120	May 1997	0.120
FTSE all share	N/A	N/A	February 1996	0.064
Nasdaq composite	January 1996	0.060	Mar. 1996	0.068
Nikkei 225	N/A	N/A	April 2000	0.228
S&P 500	December 1995	0.056	October 1995	0.048

Note: The second and fourth columns report the dates in the monitoring period when explosive behaviour is first detected by the SEQ_m and MAX_m procedures. The third and fifth columns report the empirical FPRs of each procedure at these dates.

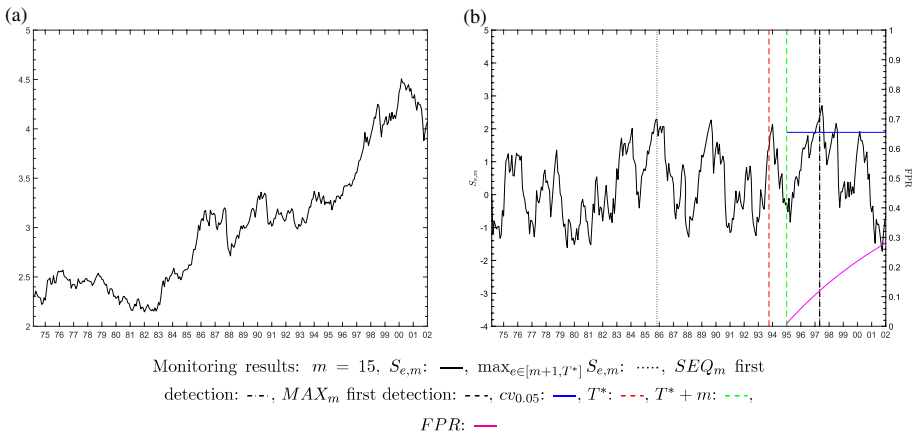
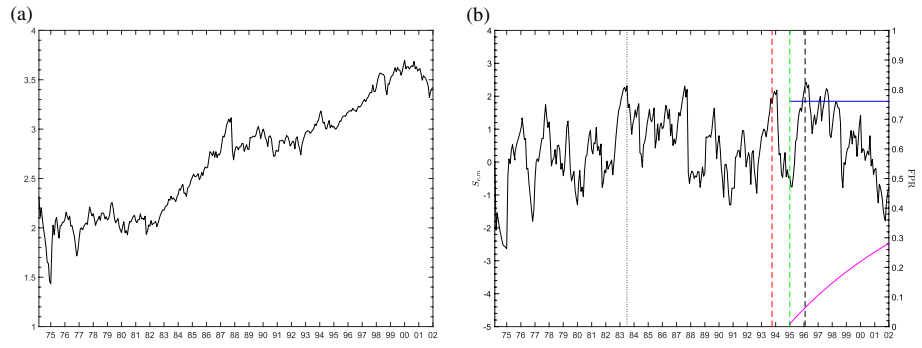


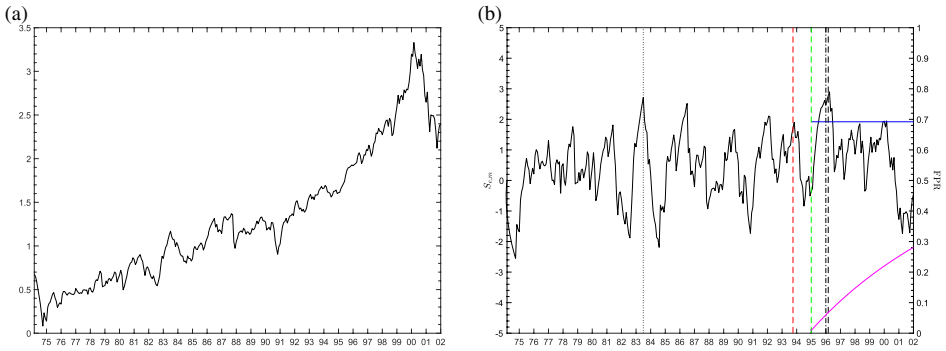
Figure 16. (a) DAX 30 index (b) Monitoring results [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12409



