Journal of Financial Econometrics, 2021, 1–41
doi: 10.1093/jjfinec/nbab009
Article

OXFORD

CUSUM-Based Monitoring for Explosive Episodes in Financial Data in the Presence of Time-Varying Volatility*

Sam Astill[1], David I. Harvey [2], Stephen J. Leybourne[3], A.M. Robert Taylor[4] and Yang Zu[5]

[1]Essex Business School, University of Essex, [2]Granger Centre for Time Series Econometrics and School of Economics, University of Nottingham, [3]Granger Centre for Time Series Econometrics and School of Economics, University of Nottingham, [4]Essex Business School, University of Essex and [5]Granger Centre for Time Series Econometrics and School of Economics, University of Nottingham

Address correspondence to David I. Harvey, School of Economics, University of Nottingham, University Park, Nottingham NG7 2RD, UK, or e-mail: dave.harvey@nottingham.ac.uk

Received March 12, 2020; revised February 22, 2021; editorial decision February 28, 2021; accepted March 12, 2021

Abstract

We generalize the Homm and Breitung (2012) CUSUM-based procedure for the real-time detection of explosive autoregressive episodes in financial price data to allow for time-varying volatility. Such behavior can heavily inflate the false positive rate (FPR) of the CUSUM-based procedure to spuriously signal the presence of an explosive episode. Our modified procedure involves replacing the standard variance estimate in the CUSUM statistics with a nonparametric kernel-based spot variance estimate. We show that the sequence of modified CUSUM statistics has a joint limiting null distribution which is invariant to any time-varying volatility present in the innovations and that this delivers a real-time monitoring procedure whose theoretical FPR is controlled. Simulations show that the modification is effective in controlling the empirical FPR of the procedure, yet sacrifices only a small amount of power to detect explosive episodes, relative to the standard procedure, when the shocks are homoskedastic. An empirical illustration using Bitcoin price data is provided.

Key words: CUSUM, explosive autoregression, nonparametric spot volatility estimator, rational bubble, real-time monitoring

JEL classification: C22, C12, G14

* The authors thank the Co-Editor, Dacheng Xiu, and two anonymous referees for helpful and constructive comments.

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 examples 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 has been argued to constitute a more 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 their collapse, 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.

Accordingly, a large literature has developed in the last decade or so around testing for the presence of historical explosive rational asset bubbles in price series. An early contribution is Diba and Grossman (1988) who applied orthodox left-tailed unit root tests (i.e., tests against stationary autoregressive alternatives) to the price and dividend series in levels and first-differenced forms to investigate the presence of asset bubbles in stock price data. They adopt this approach based on the observation that if the bubble component of the stock price evolves as an explosive autoregressive process then, as an explosive autoregressive process cannot be differenced to stationarity, a finding of non-stationarity for the price and dividend series when the series are in levels, but stationarity when the series are in first differences, is indicative that an explosive rational bubble does not exist. However, Evans (1991) argues that the tests adopted in Diba and Grossman (1988) will have little, if any, power to detect periodically collapsing bubbles. Consequently, the recent focus in the literature has been on the use of right-tailed unit root tests, that is, tests against explosive autoregressive alternatives, applied to the levels of a series. The first such contribution was made by Phillips, Wu, and Yu (2011), who developed a test of the null of no explosive behavior against the alternative of explosivity based on a sequence of forward recursive right-tailed augmented Dickey–Fuller (DF) statistics. Further contributions using sub-sample testing methods have been developed in Homm and Breitung (2012), Harvey, Leybourne, and Sollis (2015), Harvey et al. (2016), Harvey, Leybourne, and Zu (2019, 2020), Phillips, Shi, and Yu (2015), Astill et al. (2017), Phillips and Shi (2018), among others. Applications of these methods have uncovered evidence of historical asset price bubbles in stock prices, commodities futures prices, real estate prices, exchange rates, and many other price series; see Homm and Breitung (2012) for a detailed review.

A feature of the procedures outlined above, however, is that they are designed to detect speculative bubbles within a fixed historical dataset. In practice, it would seem to be of much greater practical relevance to sequentially monitor for the emergence of an asset price bubble as new data points are obtained using a real-time monitoring procedure. While sequential application of the tests of Phillips, Shi, and Yu (2015) or Astill et al. (2017), both of which are designed to detect an end-of-sample explosive autoregressive episode, could be used to do this, one could not use the critical values appropriate for their use as one-shot tests in such a monitoring exercise as these would not be size controlled. In particular, the overall false-positive rate (FPR) of such a procedure would be unknown and, as discussed in the context of a generic monitoring exercise in Chu, Stinchcombe, and White (1996) and also for the specific case of monitoring for the emergence of an explosive episode in Homm and Breitung (2012) and Astill et al. (2018), would increase monotonically as the monitoring horizon grows. Here, we define the FPR as the probability of at least one test in the monitoring sequence rejecting when the null was true and, hence, no explosive episode was

present. Homm and Breitung (2012) and Astill et al. (2018) develop real-time monitoring procedures for explosive episodes which are such that the theoretical FPR can be controlled by the practitioner. Astill et al. (2018) develop a real-time monitoring procedure based on sequential application of the end of sample test of Astill et al. (2017). Their preferred procedure signals the presence of an explosive episode if any statistic in the monitoring period exceeds the largest value of the statistic calculated over a training period of data. Homm and Breitung (2012) propose two real-time monitoring procedures, one based on standard cumulative sum (CUSUM) statistics and the other on (unaugmented) DF unit root statistics. Adopting the methodology of Chu, Stinchcombe, and White (1996), the CUSUM and DF statistics are calculated sequentially across a given monitoring period with a decision rule designed to control the theoretical FPR of the procedure.

A key assumption underlying the large sample validity of the real-time monitoring procedures of Homm and Breitung (2012) is that the shocks driving the series being monitored are unconditionally homoskedastic. This assumption is not innocuous and indeed is likely to be infeasible for many financial price series which display clear patterns of time-varying volatility. In particular, many applied studies have found strong evidence of structural breaks in the unconditional variance of asset returns, often linked to major financial and macroeconomic crises such as the 1970s oil price shocks, the East Asian currency crisis in the late-1990s, the dot-com crash in 2001, and the recent global financial crisis in 2007–2009. In a number of these studies very large structural breaks have been detected; for example, Rapach, Strauss, and Wohar (2008) and McMillan and Wohar (2011) detect breaks in the unconditional variance of the returns of some major stock market indices and sectoral stock price indices, finding that the unconditional variance in some sub-samples can be larger than that in other sub-samples by a factor of about 10. For commodity returns, both Calvo-Gonzalez, Shankar, and Trezzi (2010) and Vivian and Wohar (2012) find statistically significant evidence of structural breaks in unconditional volatility. 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 being present. It is therefore important to develop reliable methods for detecting an emerging explosive period in a series that is robust to the presence of time-varying volatility.

Using Monte Carlo simulation, Astill et al. (2018) show that the empirical FPR of the CUSUM-based procedure of Homm and Breitung (2012) cannot be adequately controlled in the presence of time-varying volatility and can differ quite drastically from the theoretical FPR which obtains under homoskedasticity. In contrast they show that the empirical FPR of their maximum-based procedure is robust to a wide range of time-varying patterns of volatility. However, as we show in the simulation results in this paper, in the case where the innovations are homoskedastic, such that its FPR is controlled, the CUSUM-based procedure displays a very clear advantage over the procedure of Astill et al. (2018) in terms of its empirical true positive rate (TPR) to detect an emergent explosive episode, where the TPR is defined as the probability of at least one test in the monitoring sequence rejecting when an explosive period is present. Given that our aim is to develop real-time monitoring procedures which have both a controlled FPR and strong power to detect an emerging explosive episode, it therefore seems worthwhile developing a heteroskedasticity-robust version of the CUSUM-based procedure. To that end, we propose a modification to the CUSUM-based procedure which replaces the standard full sample first-difference-based

variance estimate used by Homm and Breitung (2012) in calculating the CUSUM statistics with a nonparametric kernel-based spot variance estimate, designed to model the *unknown* variance path of the underlying innovations.

Under quite general conditions we show that the resulting sequence of modified CUSUM statistics has a joint limiting null distribution which is invariant to any time-varying volatility present in the innovations and that, as a result, this delivers a real-time monitoring procedure whose theoretical FPR is controlled. Indeed, these quantities are shown to coincide with those which obtain for the standard CUSUM procedure in the case of homoskedastic innovations. Monte Carlo methods are used to examine the empirical FPR and TPR of our proposed monitoring procedure. These results show that the empirical FPR of the modified procedure is well controlled in practice. Moreover, the efficacy of the modified procedure to detect an explosive episode, as measured by the empirical TPR, is shown to be little altered in the homoskedastic case, so that the cost (in terms of ability to detect an emerging explosive episode) of this additional robustness to time-varying volatility appears relatively small. We also show here that the presence of an explosive episode prior to the start of the monitoring period has little impact on the properties of our modified CUSUM procedure but can very substantially lower the empirical TPR of both the CUSUM-based procedure and the procedure of Astill et al. (2018).

The remainder of the paper is organized as follows. Section 1 outlines the autoregressive data generating process (DGP) we work with and outlines the assumptions under which our analysis will be conducted. In Section 2, we briefly review the CUSUM-based procedure of Homm and Breitung (2012) and demonstrate that it does not, in general, have a controlled FPR when time-varying volatility is present in the innovations. We then outline our modified CUSUM procedure and establish the large sample validity of this procedure. Issues concerning its practical implementation, including the selection of the bandwidth and kernel used in the context of the nonparametric spot variance estimator, are also discussed in this section. Our Monte Carlo study is reported in Section 3. An empirical illustration of our modified CUSUM monitoring procedure, using Bitcoin price data, is provided in Section 4. Section 5 concludes.

1 The Model and Assumptions

We consider the time series process $\{y_t\}$ generated according to the following autoregressive DGP:

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

$$u_t = \begin{cases} u_{t-1} + e_t, & t = 1, \ldots, T, \\ u_{t-1} + e_t, & t = T+1, \ldots, \lfloor \tau_1 T \rfloor, \\ (1+\delta)u_{t-1} + e_t, & t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor, \\ u_{t-1} + e_t, & t = \lfloor \tau_2 T \rfloor + 1, \ldots, \lfloor \lambda T \rfloor \end{cases} \tag{2}$$

where $1 \leq \tau_1 < \tau_2 \leq \lambda$ and $\lambda > 1$. Here, $\lfloor . \rfloor$ is used to denote the integer part of its argument, u_0 is taken to be an $O_p(1)$ variate, and e_t is a possibly heteroskedastic error process whose properties are discussed in detail below.

The specification of the DGP in (1) and (2) defines the series y_t separately over two subsample periods: the period $t = 1, \ldots, T$ which will later form the *training period* in our

analysis, and the period $t = T+1, \ldots, \lfloor \lambda T \rfloor$ which will form the *monitoring period* for our procedure. Our model is such that y_t follows a unit root process over the training period $t = 1, \ldots, T$, while over the monitoring period, y_t again follows a unit root process over the sub-periods $t = T+1, \ldots, \lfloor \tau_1 T \rfloor$ and $t = \lfloor \tau_2 T \rfloor + 1, \ldots, \lfloor \lambda T \rfloor$, but crucially is subject to potentially explosive behavior in the period $t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor$ when $\delta > 0$. In total there are $\lfloor \lambda T \rfloor$ observations with $\lambda > 1$ a fixed constant. When $\delta > 0$, if $\tau_1 = 1$ then the explosive regime will begin at the start of the monitoring period, while if $\tau_2 = \lambda$, the explosive regime will still be on-going at the end of the monitoring period. In the context of monitoring for explosive autoregressive behavior during the monitoring period, our null hypothesis is given by $H_0 : \delta = 0$, with the corresponding alternative hypothesis being $H_1 : \delta > 0$.

Remark 1. The model considered in (1) and (2) does not allow for a collapse following the termination of the explosive regime. The model could easily be extended to allow for either an instantaneous collapse (as in, e.g., Phillips, Wu, and Yu, 2011), or a stationary collapse regime (as in, e.g., Harvey et al., 2016). However, when monitoring for an emerging explosive regime in real time, the nature of any post-explosive collapse has no bearing on the detection properties of the monitoring procedures. While some differences will arise when monitoring beyond the point at which an explosive regime terminates, this is a secondary consideration for the purposes of this paper and, as a consequence, we focus on the case of a non-collapsing explosive period for simplicity. Simulation results for models with collapse regimes in the monitoring period are available on request.

With respect to the error, e_t, we allow for the possibility of non-constancy in its unconditional volatility by setting $e_t = \sigma_t \varepsilon_t$, such that σ_t^2 is the unconditional (spot) variance of e_t and where ε_t is a homoskedastic innovation sequence. Precisely, we make the following assumptions regarding ε_t and σ_t, respectively:

Assumption 1. *ε_t is a martingale difference sequence with respect to the natural filtration generated by the sequence of ε_t, $\{\mathcal{F}_t\}$, such that $\mathrm{Var}(\varepsilon_t | \mathcal{F}_{t-1}) = 1$ and $E(\varepsilon_t^4) < \infty$.*

Assumption 2. *For $t = 1, \ldots, T, \ldots, \lfloor \lambda T \rfloor$, σ_t is non-stochastic and satisfies $\sigma_t := \sigma(t/T)$. The function $\sigma(.)$ has support $[0, \lambda]$ and is strictly positive, continuously differentiable and uniformly bounded by a constant M. Furthermore, the derivative of $\sigma(.)$ is Lipschitz continuous over $(0, \lambda)$.*

Remark 2. Assumption 1 imposes conditional homoskedasticity on the innovation sequence ε_t. This assumption is standard in the time-varying volatility literature; see, for example, Hansen (1995), Phillips and Xu (2006), Xu and Phillips (2008), Harris and Kew (2017), Boswijk and Zu (2018), Harvey, Leybourne, and Zu (2020), Harris, Kew, and Taylor (2020), and Boswijk and Zu (2021). The moment condition, $E(\varepsilon_t^4) < \infty$, imposed by Assumption 1 is weaker than is usually made in the literature, where an assumption of the existence of the 8th moment is standard; an exception is Beare (2018) who makes a comparable finite $4 + \delta$, $\delta > 0$, moment assumption in connection with the unit root tests he develops for cases where the errors display time-varying volatility.

Remark 3. Assumption 2 allows for time-varying behavior in the unconditional volatility of e_t including, among other things, smooth transition single or multiple level shifts and trending volatility which may also be subject to smooth breaks in the trend coefficient. The

case of constant volatility, where $\sigma_t = \sigma$, for all t, also satisfies Assumption 2 because here $\sigma(s) = \sigma$ for all s. Discrete jumps in volatility are formally ruled out under Assumption 2 which imposes continuity on the volatility path $\sigma(\cdot)$. This smoothness requirement on the volatility function is needed to obtain the uniform consistency results for our nonparametric kernel-based volatility estimator, which is in turn needed for our main result given in Theorem 1 to hold. The smoothness assumption is not restrictive in practice, because one can always approximate discontinuities in $\sigma(\cdot)$ arbitrarily well using smooth transition functions. The conditions imposed on the errors, e_t, by Assumptions 1 and 2 are therefore considerably weaker than those of Homm and Breitung (2012) who assume that e_t is independent and identically distributed (IID) with mean zero and constant variance, σ^2.

2 Real-Time Explosive Episode Detection Procedures

In this section, we briefly review the CUSUM monitoring procedure of Homm and Breitung (2012) and propose a modification to this procedure to allow for the possibility of time-varying volatility in the innovations. As mentioned in the introduction, Homm and Breitung (2012) also propose a second monitoring procedure, which they label FLUC, based on sequential DF statistics; see Equations (27) and (31) of Homm and Breitung (2012). We will not consider this DF-based procedure any further in this paper for the following reason. Where no detrending is undertaken, one could use the approach taken in Beare (2018) to develop a heteroskedasticity-robust version of the DF statistic used by Homm and Breitung (2012), based on the same estimator of σ_t used to modify the CUSUM statistic in Section 2.2. The resulting FLUC procedure would then share the same large sample properties as attained by the standard FLUC procedure under homoskedasticity. However, this approach does not seem extendable to the case where detrending is used as in Remark 10. In particular, Beare (2018) demonstrates that the limiting null distribution of his modified DF statistic in that case still depends on the volatility path, $\sigma(\cdot)$. It would then appear infeasible to standardize this statistic in the way done in the homoskedastic case by Homm and Breitung (2012, p. 212) to ensure the boundary function κ_t used in Equation (31) of Homm and Breitung (2012) is positive when detrending is undertaken.

2.1 The Homm–Breitung CUSUM-Based Procedure

Under the additional assumption that e_t is an IID process with mean zero and variance σ^2, and assuming a training period of $t = 1, \ldots, T$ as in (1) and (2), Homm and Breitung (2012) propose testing for explosive behavior in the monitoring period using the following CUSUM statistic:

$$S_T^t := \frac{1}{\tilde{\sigma}_t} \sum_{j=T+1}^{t} \Delta y_j \tag{3}$$

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_T^t, is computed sequentially at dates $t = T+1, \ldots, \lfloor \lambda T \rfloor$, then under the null hypothesis, H_0, of no explosive behavior, then for any $\lambda > 1$

$$T^{-1/2}S_T^{\lfloor Tr \rfloor} \Rightarrow W(r) - W(1), \quad 1 < r \leq \lambda \tag{4}$$

and, hence, from Theorem 3.4 of Chu, Stinchcombe, and White (1996) that

$$\lim_{T\to\infty} \Pr(|S_T^t| > c_t\sqrt{t} \text{ for some } t \in \{T+1, \ldots, \lfloor \lambda T \rfloor\}) \leq \exp(-b_\alpha/2), \tag{5}$$

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 $b_\alpha = 4.6$. Henceforth, we will refer to a monitoring procedure based on the S_T^t statistic as the (standard) CUSUM monitoring procedure.

2.2 A Time Varying Volatility-Robust CUSUM Procedure

A major drawback with this CUSUM monitoring procedure in practice is that the variance estimator, $\tilde{\sigma}_t^2$, is only appropriate for the homoskedastic case. More specifically, under heteroskedasticity of the form given in Assumption 2, the result in Equation (4) no longer holds. In particular, under H_0, as shown in Cavaliere (2004, Lemma 1) $T^{-1/2}S_T^{\lfloor Tr \rfloor} \Rightarrow \int_1^r \sigma(s)\mathrm{d}W(s)/\sqrt{\int_0^r \sigma^2(s)\mathrm{d}s}$, $1 < r \leq \lambda$. This limiting process clearly depends on the time-varying volatility process, $\sigma(\cdot)$, reducing to the result in Equation (4) only in the case where $\sigma(\cdot)$ is constant. As a consequence, Theorem 3.4 of Chu, Stinchcombe, and White (1996) can no longer be applied to give the result in Equation (5). Monte Carlo simulation evidence provided in Astill et al. (2018) confirms this asymptotic prediction with the empirical FPR of the CUSUM procedure shown to be severely impacted under various patterns of time-varying volatility. The simulation results we report in Section 3 provide further confirmation of this.

In order to address this problem, we propose robustifying the standard CUSUM procedure to time-varying volatility by modifying the CUSUM statistic in Equation (3) such that each observation on Δy_j in Equation (3) is standardized by a Nadaraya–Watson-type kernel-based nonparametric spot variance estimator as is done in, among others, Hansen (1995), Xu and Phillips (2008), and Beare (2018). Specifically, we consider the following heteroskedasticity-robust version of the CUSUM statistic

$$SV_T^t := \sum_{j=T+1}^{t} \frac{\Delta y_j}{\hat{\sigma}_{j,N}}, \; t > T, \tag{6}$$

where $\hat{\sigma}_{j,N}^2$ is a kernel smoothing estimator for the spot variance $\sigma_j^2 := \sigma^2(j/T)$. The kernel smoothing estimator is defined as follows for $j \geqslant N+1$:

$$\hat{\sigma}_{j,N}^2 := \sum_{s=0}^{N} w_s \Delta y_{j-s}^2, \quad \text{with} \quad w_s := \frac{K\left(\frac{s}{N}\right)}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}, \tag{7}$$

where $K(.)$ is a kernel function and where N denotes the bandwidth (such that the number of observations used in the kernel smoothing is $N+1$). For completeness, we define $\hat{\sigma}_{j,N}^2 =$

$\hat{\sigma}^2_{N+1}$ for $j \leqslant N$, although in practical situations such definitions are not generally required as the point where monitoring begins ($t = T + 1$) is typically (much) larger than the bandwidth parameter N. Henceforth, we will refer to a monitoring procedure based on the modified CUSUM statistic, SV^t_T, as the CUSUM$_V$ monitoring procedure.

We will make the following technical assumptions on the kernel function used in Equation (7):

Assumption 3. *$K(.)$ is continuously differentiable over the interval (0, 1), with $K(x) = 0$ for $x \leq 0$ and $x \geq 1$. Also, $\int_0^1 K(x)\mathrm{d}x > 0$, $\int_0^1 |K(x)|\mathrm{d}x < \infty$, $\int_0^1 |K(x)x|\mathrm{d}x < \infty$ and the characteristic function $\phi(t) = \int_{-\infty}^{\infty} \exp(\mathrm{i}tx)K(x)\mathrm{d}x$ of K satisfies $\int_{-\infty}^{\infty} |\phi(t)|\mathrm{d}t < \infty$. $K'(.)$, the derivative of the $K(.)$ function, also has a characteristic function that is absolutely integrable.*

Remark 4. Our assumption of a one-sided kernel function $K(.)$ which is positive on the unit interval, (0, 1), implies that we are using a rolling window type filter to estimate the spot variance σ^2_j. This one-sided kernel assumption should not be viewed as a restriction, however, because in our real-time monitoring context we would obviously not have access to future data. Allowing for a kernel which is positive on $(0, \infty)$ is technically possible, but we prefer to work with a rolling window using recent lagged data as this has computational advantages in our monitoring context, especially at long monitoring horizons. Notice that our conditions on the kernel function impose that $K(0) = 0$ and $K(1) = 0$, which implies that the current observation is left out when estimating the volatility. These restrictions are necessary for proving our crucial intermediate result in Lemma 2. Examples of kernels which satisfy Assumption 3 include the rectangular, Epanechnikov, Bartlett, and truncated Gaussian kernels (with their boundary values adjusted to 0 in each case).

Before deriving our main result in Theorem 1, in Lemmas 1 and 2 we first provide two important intermediate results relating to the large sample properties of the nonparametric variance estimator, $\hat{\sigma}^2_j$.

Lemma 1. *Let y_t be generated according to Equations (1) and (2) under $H_0 : \delta = 0$ and let Assumptions 1–3 hold. Then, if $T, N \to \infty$ such that $N/T \to 0$ and $N^2/T \to \infty$, then*

$$\max_{T+1 \leqslant j \leqslant \lfloor \lambda T \rfloor} |\hat{\sigma}^2_{j,N} - \sigma^2_j| = o_p(1).$$

Lemma 2. *Let y_t be generated according to Equations (1) and (2) under $H_0 : \delta = 0$ and let Assumptions 1–3 hold. Then if $T, N \to \infty$ such that $N/T \to 0$ and $N^{3/2}/T \to \infty$, then*

$$\max_{T+1 \leqslant j \leqslant \lfloor \lambda T \rfloor} |(\hat{\sigma}^2_{j-1,N} - \sigma^2_{j-1}) - (\hat{\sigma}^2_{j,N} - \sigma^2_j)| = o_p(T^{-1}).$$

Remark 5. Lemma 1 establishes the necessary rate condition on the bandwidth, N, such that our nonparametric variance estimator, $\hat{\sigma}^2_{j,N}$ of Equation (7), is uniformly consistent. Lemma 2 establishes a similar result for the numerical derivative of the nonparametric variance estimator, showing that a stronger rate condition is needed on the bandwidth than was needed for the consistency result in Lemma 1.

Theorem 1. *Let the conditions of Lemma 2 hold. Then,*

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

Remark 6. The result in Theorem 1 demonstrates that for any pattern of time-varying volatility satisfying Assumption 2, the joint limiting null distribution of the sequence of modified CUSUM statistics in Equation (6) coincides with that which obtains for the standard CUSUM statistics in Equation (3) when e_t is homoskedastic. An immediate consequence of this, given in Corollary 1 below, is that we can therefore apply Theorem 3.4 of Chu, Stinchcombe, and White (1996) to obtain a result corresponding to Equation (5) with the implication that our modified CUSUM monitoring procedure will have a controlled theoretical FPR even in the presence of time-varying volatility of the form specified by Assumption 2.

Remark 7. The proof strategy used to establish the result in Theorem 1 extends the approach used in Beare (2018), which is based on a uniform consistency result for the nonparametric variance estimator and its derivative (cf., Lemma 4.1 and Lemma 4.2 of Beare, 2018). By construction, our nonparametric variance estimator $\hat{\sigma}^2_{\lfloor sT \rfloor,N}$ is not differentiable with respect to s and, hence, the proof strategy used in Beare (2018) is not directly applicable. We therefore extend the proof strategy used in Beare (2018) and build our proof upon a corresponding result for the numerical derivative of $\hat{\sigma}^2_{\lfloor sT \rfloor,N}$, as given in Lemma 2.

By Theorem 3.4 of Chu, Stinchcombe, and White (1996), Theorem 1 implies the following result for SV_T^t.

Corollary 1. *Under the same conditions as Theorem 1*

$$\lim_{T\to\infty} \Pr(|SV_T^t| > c_t\sqrt{t} \text{ for some } t \in \{T+1, \ldots, \lfloor \lambda T \rfloor\}) \leq \exp(-b_\alpha/2).$$

Remark 8. The result given in Corollary 1 has two main implications. First, in the case where the innovations are homoskedastic, using the nonparametric spot volatility estimator $\hat{\sigma}_j$ leads to the same limiting null distribution and crossing probabilities for both the CUSUM and CUSUM$_V$ procedures. Second, and more importantly, when the innovations display time-varying volatility of the form outlined in Assumption 2, both the limiting null distribution and crossing probability for the CUSUM$_V$ procedure are unchanged relative to those which obtain in the homoskedastic case. This stands in contrast to the CUSUM procedure which requires homoskedasticity for Equation (4), and therefore Equation (5), to hold. The CUSUM$_V$ procedure therefore offers robustness to time varying volatility in a way that the standard CUSUM procedure does not.

Remark 9. It is interesting to note the asymptotic results given in Theorem 1 and Corollary 1 remain valid in the case where a finite number of level breaks are present in the DGP for y_t in Equation (1). This holds by virtue of the fact that SV_T^t is calculated using the first differences of y_t. Consequently, any level breaks present in the levels data, y_t, are transformed to one-period outliers in the first differenced data which therefore have no impact on the large sample properties of SV_T^t.

Remark 10. The model in Equations (1) and (2) allows for a non-zero mean in y_t through the presence of μ in Equation (1). In most applications of tests for explosive episodes,

allowing for a non-zero mean in y_t is considered sufficient. However, in some circumstances it may be desirable to allow for the possibility that the expected value of y_t follows a linear trend. In this case, we would replace μ in Equation (1) by $\mu + \beta t$. As discussed in Homm and Breitung (2012, p. 212), the CUSUM-type procedures discussed in Section 2.1 can be modified to allow for the presence of a linear trend in Equation (1) by replacing Δy_j in the CUSUM statistic in Equation (3) with the standardized recursive residuals $\sqrt{\frac{j-1}{j}}(\Delta y_j - \hat{\mu}_{j-1})$ and redefining the variance estimator as $\tilde{\sigma}_t^2 := (t-2)^{-1} \sum_{j=2}^{t} (\Delta y_j - \hat{\mu}_t)^2$ where, for any $k \geq 1$, $\hat{\mu}_k := (k-1)^{-1} \sum_{l=2}^{k} \Delta y_l$. The large sample properties of the resulting sequence of CUSUM statistics are unchanged from that given in Section 2.1. The heteroskedasticity-robust CUSUM statistic defined in Equation (6) can be similarly modified to allow for a linear trend by again replacing Δy_j in the numerator of Equation (6) by $\sqrt{\frac{j-1}{j}}(\Delta y_j - \hat{\mu}_{j-1})$ and replacing Δy_{j-s}^2 in Equation (7) by $(\Delta y_{j-s} - \hat{\mu}_j)^2$, and similarly in Equations (8) and (9) in Section 2.3; again this will not affect the large sample results stated in Section 2.2. We repeated all of the simulation experiments reported in Section 3 (which only allow for a non-zero mean) with this correction for a linear trend implemented and found these results to be almost identical to those reported. These results are available on request.