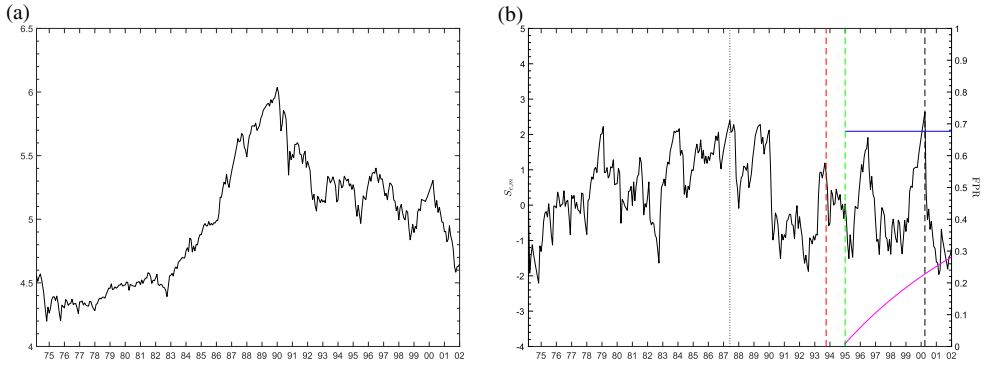
Monitoring results: $m = 15$, $S_{e,m}$: ——, $\max_{e\in[m+1,T^*]} S_{e,m}$: ·····, MAX_m first detection: - - -, $cv_{0.05}$: ——, T^*: - - -, $T^* + m$: - - -, FPR: ——

Figure 17. (a) FTSE all share index (b) Monitoring results [Color figure can be viewed at wileyonlinelibrary.com]



Monitoring results: $m = 15$, $S_{e,m}$: ——, $\max_{e\in[m+1,T^*]} S_{e,m}$: ·····, SEQ_m first detection: - - -, MAX_m first detection: - - -, $cv_{0.05}$: ——, T^*: - - -, $T^* + m$: - - -, FPR: ——

Figure 18. (a) Nasdaq composite index (b) Monitoring results [Color figure can be viewed at wileyonlinelibrary.com]



Monitoring results: $m = 15$, $S_{e,m}$: ——, $\max_{e\in[m+1,T^*]} S_{e,m}$: ·····, MAX_m first detection: - - -, $cv_{0.05}$: ——, T^*: - - -, $T^* + m$: - - -, FPR: ——

Figure 19. (a) Nikkei 225 index (b) Monitoring results [Color figure can be viewed at wileyonlinelibrary.com]

DOI: 10.1111/jtsa.12409

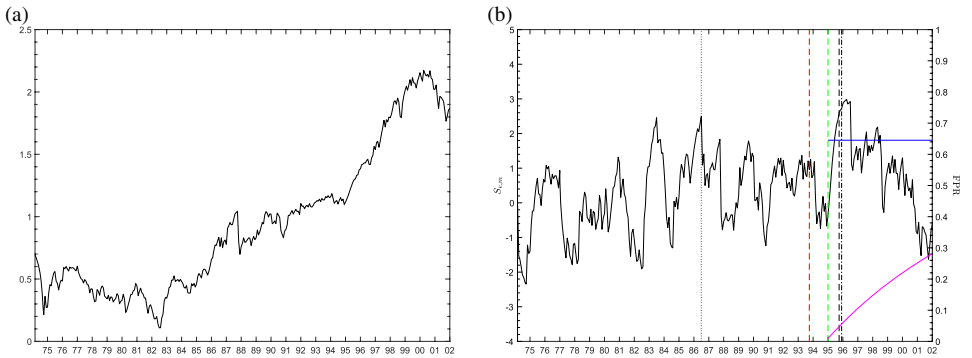


Figure 20. (a) S&P 500 index (b) Monitoring results [Color figure can be viewed at wileyonlinelibrary.com]

index when $m = 5$ (the S&P 500 index). Note that this pattern of results is consistent with the results obtained from the simulations discussed in Section 4, in the sense that in the simulations with $m = 15$ we found that the MAX_m typically has a higher TPR than SEQ_m, and the overall TPR is higher for the larger values of m.

It can also be seen in Table III that when explosive behaviour is detected by our procedures, the first detection point occurs in the early years of the Dot-Com bubble period, apart from for the Nikkei 225 index. Consider, for example, the results for the DAX 30 index in Figure 16(a, b). In this case, depending on the value of m used, explosive behaviour is first detected by SEQ_m in May 1997 or June 1997. Similarly, for the Nasdaq Composite index, explosive behaviour is first detected by SEQ_m in January 1996 and by MAX_m in March 1996. The Nikkei 225 index behaves quite differently to the other indices over this period and explosive behaviour is not detected by MAX_m until April 2000, at the peak of the Dot-Com bubble. When explosive behaviour is detected for more than one m setting for a given series, the first date at which explosivity is identified is earlier for the smaller values of m, in line with the TPR results in our simulations.