Remark 11. Thus far e_t in Equation (2) has been assumed serially uncorrelated. This can be weakened to allow for the case where e_t admits a finite-order autoregression of the form $\phi(L)e_t = \sigma_t \varepsilon_t$ with $\phi(z) := 1 - \sum_{j=1}^{p} \phi_j z^j$, such that $\phi(z) \neq 0$ for all $|z| \leq 1$, and where ε_t and σ_t continue to satisfy Assumptions 1 and 2, respectively. The CUSUM-type procedures discussed in Sections 2.1 and 2.2 can be modified to allow for such serial correlation by using pre-whitening. This is done by replacing Δy_j in Equations (3) and (6) by $\tilde{e}_j := \Delta y_j - \hat{\phi}_{1,j} \Delta y_{j-1} - \cdots - \hat{\phi}_{p,j} \Delta y_{j-p}$, where $\hat{\phi}_{i,j}$, $i = 1, \ldots, p$, are the OLS autoregressive lag estimates from the prewhitening regression of Δy_s on $\{\Delta y_{s-i}\}_{i=1}^{p}$, over the sample data $s = p+2, \ldots, j$. Similarly, redefine the variance estimator used in connection with Equation (3) as $\tilde{\sigma}_t^2 := (t-2p-1)^{-1} \sum_{j=p+2}^{t} (\Delta y_j - \hat{\phi}_{1,t} \Delta y_{j-1} - \cdots - \hat{\phi}_{p,t} \Delta y_{j-p})^2$. With regard to the kernel-based estimator in Equation (7) we need to replace Δy_{j-s}^2 by $(\Delta y_{j-s} - \hat{\phi}_{1,j} \Delta y_{j-s-1} - \cdots - \hat{\phi}_{p,j} \Delta y_{j-s-p})^2$, $s = 0, \ldots, N$, and similarly in the associated cross-validation criteria in Equations (8) and (9) in Section 2.3.[1] If the lag order, p, is known, then the estimates of the autoregressive lag coefficients defined above are consistent at rate $T^{1/2}$ under H_0; see Phillips and Xu (2006). As a result, we conjecture that the limiting results given previously will continue to hold under this modification. In most

1 In fact, provided e_t is β-mixing with exponentially decaying coefficients, then the kernel-based variance estimators discussed in this section, which are based on smoothing squared differences of data, remain consistent without the need for pre-whitening. According to the discussion in Fan and Yao (2003, p. 69), the AR(p) case discussed here satisfies this mixing condition provided ε_t is an IID sequence of continuously distributed random variables.

practical applications a very low autoregressive order, either $p=0$ or $p=1$, is typically assumed. In practice, p could in principle be determined using any consistent model selection criterion, an obvious example being the Bayesian information criterion (BIC) of Schwarz (1978).[2]

The next theorem establishes the large sample behavior of SV_T^t under the alternative hypothesis:

Theorem 2. *Under the same conditions as Theorem 1, but under* $H_1 : \delta > 0$,

$$\lim_{T\to\infty} \Pr(|SV_T^t| > c_t\sqrt{t}, \text{ for some } t \in \{T+1,\ldots,\lfloor \lambda T \rfloor\}) = 1.$$

The result in Theorem 2 demonstrates that our modified CUSUM monitoring procedure is consistent under H_1, rejecting the false null of no explosivity with probability one in the limit.

2.3 Implementation Issues: Bandwidth and Kernel Selection

The practical implementation of SV_T^t requires choices to be made for both the kernel and bandwidth used in constructing the nonparametric estimator $\hat{\sigma}_{j,N}^2$. We will now discuss these two choices, providing recommendations for each.

In general, the choice of kernel tends to be much less crucial for the finite sample performance of nonparametric kernel-based estimators than is the bandwidth, and we found this general rule to hold true for our particular nonparametric estimator, $\hat{\sigma}_{j,N}^2$. We conducted finite sample simulations using a number of kernels (in particular the boundary value adjusted truncated Gaussian, rectangular, Epanechnikov, and Bartlett kernels), and found little difference between the empirical FPR and TPR profiles of our proposed monitoring procedure across these different choices. Throughout the remainder of the paper, results are reported for the truncated Gaussian kernel; results for the other kernels mentioned above are available on request.

In practice, it is the choice of bandwidth that is crucial to the performance of nonparametric estimators such as $\hat{\sigma}_{j,N}^2$. Other things being equal, adopting too large a bandwidth results in oversmoothing which leads to increased bias in the volatility estimator, while using too small a bandwidth results in undersmoothing which leads to an increased variance in the resulting volatility estimator, both of which will have a detrimental impact on the empirical FPR and TPR of the resulting procedure. As is commonly done in the literature, we adopt a data-driven method for selecting the bandwidth in order to automate the decision on how to trade off the bias and variance of the estimator. To this end, we propose selecting the bandwidth according to a standard cross-validation procedure. Specifically, for a given time period in the monitoring period $t = T+1,\ldots,\lfloor \lambda T \rfloor$, first define the cross validation criterion

2 A linear trend can also be allowed for as described in Remark 10, by analogous demeaning of the prewhitened residuals. Here, an intercept also needs to be included in the pre-whitening regression.

$$CV_t(N) := \frac{1}{t-T}\sum_{j=T+1}^{t}(\hat{\sigma}^2_{j,N} - \Delta y_j^2)^2. \tag{8}$$

The $CV_t(N)$ criterion is essentially an estimate of the mean integrated squared error (MISE) of the variance estimator for a given bandwidth N. The automated bandwidth, denoted N_t^{cv}, is then chosen to be the bandwidth that minimizes the (estimated) MISE and is therefore defined as $N_t^{cv} := \text{argmin}_N CV_t(N)$. Following Härdle, Hall, and Marron (1988), $N_t^{cv} = O(T^{4/5})$, and so this choice of bandwidth easily satisfies the rate restriction placed on the bandwidth in Theorem 1.

The MISE-minimizing cross-validation procedure discussed above can be interpreted as a "global" procedure in that it attempts to minimize the error we make in estimating the spot variance from time $j = T+1$ to the current period $j = t$. In the context of monitoring, however, minimizing $CV_t(N)$ across the full range of $\hat{\sigma}^2_{j,N}$, $j = T+1, \ldots, t$, may not be appropriate. What is important is not how well we estimate the spot variance using the entire monitoring period, but how well this variance is estimated in the immediate neighborhood of the monitoring time period t. As such, we may also consider a "local" cross-validation procedure where $CV_t(N)$ is instead defined as

$$CV_t^*(N) := \frac{1}{H}\sum_{j=t-H+1}^{t}(\hat{\sigma}^2_{j,N} - \Delta y_j^2)^2 \tag{9}$$

and we select the bandwidth, denoted N_t^{cv}, according to $N_t^{cv*} := \text{argmin}_{N\in[1,H]} CV_t^*(N)$ so that the bandwidth is instead selected to minimize the estimation error of the spot variance over the most recent H observations; cf., Hall and Schucany (1989). We recommend the use of the automated bandwidth N_t^{cv*} in practice and will use this choice in both our numerical simulations and empirical exercise. Implementation of N_t^{cv*} requires the user to make a choice for the tuning parameter, H; this will be further explored in Section 3.

3 Finite-Sample Simulations

In this section, we compare the finite sample performance of our proposed CUSUM_V monitoring procedure with the standard CUSUM monitoring procedure of Homm and Breitung (2012) and also with the MAX_m monitoring procedure of Astill et al. (2018); the latter is, like CUSUM_V, robust to the presence of time-varying volatility in the errors.

The MAX_m monitoring procedure of Astill et al. (2018) is based on the sequential application of the statistic

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

where m is a user chosen window width. The MAX_m procedure then 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 $S_{t,m}$ exceeds the maximum value across the corresponding sequence of statistics $S_{j,m}$, $j = m+1, \ldots, T-m+1$, calculated over the training period; that is, $H_0 : \delta = 0$ is

rejected if $\max_{t\in[T+1,\lfloor\lambda T\rfloor]} S_{t,m} > \max_{t\in[m+1,T-m+1]} S_{t,m}$. Astill et al. (2018) demonstrate that an approximation to the FPR of this procedure is given by

$$\alpha := \frac{\lfloor\lambda T\rfloor - T}{\lfloor\lambda T\rfloor - 2m + 1}. \tag{10}$$

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 Equation (10) by replacing $\lfloor\lambda T\rfloor$ with t. Observe that this FPR is a function of the length of the training period, T and the window width, m, used in the $S_{t,m}$ statistics.

In our simulations, data were generated according to the DGP Equations (1) and (2) with $\varepsilon_t \sim \text{NIID}(0,1)$, setting $\mu = 0$ without loss of generality, and with the DGP initialized at $u_0 = 100$ so that generated data remain positive and thereby any explosive episodes will typically appear as having upward trajectories, thus mimicking what is observed in an asset price bubble. We assume that monitoring begins at time $t = 220$ and set the training period sample size to $T = 219$. We set $\lambda T = 255$, such that we have a (maximum) monitoring period of 36 observations. For the MAX_m procedure, we use a single window width of $m = 10$ as recommended by Astill et al. (2018). By assuming a common monitoring start date for all procedures of $T = 220$, we treat the sample $t = 1,\ldots,T$ as the training period for the CUSUM and CUSUM$_V$ procedures, while the training period for the MAX_{10} procedure is given by the sample $t = 1,\ldots,T-m+1$. Homm and Breitung (2012) show that the asymptotic critical values for the CUSUM procedures implied by Equation (5) are very conservative in practice, being based on an assumption of a monitoring period of infinite length. As such, and to aid comparison with the MAX_{10} procedure, finite sample critical values for the CUSUM and CUSUM$_V$ monitoring procedures were obtained by choosing a value of b_α such that for a homoskedastic DGP ($\sigma_t = 1$), the empirical FPR of these procedures is equal to the FPR of the MAX_{10} procedure, determined by Equation (10), when the latter has a FPR of 0.10. Therefore, in the simulations that follow the CUSUM and CUSUM$_V$ monitoring procedures were performed using $b_\alpha = 0.147$ and $b_\alpha = 0.177$, respectively. All simulations were conducted in Gauss 9.0 using 10,000 replications.

The bandwidth used in connection with the kernel-based spot variance estimator used in the $CUSUM_V$ procedure was selected at each point in the monitoring period using the local cross validation procedure in Equation (9). We experimented with the tuning parameter H and found the robustness of the empirical FPR of the CUSUM$_V$ procedure to time-varying volatility to be decreasing in the choice of H, whereas the empirical TPR of the procedure was found to be increasing in H. Our experiments suggested that setting $H = 20$ delivered a procedure with the best trade-off between these two considerations and we will use this choice in all of the numerical and empirical work that follows. As discussed in Section 2.3, the reported results are for a truncated Gaussian kernel.

3.1 Empirical FPRs under $H_0 : \delta = 0$

We first simulate the FPRs of the CUSUM, CUSUM$_V$, and MAX_{10} procedures under the null hypothesis $H_0 : \delta = 0$ for cases where: (i) the errors e_t are homoskedastic ($\sigma_t = 1$) and (ii) where the errors exhibit time-varying volatility. In the latter case, we first consider smooth shifts in volatility of the form

$$\sigma_t := 1 + a[1 + \exp(-\theta(t - T_b))]^{-1} \tag{11}$$

that is, a logistic smooth transition in volatility from 1 to $1 + a$ when $\theta > 0$, and from $1 + a$ to 1 when $\theta < 0$. In each case, the transition speed and timing of the transition are governed by $|\theta|$ and T_b, respectively.

Figure 1 reports the empirical FPRs of the three procedures when a smooth shift in volatility occurs, with $|\theta| = 0.25$ with $T_b = T$ so that the transition is centered around the starting date of the monitoring period. For each time point e, $T + 1 \leq e \leq \lfloor T\lambda \rfloor$, the corresponding point on the curves in the figure represents the empirical FPR of the particular procedure run from time $t = T + 1$ until time $t = e$. We consider the cases $a = \{0, \sqrt{2} - 1, \sqrt{3} - 1, \sqrt{4} - 1\}$; here, $a = 0$ represents the benchmark case of homoskedasticity (in which case the value of θ is irrelevant), while for $a \neq 0$, the variance changes (increases) from 1 to 2, 3, or 4 when $\theta = 0.25$, and (decreases) from 2, 3, or 4 to 1 when $\theta = -0.25$. The red vertical dashed line on each graph represents the time at which the FPRs of the CUSUM, CUSUM$_V$, and MAX_{10} procedures are equal to 0.10 for the case of a homoskedastic DGP.

Figure 1(a) reports the homoskedastic case ($a = 0$) and here all three procedures have very similar FPR profiles across the range of end-of-monitoring dates e, $T + 1 \leq e \leq \lfloor T\lambda \rfloor$. Figure 1(b)–(d) reports the FPR of the procedures when an upward shift in volatility occurs with $\theta = 0.25$. The FPR of the CUSUM procedure is seen to be inflated to a large degree relative to the homoskedastic case; for the largest value of a considered this FPR exceeds 0.33 at time $e = 241$, this being the point at which the procedure is calibrated to have an FPR of 0.10 under homoskedasticity. The FPR of the CUSUM$_V$ procedure is inflated to some extent relative to the case where the errors are homoskedastic, but to nowhere near the extent of the CUSUM procedure. For the largest value of a considered, the FPR of the CUSUM$_V$ procedure at time $e = 241$ is about 0.13. The FPR of the MAX_{10} procedure is barely impacted by any shifts in volatility. Figure 1(e)–(g) reports the FPR of the procedures for the smooth downward shift in volatility cases with $\theta = -0.25$. The FPR of the CUSUM procedure is severely deflated relative to the homoskedastic case, with this feature again most apparent for larger values of a. For the largest a considered, the FPR of the CUSUM procedure does not exceed 0.05 even by the very end of the monitoring period. The FPR of the CUSUM$_V$ procedure is slightly deflated relative to the homoskedastic case, but again this is modest in comparison with the CUSUM procedure. As with upward volatility shifts, the FPR of the MAX_{10} procedure is little affected.

We next consider cases where the volatility shift is centered around a date before or after the monitoring period commences. Figure 2 reports the FPR of the three procedures when $T_b = T + 10$ so that the mid-point of the volatility shift occurs shortly after the start of monitoring. Figure 2(a) is a repeat of Figure 1(a) for reference purposes, while Figure 2(b)–(d) reports the FPRs when the volatility shift is upward. The relative FPR inflation exhibited by the test procedures is broadly similar to the case where the smooth volatility shift is centered around the start of the monitoring period, with the CUSUM procedure displaying the largest degree of FPR inflation and the CUSUM$_V$ procedure again displaying only a modest degree of FPR inflation relative to the homoskedastic case. The FPR of the MAX_{10} procedure is again almost unchanged from the homoskedastic case. Figure 2(e)–(g) reports the FPRs for the cases of downward shifts in volatility. Again we see broad similarity with the corresponding panels of Figure 1. Finally, Figure 3 reports results for $T_b = T - 10$ so that

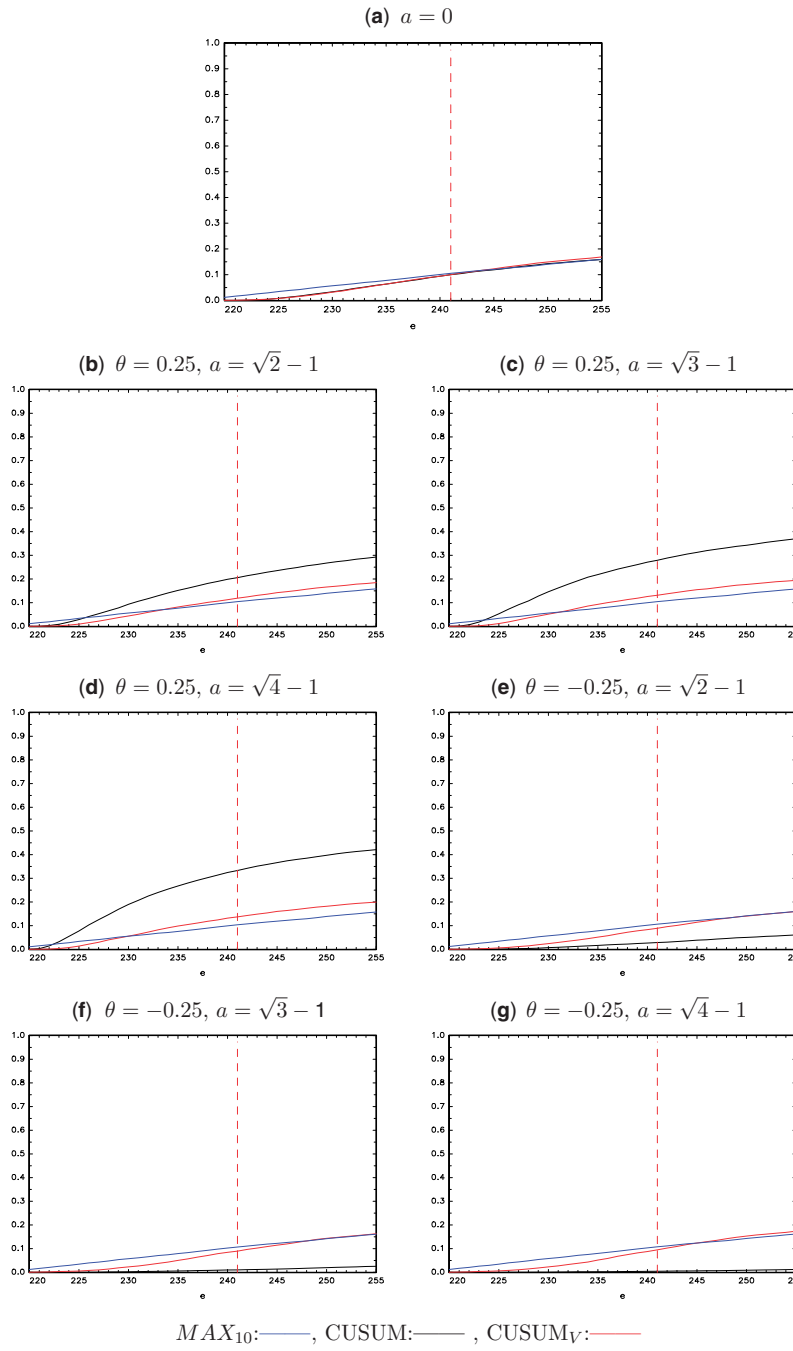

Figure 1. FPR—smooth volatility shift ($T_b = T$). (a) $a = 0$. (b) $\theta = 0.25$, $a = \sqrt{2} - 1$. (c) $\theta = 0.25$, $a = \sqrt{3} - 1$. (d) $\theta = 0.25$, $a = \sqrt{4} - 1$. (e) $\theta = -0.25$, $a = \sqrt{2} - 1$. (f) $\theta = -0.25$, $a = \sqrt{3} - 1$. (g) $\theta = -0.25$, $a = \sqrt{4} - 1$.

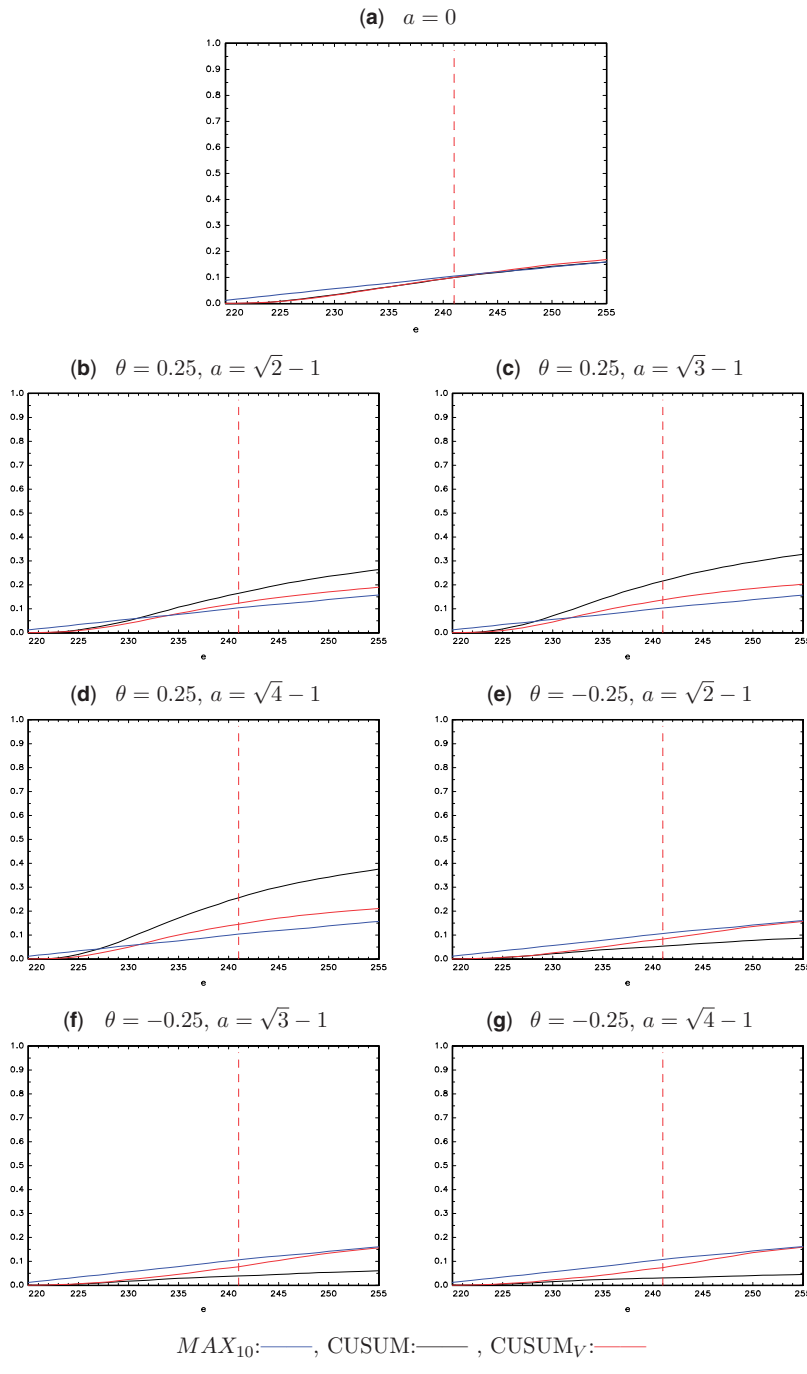

Figure 2. FPR—smooth volatility shift ($T_b = T + 10$). (a) $a = 0$. (b) $\theta = 0.25$, $a = \sqrt{2} - 1$. (c) $\theta = 0.25$, $a = \sqrt{3} - 1$. (d) $\theta = 0.25$, $a = \sqrt{4} - 1$. (e) $\theta = -0.25$, $a = \sqrt{2} - 1$. (f) $\theta = -0.25$, $a = \sqrt{3} - 1$. (g) $\theta = -0.25$, $a = \sqrt{4} - 1$.

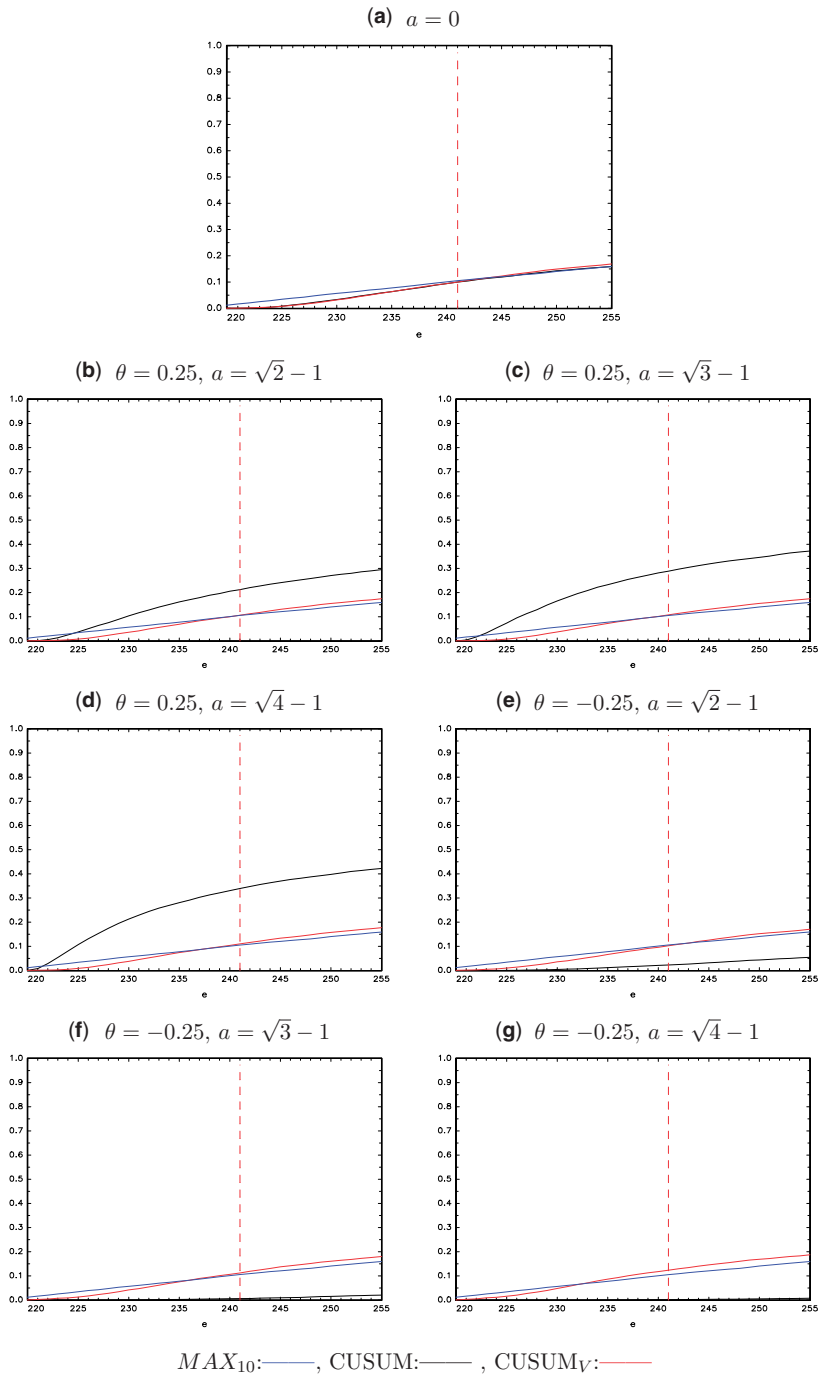

Figure 3. FPR—smooth volatility shift ($T_b = T - 10$). (a) $a = 0$. (b) $\theta = 0.25$, $a = \sqrt{2} - 1$. (c) $\theta = 0.25$, $a = \sqrt{3} - 1$. (d) $\theta = 0.25$, $a = \sqrt{4} - 1$. (e) $\theta = -0.25$, $a = \sqrt{2} - 1$. (f) $\theta = -0.25$, $a = \sqrt{3} - 1$. (g) $\theta = -0.25$, $a = \sqrt{4} - 1$.

the volatility shift mid-point occurs shortly before the commencement of monitoring. Once again, the FPR of the CUSUM procedure is severely impacted, with a large degree of FPR inflation or deflation exhibited for $\theta = 0.25$ and $\theta = -0.25$, respectively. The FPR of the CUSUM_V procedure is less impacted by a smooth shift in volatility centered at this point in time than for shifts centered at dates where monitoring has already commenced, with this result explained by the fact that volatility shifts that occur before the commencement of monitoring allow greater time for the spot variance estimate in Equation (7) to adapt to the transitioning volatility path. Again, the FPR of the MAX_{10} procedure is very similar to the homoskedastic case.