Interestingly, although the Dot-Com bubble is associated with information technology stocks, the earliest indication of explosive behaviour from our procedures around the time when the Dot-Com bubble is thought to have started is for the broad S&P 500 index rather than for the information technology-focused Nasdaq Composite index. For the S&P 500 index when $m = 5$, MAX_m detects explosive behaviour in May 1995, while for the Nasdaq Composite index explosive behaviour is detected in January 1996 (when $m = 10$). When $m = 15$, explosive behaviour is detected in the S&P 500 index in October 1995 by MAX_m and in December 1995 by SEQ_m. For the Nasdaq Composite index, when $m = 15$, explosive behaviour is also detected later in the sample than for the S&P 500 index, in January 1996 and March 1996 for SEQ_m and MAX_m respectively. Naturally, the earlier that explosive behaviour is detected in the monitoring period, the smaller the associated FPR is for the test procedure at that date. For example, Table III shows that for the S&P 500 index when $m = 5$, the associated FPR for the MAX_m detection of explosive behaviour in May 1995 is just 0.023, while for the Nasdaq Composite index with $m = 10$, the SEQ_m detection of explosive behaviour in January 1996 has an FPR of 0.056. Since it occurs much later in the monitoring period (April 2000), the FPR associated with MAX_m for the Nikkei 225 index when $m = 15$ is considerably larger at 0.228.

6. CONCLUSIONS

We have proposed monitoring procedures that can be used by practitioners to detect the emergence of asset price bubbles in real time. Our procedures involve sequential computation of sub-sample-based test statistics from a training period of data. Our first procedure signals the presence of a bubble if any statistic in the monitoring period exceeds the largest sub-sample statistic calculated in the training period, whereas our second procedure signals

DOI: 10.1111/jtsa.12409

the presence of a bubble in the monitoring period when the number of contiguous rejections in the monitoring period exceeds the number of contiguous rejections in the training period, using a critical value obtained from the training sample statistics. We also proposed a union-of-rejections approach in which a bubble is detected if either of the procedures reject the null of no bubble. We have shown that a practitioner can determine the theoretical FPR of the procedures for any given monitoring horizon, or can ensure that the FPR does not exceed a specified level, by setting a maximum monitoring horizon. A Monte Carlo exercise comparing the empirical FPR of our procedures with those of the CUSUM procedure of Homm and Breitung (2012) showed that only the procedures developed in this article were empirically robust to time-varying volatility and serial correlation in the shocks. Further simulations showed that our proposed procedures are able to rapidly detect an emerging bubble in real time, and our results showed that this can even be the case when a past bubble is present in the training period data. An empirical application to five major stock market indices found that our monitoring procedures would, as part of a real-time monitoring exercise, in some form have signalled the presence of bubbles in each index if run from January 1995 to January 2002. While our focus in this article has been on financial asset price bubbles, the model and our proposed bubble monitoring procedures can equally be applied in other contexts, allowing real-time bubble detection in a wide range of situations.

REFERENCES

Andrews DWK. 2003. End-of-sample instability tests. *Econometrica* **71**: 1661–1694.

Andrews DWK, 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 DI, Leybourne SJ, Taylor AMR. 2017. Tests for an end-of-sample bubble in financial time series. *Econometric Reviews* **36**: 651–666.

Cavaliere G, Taylor AMR. 2008. Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis* **29**: 300–330.

Diba BT, Grossman HI. 1988. Explosive rational bubbles in stock prices? *American Economic Review* **78**: 520–530.

Engle RF. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**: 987–1007.

Harvey DI, Leybourne SJ, Sollis R. 2017. Improving the accuracy of asset price bubble start and end date estimators. *Journal of Empirical Finance* **40**: 121–138.

Harvey DI, Leybourne SJ, Sollis R, Taylor AMR. 2016. Tests for explosive financial bubbles in the presence of non-stationary volatility. *Journal of Empirical Finance* **38**: 548–574.

Harvey DI, Leybourne SJ, Sollis R, Taylor AMR. 2018. Detecting regimes of predictability in the U.S. equity premium, Nottingham, UK: University of Nottingham. Discussion Paper.

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

Jarque CM, Bera AK. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* **6**: 255–259.

Phillips PCB, Wu Y, Yu J. 2011. Explosive behavior in the 1990s Nasdaq: when did exuberance escalate stock values? *International Economic Review* **52**: 201–226.

Phillips PCB, Shi S-P, Yu J. 2015. Testing for multiple bubbles: historical episodes of exuberance and collapse in the S&P 500. *International Economic Review* **56**: 1043–1078.

Shorack GR, Wellner JA. 2014. *Empirical Processes with Applications to Statistics*. Berlin, Germany: Springer.

DOI: 10.1111/jtsa.12409