3.2 Empirical TPRs under $H_1 : \delta > 0$

We now turn to an examination of the empirical TPRs of the CUSUM, CUSUM_V, and MAX_{10} procedures to detect an emergent explosive episode in the monitoring period. We initially concentrate on the homoskedastic case $a = 0$. We will consider two possible starting dates for the explosive regime in Equations (1) and (2), namely $\lfloor \tau_1 T \rfloor = \{220, 230\}$ and generate explosive regimes of length 25 observations, so that $\lfloor \tau_2 T \rfloor = \lfloor \tau_1 T \rfloor + 25$. We also vary the magnitude of the offset to the autoregressive parameter driving the explosive regime by considering the settings $\delta = \{0.004, 0.006, 0.008, 0.010\}$. The results are reported in Figure 4. Henceforth, the time periods over which an explosive regime is present are identified by gray shaded areas in each figure. As would be expected, the best overall TPR profile is displayed by the CUSUM procedure, as this procedure is specifically calibrated for data generated from the homoskedastic case. The TPR of the CUSUM_V procedure is, encouragingly, very close to that of the CUSUM procedure, so it appears that the FPR robustness of the CUSUM_V procedure to time-varying volatility in the errors does not come at the expense of significantly reduced power to detect an explosive regime. The TPR profile of the CUSUM_V procedure is far superior overall than that of the MAX_{10} procedure, with the TPR of MAX_{10} only marginally higher than that of the CUSUM-based procedures for a small number of observations at the beginning of the explosive episode, at which point all of the TPRs are still very close to the corresponding FPRs.

We next consider the TPRs of the three procedures when a smooth volatility shift is present in the data. Figure 5 reports the TPRs when the volatility shift is upward and centered at the commencement of the explosive episode; this timing ensures that the start of the explosive episode coincides with the middle of the volatility transition, a situation that is arguably of substantial empirical relevance given that periods of explosivity are often accompanied by large changes in volatility. Specifically, we set $\lfloor \tau_1 T \rfloor = \{220, 230\}$, $\lfloor \tau_2 T \rfloor = \lfloor \tau_1 T \rfloor + 25$, $\theta = 0.25$, and $T_b = \lfloor \tau_1 T \rfloor$, focusing on the case $\delta = 0.007$ (the average of the set of δ values considered in Figure 4). Results are reported for the same set of values of a as considered in Figures 1–3 (including the homoskedastic case of $a = 0$). When a smooth upward volatility shift occurs the TPR of the CUSUM procedure is much higher than the other two procedures, but this result is of course purely an artifact of the significant FPR inflation exhibited by the CUSUM procedure when such a volatility shift is present in the data. Of the two procedures which offer broad FPR robustness to an upward volatility shift, namely CUSUM_V and MAX_{10}, it is the CUSUM_V procedure which exhibits by far the superior TPR profile.

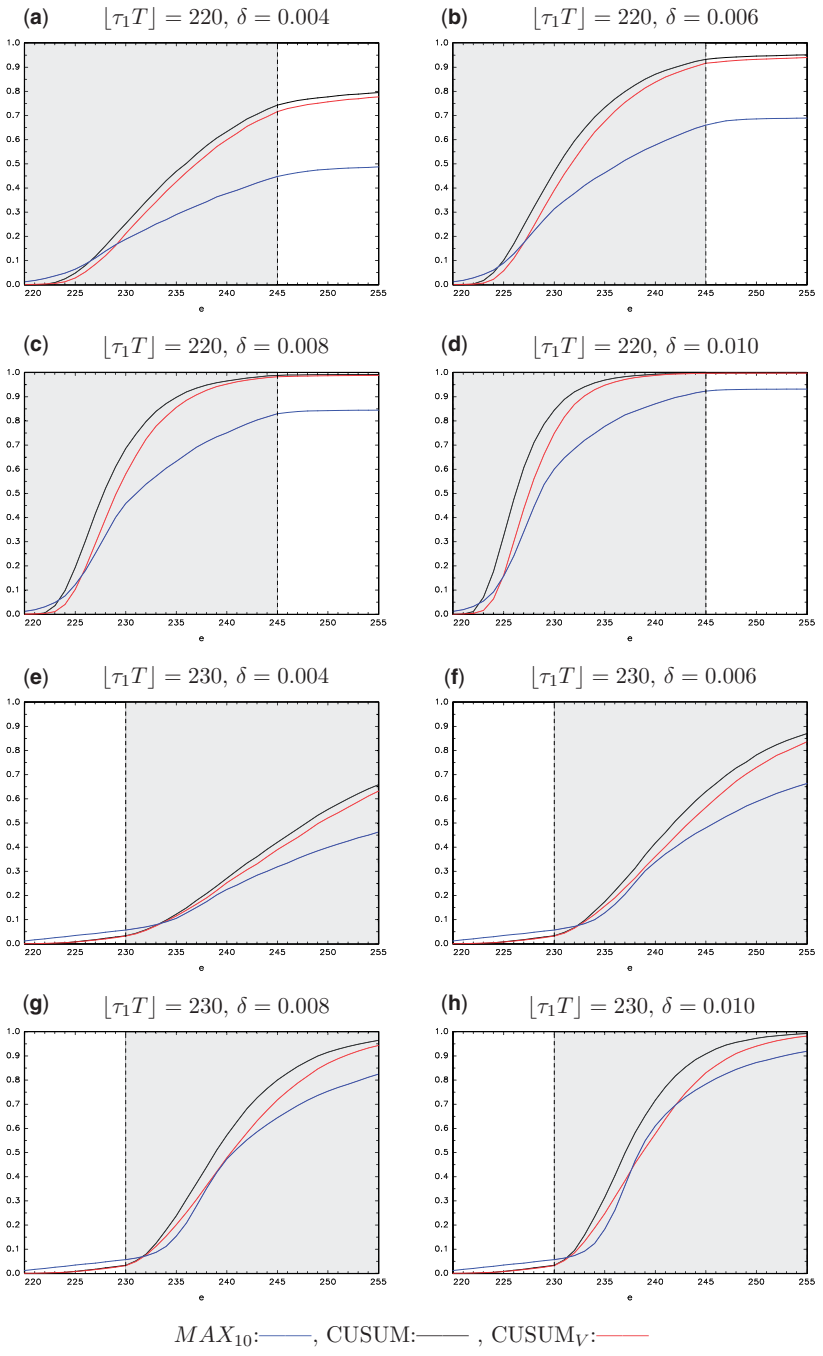

Figure 4. TPR—$a = 0$. (a) $\lfloor \tau_1 T \rfloor = 220$, $\delta = 0.004$. (b) $\lfloor \tau_1 T \rfloor = 220$, $\delta = 0.006$. (c) $\lfloor \tau_1 T \rfloor = 220$, $\delta = 0.008$. (d) $\lfloor \tau_1 T \rfloor = 220$, $\delta = 0.010$. (e) $\lfloor \tau_1 T \rfloor = 230$, $\delta = 0.004$. (f) $\lfloor \tau_1 T \rfloor = 230$, $\delta = 0.006$. (g) $\lfloor \tau_1 T \rfloor = 230$, $\delta = 0.008$. (h) $\lfloor \tau_1 T \rfloor = 230$, $\delta = 0.010$.

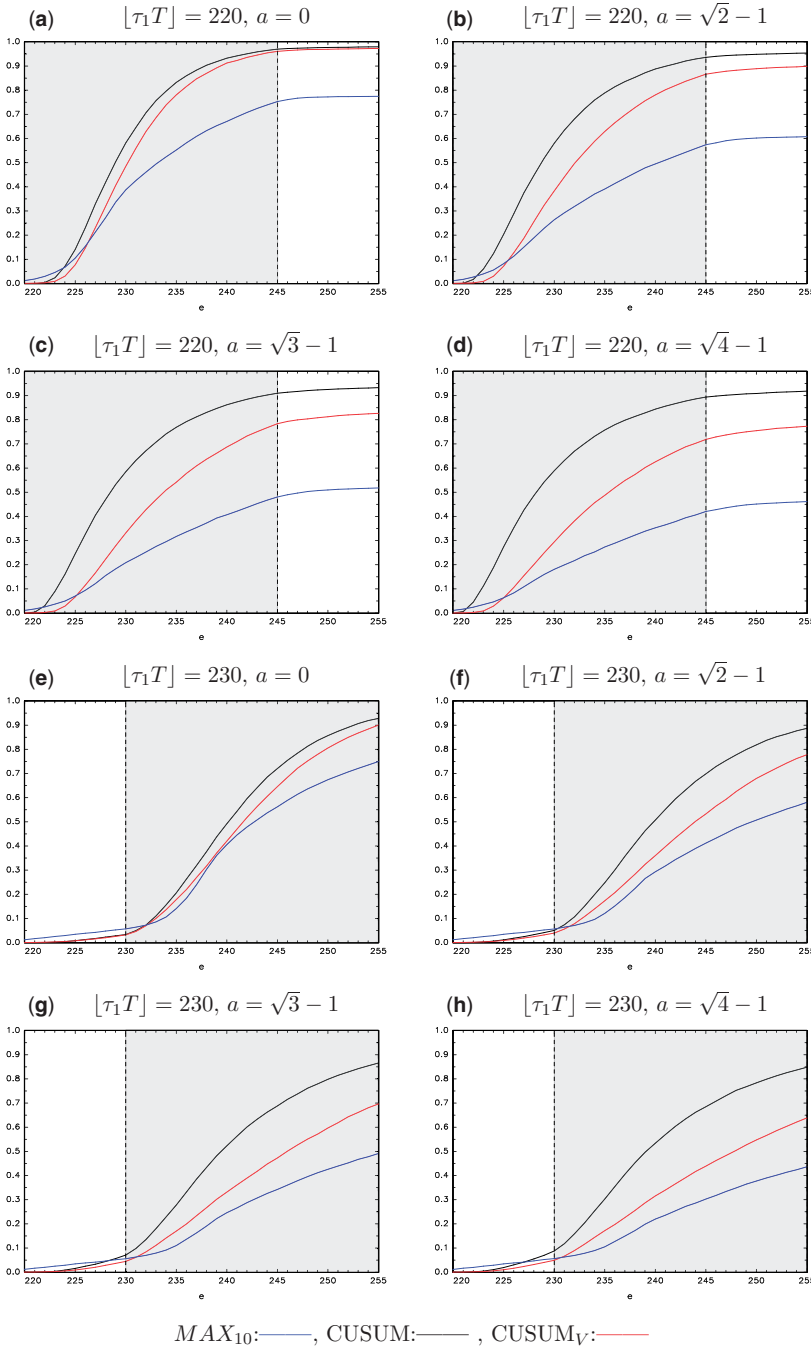

Figure 5. TPR—smooth upward volatility shift. (a) $\lfloor \tau_1 T \rfloor = 220$, $a = 0$. (b) $\lfloor \tau_1 T \rfloor = 220$, $a = \sqrt{2} - 1$. (c) $\lfloor \tau_1 T \rfloor = 220$, $a = \sqrt{3} - 1$. (d) $\lfloor \tau_1 T \rfloor = 220$, $a = \sqrt{4} - 1$. (e) $\lfloor \tau_1 T \rfloor = 230$, $a = 0$. (f) $\lfloor \tau_1 T \rfloor = 230$, $a = \sqrt{2} - 1$. (g) $\lfloor \tau_1 T \rfloor = 230$, $a = \sqrt{3} - 1$. (h) $\lfloor \tau_1 T \rfloor = 230$, $a = \sqrt{4} - 1$.

Figure 6 reports the TPRs for the case of a smooth downward shift in volatility, using the same settings as for Figure 5 but with $\theta = -0.25$. Here, we observe that the TPR of the CUSUM procedure is severely impacted, which is again a consequence of its corresponding FPR deflation in the case of a downward volatility shift, with this impact being more pronounced the larger is the value of a. The best overall TPR profile is arguably displayed by the CUSUM_V procedure, although as a increases the region where the MAX_{10} procedure offers TPR advantages over CUSUM_V for observations early in the explosive regime become somewhat more pronounced; this feature is in line with the modest FPR deflation exhibited by the CUSUM_V procedure in the presence of a downward volatility shift.

3.3 The Impact of an Explosive Episode in the Training Period

We next assess the impact that a single collapsed explosive episode in the training period has on both the empirical FPR and empirical TPR of the CUSUM, CUSUM_V, and MAX_{10} detection procedures. To that end, data were generated according to $y_t = u_t$ with

$$u_t = \begin{cases} u_{t-1} + e_t, & t = 1, \ldots, \lfloor \tau_{1,p} T \rfloor, \\ (1+\delta_p)u_{t-1} + e_t, & t = \lfloor \tau_{1,p} T \rfloor + 1, \ldots, \lfloor \tau_{2,p} T \rfloor, \\ u_{\lfloor \tau_{1,p} T \rfloor} + e_t & t = \lfloor \tau_{2,p} T \rfloor + 1, \\ u_{t-1} + e_t, & t = \lfloor \tau_{2,p} T \rfloor + 2, \ldots, T, \\ u_{t-1} + e_t, & t = T+1, \ldots, \lfloor \tau_1 T \rfloor, \\ (1+\delta)u_{t-1} + e_t, & t = \lfloor \tau_1 T \rfloor + 1, \ldots, \lfloor \tau_2 T \rfloor, \\ u_{t-1} + e_t, & t = \lfloor \tau_2 T \rfloor + 1, \ldots, \lfloor \lambda T \rfloor \end{cases} \tag{12}$$

and $e_t = \varepsilon_t \sim \text{NIID}(0,1)$, thereby focusing on the homoskedastic case (cf., $a = 0$ in the previous subsections). The series y_t therefore admits a single collapsed explosive episode in the training period of length $\lfloor \tau_{2,p} T \rfloor - \lfloor \tau_{1,p} T \rfloor := l_p$ driven by an explosive offset of δ_p. A further explosive episode will occur in the monitoring period if $\delta > 0$.

Figure 7 reports the empirical FPRs of the procedures in the case where $\lfloor \tau_{1,p} T \rfloor = 95$ for two possible lengths, $l_p = \{10, 15\}$ and four explosive offsets $\delta_p = \{0.004, 0.006, 0.008, 0.010\}$. In all cases, the FPR of the CUSUM_V procedure is seen to be unaffected by the presence of these explosive episodes in the training period. This is because the local cross-validation procedure in Equation (9) used to select the bandwidth entails that these explosive observations from the training period receive zero weight in the construction of the spot variance estimator used in the CUSUM_V monitoring statistics. By contrast, the empirical FPRs of both the CUSUM and MAX_V procedures are very significantly deflated when a training period explosive episode is present, as is clear from a comparison of Figure 7 with Figure 1(a). The impact of an explosive episode in the training period is seen to be greater the longer the length of that training period episode and the larger the explosive offset driving this episode.

Figure 8 reports the TPRs of the three procedures to detect an explosive episode in the monitoring period in the case where an explosive episode occurred during the training period. We report results for a single setting for the training period episode for $l_p = 15$ and $\delta_p = 0.007$; qualitatively similar patterns emerge for other settings. The explosive episode generated in the monitoring period for this setting are set to be identical to those reported in Figure 4. A comparison of the results in Figure 8 with those in Figure 4 shows that the

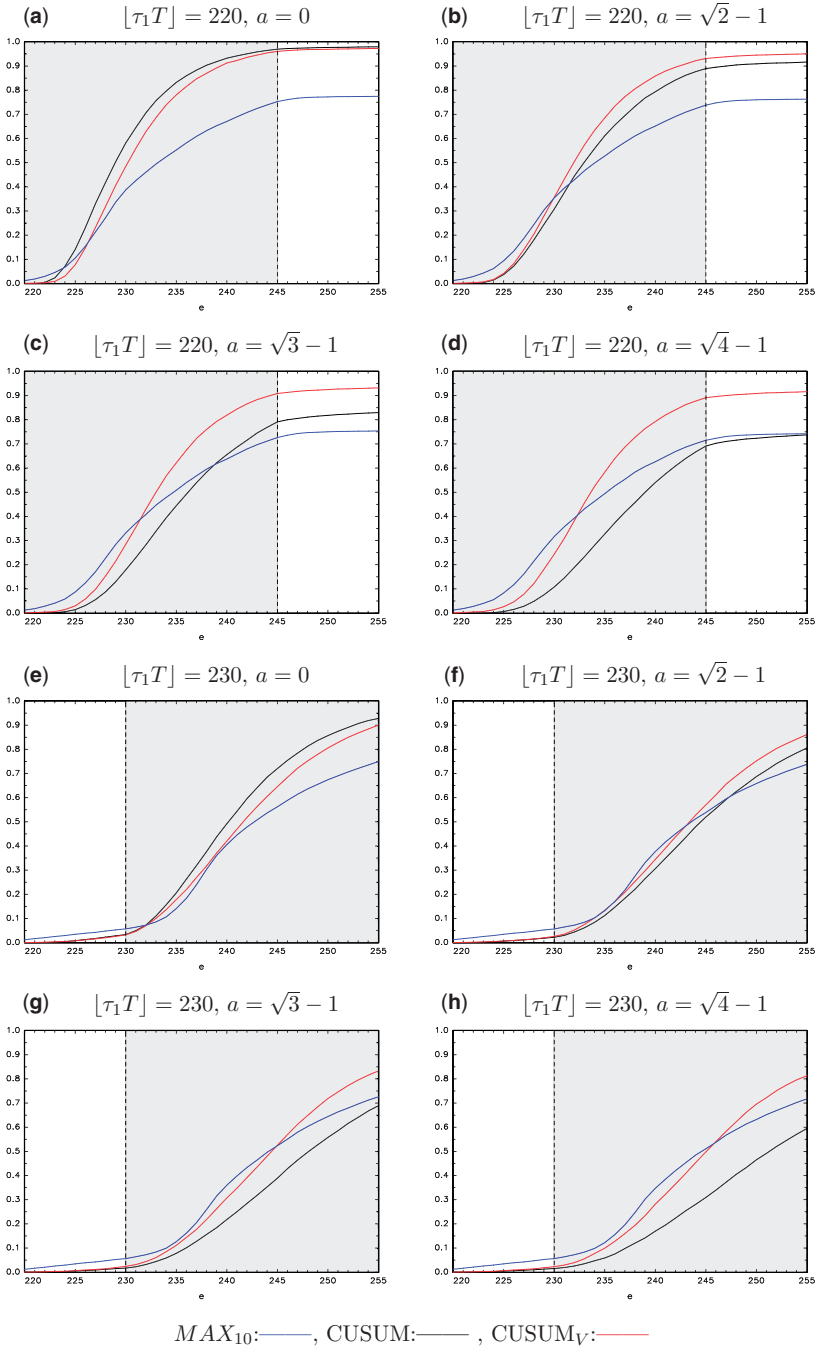

Figure 6. TPR—smooth downward volatility shift. (a) $\lfloor \tau_1 T \rfloor = 220$, $a = 0$. (b) $\lfloor \tau_1 T \rfloor = 220$, $a = \sqrt{2} - 1$. (c) $\lfloor \tau_1 T \rfloor = 220$, $a = \sqrt{3} - 1$. (d) $\lfloor \tau_1 T \rfloor = 220$, $a = \sqrt{4} - 1$. (e) $\lfloor \tau_1 T \rfloor = 230$, $a = 0$. (f) $\lfloor \tau_1 T \rfloor = 230$, $a = \sqrt{2} - 1$. (g) $\lfloor \tau_1 T \rfloor = 230$, $a = \sqrt{3} - 1$. (h) $\lfloor \tau_1 T \rfloor = 230$, $a = \sqrt{4} - 1$.

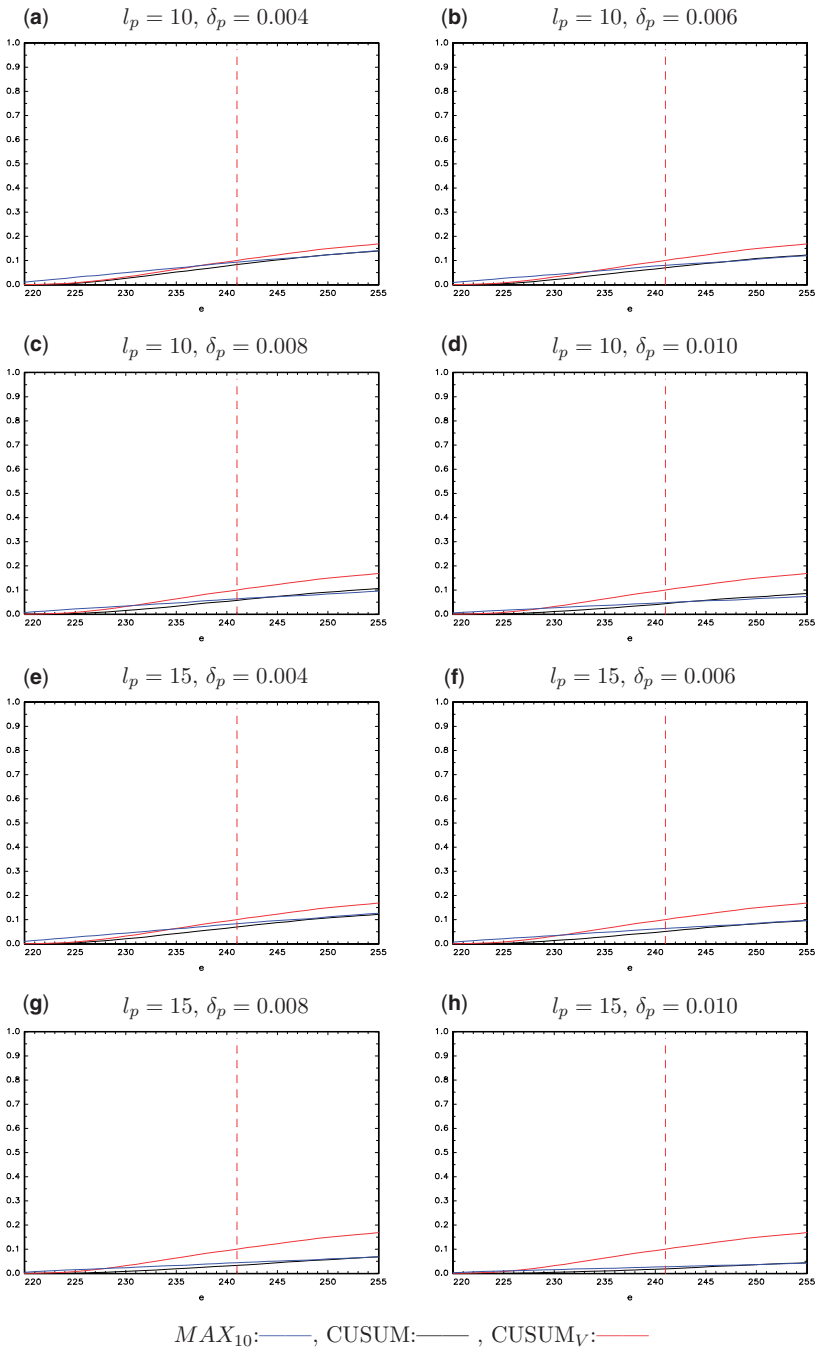

Figure 7. FPR—past bubble. (a) $l_p = 10$, $\delta_p = 0.004$. (b) $l_p = 10$, $\delta_p = 0.006$. (c) $l_p = 10$, $\delta_p = 0.008$. (d) $l_p = 10$, $\delta_p = 0.010$. (e) $l_p = 15$, $\delta_p = 0.004$. (f) $l_p = 15$, $\delta_p = 0.006$. (g) $l_p = 15$, $\delta_p = 0.008$. (h) $l_p = 15$, $\delta_p = 0.010$.

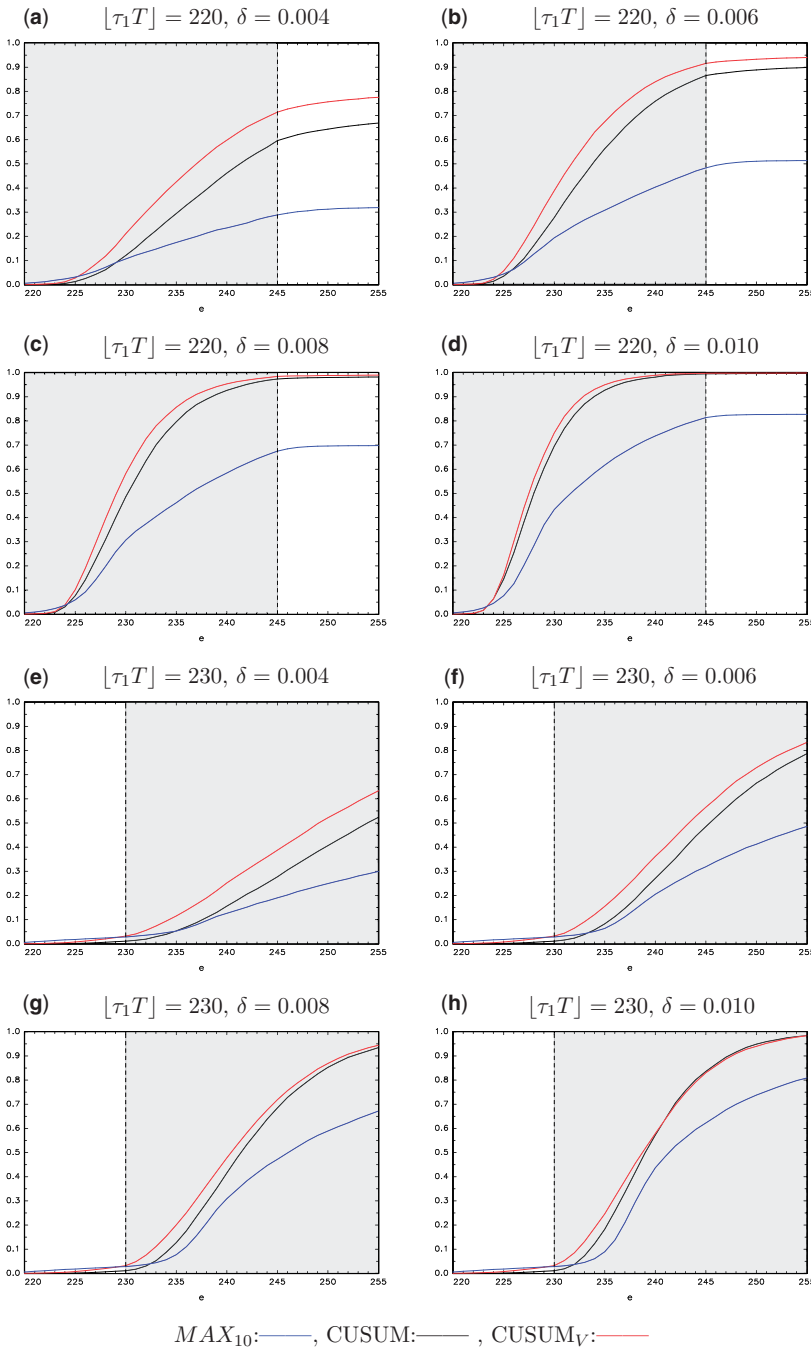

Figure 8. TPR—past bubble. (a) $\lfloor \tau_1 T \rfloor = 220$, $\delta = 0.004$. (b) $\lfloor \tau_1 T \rfloor = 220$, $\delta = 0.006$. (c) $\lfloor \tau_1 T \rfloor = 220$, $\delta = 0.008$. (d) $\lfloor \tau_1 T \rfloor = 220$, $\delta = 0.010$. (e) $\lfloor \tau_1 T \rfloor = 230$, $\delta = 0.004$. (f) $\lfloor \tau_1 T \rfloor = 230$, $\delta = 0.006$. (g) $\lfloor \tau_1 T \rfloor = 230$, $\delta = 0.008$. (h) $\lfloor \tau_1 T \rfloor = 230$, $\delta = 0.010$.

TPR of the CUSUM_V procedure is unchanged relative to the case where no training period episode is present, for the same reasons discussed above in the context of the robustness of its empirical FPR to training period explosive episodes. Contrastingly, the TPRs of both the CUSUM and MAX_{10} procedures are significantly negatively impacted by the presence of a training period explosive episode, so that the CUSUM_V procedure easily offers the best TPR profile of the three procedures considered in such cases.

3.4 Additional Simulations

In addition to the simulation results discussed in Sections 3.1–3.3 we also performed a large number of further experiments which we do not report in detail here, but which are available on request. We now summarize these findings:

i. The simulations reported in Sections 3.1–3.3 all set ε_t to be an NIID(0, 1) process. We repeated the experiments for a number of other distributions for ε_t, including highly skewed distributions such as $(\chi^2_1 - 1)$, and found these all to yield very similar results to those reported.

ii. We also examined the impact of other forms of time-varying volatility on the finite sample properties of the CUSUM, CUSUM_V, and MAX_{10} procedures. In particular, we considered linearly trending volatility paths, abrupt shifts in volatility, and multiple smooth or abrupt volatility shifts. Upward (downward) trending volatility paths lead to FPR inflation (deflation) for the CUSUM procedure but had little impact on the FPR of the CUSUM_V procedure. Abrupt upward (downward) shifts in volatility again lead to significant FPR inflation (deflation) for the CUSUM procedure but only had a modest impact on the FPR of the CUSUM_V procedure, in spite of such abrupt volatility shifts not being permitted by Assumption 2. With multiple smooth or abrupt volatility shifts, the FPR of the CUSUM monitoring procedures was governed by the relative average volatility in the training and monitoring periods implied by each volatility path. For scenarios where the average volatility in the monitoring period exceeded that in the training period the CUSUM procedures exhibited FPR inflation, whereas if the average volatility in the monitoring period was less than that in the training period the CUSUM procedures exhibited FPR deflation. In all instances, the inflation or deflation exhibited by the CUSUM procedure was far more severe than for the CUSUM_V procedure. The FPR of the MAX_{10} was unaffected by trending volatility paths, abrupt shifts in volatility, or multiple volatility shifts.

iii. When allowing for trending volatility under the alternative, the TPRs of the CUSUM_V and MAX_{10} procedures were little different to those obtained in the homoskedastic baseline. In line with the distortions in the FPR outlined above, under upward (downward) trending volatility the TPR of the CUSUM procedure was relatively higher (lower) than the other two procedures. In the case of abrupt upward variance shifts the power ordering of the three test procedures was unchanged relative to the homoskedastic baseline, albeit with the difference in power between the test procedures being amplified somewhat due to the modest FPR inflation displayed by the CUSUM_V procedure and the severe FPR inflation exhibited by the CUSUM procedure. For abrupt downward variance shifts under the alternative, the TPR of the MAX_{10} procedure was largely unchanged, whereas the powers of the CUSUM and CUSUM_V procedures were

deflated relative to the homoskedastic baseline, modestly so in the case of the CUSUM$_V$ procedure and severely so in the case of the CUSUM procedure.

4 Empirical Application

In this section, we illustrate the methods discussed in this paper using empirical data on Bitcoin. Bitcoin is a digital cryptocurrency that, much like government-backed currencies, is envisaged as a medium of exchange. The price of Bitcoin has, however, been subject to a great deal of volatility since its inception and, as such, is regarded as a speculative asset. We apply our procedures directly to the Bitcoin price with no adjustment for fundamentals since there is no consensus as to what would be the appropriate fundamental for the price of Bitcoin or if indeed it could even be measured (e.g., the cost of mining has been suggested). Consequently, what we are doing is examining the Bitcoin data for an emerging explosive episode. While formally this doesn't allow us to determine if a bubble is present, it does nonetheless provide some evidence that a bubble might be present.

We obtained daily data on the price of Bitcoin from https://finance.yahoo.com/quote/BTC-GBP. We will concentrate attention on monitoring for explosive episodes in the year 2017 as the price of Bitcoin rose markedly over the course of this year and also appears subject to time-varying volatility.

We apply both the CUSUM and CUSUM$_V$ monitoring procedures to this dataset. Results for the MAX_{10} procedure are omitted as this procedure fails to find evidence of explosivity in either of the two example monitoring exercises we consider. To account for potential serial correlation in the data, we apply pre-whitening to the first differences of Δy_t as discussed in Remark 11, selecting the lag order, p, using BIC with a maximum lag order of 4.

Before considering the results of our monitoring exercises we first plot in Figure 9 the price and estimated volatility path of the Bitcoin series for the period January 1, 2017–November 30, 2017. The estimated volatility path is computed using the kernel smoothing estimator in Equation (7) using the same choices for the kernel function and H as in the Monte Carlo exercise reported in Section 3. It can be seen from Figure 9 that the Bitcoin price series rose a great deal over 2017, beginning the year at £809 and rising to £7,565 by November 30, 2017, leading to widespread belief that the series may have been subject to one or more explosive episodes over the period in question. Figure 9 also highlights the presence of considerable time-variation in the estimated volatility path of the Bitcoin price series. It is therefore seen to be of considerable importance to allow for the presence of time-varying volatility in the data when investigating whether or not the general upward movement in the Bitcoin price series is due to explosive episodes.

We report results for two monitoring exercises performed on the Bitcoin data. In each case, the length of the training period is set to be the same value as in the Monte Carlo simulations in Section 3. We first consider how a real-time monitoring exercise that began on July 15, 2017 and ended on August 19, 2017 would have played out, with the training period for this monitoring exercise given by data from December 8, 2016 to July 14, 2017. This monitoring period is identified by the first shaded area in Figure 9, with the solid and dashed black lines plotted within this period identifying the first point at which the CUSUM$_V$ and CUSUM monitoring procedures signal a rejection, respectively. The price of Bitcoin increases slightly from July 15–July 20 and is then relatively flat until the end of

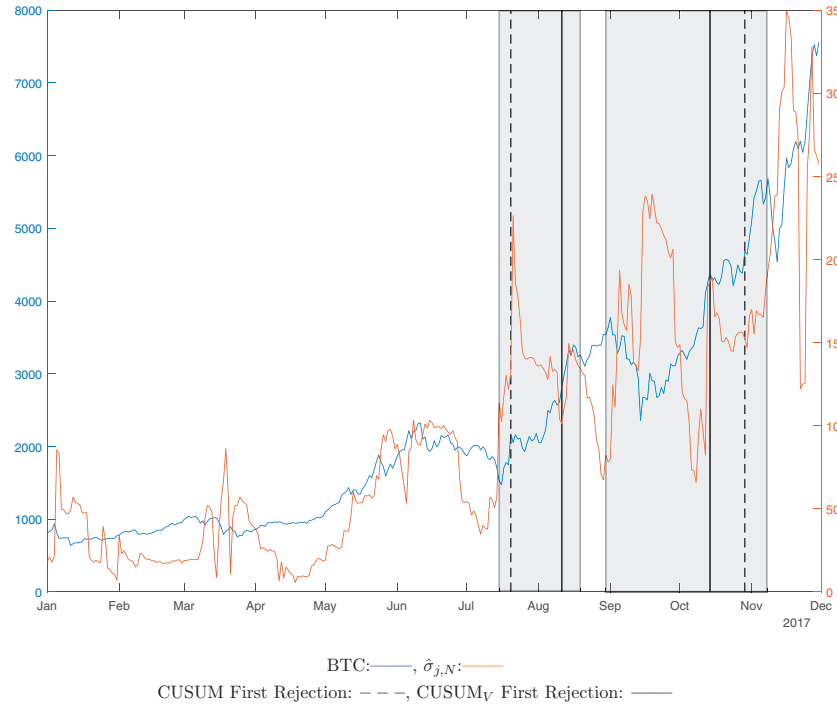

Figure 9. Application to daily Bitcoin prices.

July, at which point it begins to increase again until the end of the monitoring period. The estimated volatility path of the series shows a marked increase at the start of the monitoring period, before gradually declining over the remainder of the monitoring period. In this example, both the CUSUM and CUSUM$_V$ monitoring procedures signal an explosive episode, with the CUSUM procedure first rejecting on July 20 and CUSUM$_V$ first rejecting on August 11. In Section 3.1, we saw that upward volatility shifts can significantly increase the FPR of the standard CUSUM procedure relative to CUSUM$_V$. It is also worth noting that the spike in the volatility estimate on July 21 implies that the volatility of the series increased a great deal on July 20 which is precisely the date where the CUSUM procedure first rejects.[3] It therefore seems likely that in this example the CUSUM procedure signals an explosive episode much earlier than the CUSUM$_V$ procedure because the former is running off a much higher FPR than the latter due to an increase in volatility at the start of the monitoring period. That it signals an explosive episode at exactly the point where a large spike in volatility is observed while CUSUM$_V$ does not is further suggestive that this may be a spurious detection.

For our second illustration, we investigate how a real-time monitoring exercise that began on August 30, 2017 and ended on November 8, 2017 would have played out, with the training period for this monitoring exercise given by data from January 23, 2017 to

3 The kernel smoothing estimator in Equation (7) imposes leave-one-out so that any increases in volatility have a 1-period lag in entering the volatility estimator.

August 29, 2017. This monitoring period is identified by the second shaded area in Figure 9 with, again, the solid and dashed black lines plotted within this period identifying the first point at which the CUSUM_V and CUSUM monitoring procedures signal a rejection, respectively. The price of Bitcoin declines slightly at the start of the monitoring period, but then increases from September 14 until the end of the monitoring period. The estimated volatility path of the series first increases up until September 15 before then dropping rapidly up until October 9. At this point, volatility then increases again up until October 13 before stabilizing for the remainder of the monitoring period. In contrast to our first monitoring exercise, the CUSUM_V procedure is the first to signal an explosive episode on October 14, with the CUSUM procedure not rejecting until October 29. So, while in our first example an upward movement in volatility coincided with the CUSUM procedure rejecting first, here we observe the opposite outcome. In this scenario, the decline in volatility seen from mid-September until early October seems likely to have effected a lower rejection probability for the CUSUM procedure relative to CUSUM_V, consistent with the simulation results in Section 3.1 for the case of downward shifts in volatility, allowing the latter to deliver a far earlier rejection of the null. We also note that the results of the first monitoring exercise indicate that there may be an explosive episode present in the training period for this second monitoring exercise, which may also be contributing to the significant delay in CUSUM rejecting relative to CUSUM_V and would again be consistent with the simulation results reported in Figures 7 and 8.

5 Conclusions

We have generalized the CUSUM-based real-time explosive episode detection procedure of Homm and Breitung (2012) to allow for the presence of time-varying volatility in the innovations. Such patterns were shown to cause potentially severe inflation in the true FPR of the CUSUM procedure. Our proposed modification involves replacing the first-difference estimator of the variance used in the CUSUM statistics by a Nadaraya–Watson-type nonparametric estimator. The resulting sequence of modified CUSUM statistics was shown to have a pivotal joint limiting null distribution coinciding with that of the sequence of standard CUSUM statistics under homoskedasticity with the result that the theoretical FPR of the procedure is controlled. A discussion of the bandwidth and kernel choices associated with the nonparametric variance estimator was also provided with a cross validation choice recommended for the former, whereby it is selected to minimize the estimation error of the spot variance over the most recent observations. Simulation evidence, for a variety of time-varying volatility processes, suggested that the FPR of the modified procedure is well controlled in finite samples. Where the innovations were homoskedastic the potency of the modified procedure to detect an emergent explosive episode was shown to be only slightly lower than the standard procedure. In contrast to both the standard CUSUM procedure and the procedure of Astill et al. (2018), the modified CUSUM procedure was also shown to be robust to the presence of explosive episodes in the training period. An application to a Bitcoin price data series was used to illustrate the possible advantages of our proposed procedure relative to the standard CUSUM procedure, with our proposed procedure signaling the presence of an explosive episode sooner than the standard CUSUM procedure in a period of downward transitioning volatility and avoiding a potential early false rejection in a period of upward transitioning volatility.

APPENDIX A: MATHEMATICAL PROOFS

Throughout the proofs, unless otherwise stated, we use $\max_j$ as shorthand notation for $\max_{T+1 \leqslant j \leqslant \lfloor \lambda T \rfloor}$. We also denote $(\sigma^2)'(.)$ as the derivative of $\sigma^2(.)$.

A.1 Preparatory Lemmas

Lemma A1. Let the conditions of Theorem 1 hold. Then, under H_0,

$$\max_j \left| \sum_{s=0}^{N} w_s \sigma^2_{j-s} (\varepsilon^2_{j-s} - 1) \right| = o_p(1).$$

Proof of Lemma A1. This lemma is proved using the Fourier transformation-based method as in Theorem 2.8 of Pagan and Ullah (1999). Notice that

$$\begin{aligned} \sum_{s=0}^{N} w_s \sigma^2_{j-s} (\varepsilon^2_{j-s} - 1) &= \frac{\sum_{s=0}^{N} K(s/N) \sigma^2_{j-s} (\varepsilon^2_{j-s} - 1)}{\sum_{s=0}^{N} K(s/N)} \\ &= \frac{\sum_{l=1}^{\lfloor \lambda T \rfloor} K((j-l)/N) \sigma^2_l (\varepsilon^2_l - 1)}{\sum_{l=1}^{\lfloor \lambda T \rfloor} K((j-l)/N)}, \end{aligned} \tag{A.1}$$

where we have used the change of variable $l = j - s$. In fact we know that there are zero terms in the sum as K is only non-zero on $(0, 1)$, but we keep the sum free of the index j as a mechanism to derive the max rate.

Consider first the numerator of (A.1). For this, we have that

$$\begin{aligned} \sum_{l=1}^{\lfloor \lambda T \rfloor} K((j-l)/N) \sigma^2_l (\varepsilon^2_l - 1) &= \sum_{l=1}^{\lfloor \lambda T \rfloor} \frac{1}{2\pi} \int \exp\left(-\mathrm{i}s \left(\frac{j-l}{N} \right) \right) \sigma^2_l (\varepsilon^2_l - 1) \phi(s) \mathrm{d}s \\ &= \frac{1}{2\pi} \int \left(\sum_{l=1}^{\lfloor \lambda T \rfloor} \exp\left(\mathrm{i}s \frac{l}{N} \right) \sigma^2_l (\varepsilon^2_l - 1) \right) \exp\left(-\mathrm{i}s \frac{j}{N} \right) \phi(s) \mathrm{d}s \\ &= \frac{N}{2\pi} \int \left(\sum_{l=1}^{\lfloor \lambda T \rfloor} \exp(\mathrm{i}tl) \sigma^2_l (\varepsilon^2_l - 1) \right) \exp(-\mathrm{i}j) \phi(tN) \mathrm{d}t, \end{aligned}$$

where we have used the change of variable $s = tN$. Thus, we have

$$\begin{aligned}
&\max_j |\sum_{l=1}^{\lfloor \lambda T \rfloor} K((j-l)/N)\sigma_l^2(\varepsilon_l^2-1)| \\
\leqslant\ & \max_j |\frac{N}{2\pi}\int \Big(\sum_{l=1}^{\lfloor \lambda T \rfloor} \exp(\mathrm{i}tl)\sigma_l^2(\varepsilon_l^2-1)\Big)\exp(-\mathrm{i}j)\phi(tN)\mathrm{d}t| \\
\leqslant\ & \frac{N}{2\pi}\int |\sum_{l=1}^{\lfloor \lambda T \rfloor} \exp(\mathrm{i}tl)\sigma_l^2(\varepsilon_l^2-1)| \max_j |\exp(-\mathrm{i}j)||\phi(tN)|\mathrm{d}t \\
\leqslant\ & \frac{N}{2\pi}\int |\sum_{l=1}^{\lfloor \lambda T \rfloor} \exp(\mathrm{i}tl)\sigma_l^2(\varepsilon_l^2-1)||\phi(tN)|\mathrm{d}t,
\end{aligned}$$

where we have used the fact that $\max_j|\exp(-\mathrm{i}j)| = 1$ identically, by definition. Thus, the right-hand side now is free of the index j. Taking expectations on both sides, we then have that

$$E\max_j\left|\sum_{l=1}^{\lfloor \lambda T \rfloor} K((j-l)/N)\sigma_l^2(\varepsilon_l^2-1)\right| \leqslant \frac{N}{2\pi}\int E\left[\left|\sum_{l=1}^{\lfloor \lambda T \rfloor} \exp(\mathrm{i}tl)\sigma_l^2(\varepsilon_l^2-1)\right|\right]|\phi(tN)|\mathrm{d}t. \quad (A.2)$$

Next, we evaluate the expectation in the integral in the right member of Equation (A.2)

$$\begin{aligned}
&E|\sum_{l=1}^{\lfloor \lambda T \rfloor} \exp(\mathrm{i}tl)\sigma_l^2(\varepsilon_l^2-1)| \\
=\ & E|\sum_{l=1}^{\lfloor \lambda T \rfloor} (\cos(tl)\sigma_l^2(\varepsilon_l^2-1) + \mathrm{i}\sin(tl)\sigma_l^2(\varepsilon_l^2-1))| \\
=\ & E\Big(\Big(\sum_{l=1}^{\lfloor \lambda T \rfloor} \cos(tl)\sigma_l^2(\varepsilon_l^2-1)\Big)^2 + \Big(\sum_{l=1}^{\lfloor \lambda T \rfloor} \sin(tl)\sigma_l^2(\varepsilon_l^2-1)\Big)^2\Big)^{1/2} \\
\leqslant\ & \Big(E\Big(\sum_{l=1}^{\lfloor \lambda T \rfloor} \cos(tl)\sigma_l^2(\varepsilon_l^2-1)\Big)^2 + E\Big(\sum_{l=1}^{\lfloor \lambda T \rfloor} \sin(tl)\sigma_l^2(\varepsilon_l^2-1)\Big)^2\Big)^{1/2},
\end{aligned}$$

using Jensen's inequality which gives us that $EZ^{1/2} \leqslant (EZ)^{1/2}$ for $Z > 0$. As $\{\varepsilon_l^2 - 1\}$ is a martingale difference sequence indexed by l, it follows using Burkholder's inequality (e.g., Shiryaev, 1996, p. 499) that, for a positive constant C,

$$E\Big(\sum_{l=1}^{\lfloor \lambda T \rfloor} \cos(tl)\sigma_l^2(\varepsilon_l^2-1)\Big)^2 \leqslant CE\sum_{l=1}^{\lfloor \lambda T \rfloor} \cos^2(tl)\sigma_l^4(\varepsilon_l^2-1)^2 = O(T),$$

by the uniform boundedness of volatility function and the existence of fourth moment of ε_l. Similarly, we have

$$E\Big(\sum_{l=1}^{\lfloor \lambda T \rfloor} \sin(tl)\sigma_l^2(\varepsilon_l^2 - 1)\Big)^2 = O(T).$$

Consequently,

$$E\left[\left|\sum_{l=1}^{\lfloor \lambda T \rfloor} \exp(\mathrm{i}tl)\sigma_l^2(\varepsilon_l^2 - 1)\right|\right] = O(\sqrt{T}).$$

Plugging this result into Equation (A.2), we therefore have that

$$\begin{aligned} E\max_j\Big|\sum_{l=1}^{\lfloor \lambda T \rfloor} K((j-l)/N)\sigma_l^2(\varepsilon_l^2 - 1)\Big| &\leqslant \frac{N}{2\pi} O(\sqrt{T}) \int |\phi(tN)|\mathrm{d}t \\ &= \frac{1}{2\pi} O(\sqrt{T}) \int |\phi(s)|\mathrm{d}s \\ &= O(\sqrt{T}). \end{aligned}$$

Then by Markov's inequality, it follows straightforwardly that

$$\max_j \Big|\sum_{l=1}^{\lfloor \lambda T \rfloor} K((j-l)/N)\sigma_l^2(\varepsilon_l^2 - 1)\Big| = O_p(\sqrt{T}).$$

Turning to the denominator of Equation (A.1), $\sum_{l=1}^{\lfloor \lambda T \rfloor} K((j-l)/N)$ clearly has order $O(N)$, uniformly for all j. Combining the results for the numerator and denominator of Equation (A.1), we therefore have that

$$\max_j \Big|\sum_{s=0}^{N} w_s \sigma_{j-s}^2(\varepsilon_{j-s}^2 - 1)\Big| = \frac{O_p(\sqrt{T})}{O(N)} = o_p(1),$$

by the bandwidth assumption that $N^2/T \to \infty$. □

Lemma A2. Under H_1, when $\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor$, with $\phi = 1 + \delta$,

$$\max_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor} |NT^{-1}\phi^{-2(j-1-\lfloor \tau_1 T \rfloor)}\hat{\sigma}_{j,N}^2| = O_p(1).$$

We also have $\min_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor} |NT^{-1}\phi^{-2(j-1-\lfloor \tau_1 T \rfloor)}\hat{\sigma}_{j,N}^2|$ is nondegenerate to 0.

Proof of Lemma A2. First, by repeated backward substitution, for $\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor$,

$$u_j = e_j + \phi e_{j-1} + \cdots + \phi^{j-\lfloor \tau_1 T \rfloor - 1} e_{\lfloor \tau_1 T \rfloor + 1} + \phi^{j-\lfloor \tau_1 T \rfloor} u_{\lfloor \tau_1 T \rfloor} \tag{A.3}$$

and $u_{\lfloor \tau_1 T \rfloor}$ is the last observation in the unit root regime (and also serves as the initial value for the explosive regime). Since $\{e_j\}$ is a m.d.s., using Burkholder's inequality, we have

$$E\phi^{-2(j-\lfloor \tau_1 T \rfloor)}|e_j + \phi e_{j-1} + \cdots + \phi^{j-\lfloor \tau_1 T \rfloor - 1} e_{\lfloor \tau_1 T \rfloor + 1}|^2 = O(1),$$

and it follows that $\phi^{-(j-\lfloor \tau_1 T \rfloor)}(e_j + \phi e_{j-1} + \cdots + \phi^{j-\lfloor \tau_1 T \rfloor - 1} e_{\lfloor \tau_1 T \rfloor + 1}) = O_p(1)$ for any $\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor$. Then by Doob's maximal inequality for martingales, we also have

$$\max_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor} |\phi^{-(j-\lfloor \tau_1 T \rfloor)}(e_j + \phi e_{j-1} + \cdots + \phi^{j-\lfloor \tau_1 T \rfloor - 1} e_{\lfloor \tau_1 T \rfloor + 1})| = O_p(1).$$

For the initial value of the explosive regime, it is satisfied that

$$\max_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor} |\phi^{-(j-\lfloor \tau_1 T \rfloor)} \phi^{j-\lfloor \tau_1 T \rfloor} u_{\lfloor \tau_1 T \rfloor}| = |u_{\lfloor \tau_1 T \rfloor}| = O_p(\sqrt{T}).$$

So the effect of the initial point is dominant in Equation (A.3), and we have

$$\max_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor} |T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} u_{j-1}| = O_p(1). \tag{A.4}$$

We first prove the result for $\lfloor \tau_1 T \rfloor + N + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor$. When $\lfloor \tau_1 T \rfloor + N + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor$,

$$\begin{aligned}
\hat{\sigma}^2_{j,N} &= \sum_{s=0}^{N} w_s \Delta y^2_{j-s} \\
&= \sum_{s=0}^{N} w_s (\delta u_{j-s-1} + e_{j-s})^2 \\
&= \delta^2 \sum_{s=0}^{N} w_s u^2_{j-s-1} + \sum_{s=0}^{N} w_s e^2_{j-s} + 2\delta \sum_{s=0}^{N} w_s u_{j-s-1} e_{j-s}. \\
&= D_1 + D_2 + D_3.
\end{aligned}$$

Notice that D_1 satisfies

$$\begin{aligned}
&\max_{\lfloor \tau_1 T \rfloor + N + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor} NT^{-1} \phi^{-2(j-1-\lfloor \tau_1 T \rfloor)} D_1 \\
&= \max_{\lfloor \tau_1 T \rfloor + N + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor} \delta^2 N \phi^{-2(j-1-\lfloor \tau_1 T \rfloor)} \sum_{s=0}^{N} w_s (T^{-1} \phi^{-2(j-s-1-\lfloor \tau_1 T \rfloor)} u^2_{j-s-1}) \phi^{2(j-s-1-\lfloor \tau_1 T \rfloor)} \\
&\leqslant \delta^2 \max_{\lfloor \tau_1 T \rfloor + N + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor, 0 \leqslant s \leqslant N} |T^{-1} \phi^{-2(j-s-1-\lfloor \tau_1 T \rfloor)} u^2_{j-s-1}| (N \sum_{s=0}^{N} w_s \phi^{-2s}) \\
&= O_p(1),
\end{aligned}$$

where we have used Equation (A.4) and that $N\sum_{s=0}^{N} w_s\phi^{-2s} \leqslant N\max_0 \leqslant s \leqslant N|w_s|\sum_{s=0}^{N}\phi^{-2s} = O(1)$. D_2 is clearly dominated by D_1 and D_3 is the cross product term which cannot be the largest term of the three. Thus, D_1 is dominant and we have

$$\max_{\lfloor\tau_1 T\rfloor+N+1 \leqslant j \leqslant \lfloor\tau_2 T\rfloor} |NT^{-1}\phi^{-2(j-1-\lfloor\tau_1 T\rfloor)}\hat{\sigma}^2_{j,N}| = O_p(1).$$

Notice that D_1 also satisfies

$$\min_{\lfloor\tau_1 T\rfloor+N+1 \leqslant j \leqslant \lfloor\tau_2 T\rfloor} NT^{-1}\phi^{-2(j-1-\lfloor\tau_1 T\rfloor)}D_1$$

$$\geqslant \delta^2 \times \min_{\lfloor\tau_1 T\rfloor+N+1 \leqslant j \leqslant \lfloor\tau_2 T\rfloor, 0 \leqslant s \leqslant N} |T^{-1}\phi^{-2(j-s-1-\lfloor\tau_1 T\rfloor)}u^2_{j-s-1}|)(N\sum_{s=0}^{N} w_s\phi^{-2s}),$$

which is also $O_p(1)$ and clearly nondegenerate to $o_p(1)$, so the claimed result for $\lfloor\tau_1 T\rfloor + N + 1 \leqslant j \leqslant \lfloor\tau_2 T\rfloor$ is proved.

Now for $\lfloor\tau_1 T\rfloor + 1 \leqslant j \leqslant \lfloor\tau_1 T\rfloor + N$, we have

$$\begin{aligned}\hat{\sigma}^2_{j,N} &= \sum_{s=0}^{j-\lfloor\tau_1 T\rfloor-1} w_s(\delta u_{j-s-1} + e_{j-s})^2 + \sum_{s=j-\lfloor\tau_1 T\rfloor}^{N} w_s e^2_{j-s} \\ &= \delta^2 \sum_{s=0}^{j-\lfloor\tau_1 T\rfloor-1} w_s u^2_{j-s-1} + \sum_{s=0}^{N} w_s e^2_{j-s} + 2\delta \sum_{s=0}^{j-\lfloor\tau_1 T\rfloor-1} w_s u_{j-s-1}e_{j-s} \\ &\equiv D'_1 + D'_2 + D'_3.\end{aligned}$$

As with D_1, D'_1 is a weighted average of preceding u^2_j but with less than N terms. D'_2 is the same as D_2 while D'_3 is the cross-product term with less terms to sum than D_3. Using the same argument for analyzing D_1, D_2, and D_3, it can still be shown that D'_1 is the dominant term, and we still have the claimed lower bound and upper bound results for $\lfloor\tau_1 T\rfloor + 1 \leqslant j \leqslant \lfloor\tau_1 T\rfloor + N$ proved. Intuitively, this is because although in this regime there are less terms in D'_1 than D_1, the order of D'_1 is still determined by the last observation u_{j-1}. The previous derived lower bound for D_1 also holds for D'_1. Taken together, the lemma is proved. □

Lemma A3. Under H_1, when $\lfloor\tau_1 T\rfloor + 1 \leqslant t \leqslant \lfloor\tau_2 T\rfloor$, with $\phi = 1 + \delta$,

$$\sum_{j=\lfloor\tau_1 T\rfloor+1}^{t} |T^{-1/2}\phi^{-(j-1-\lfloor\tau_1 T\rfloor)}u_{j-1}| \sim O_p(t - \lfloor\tau_1 T\rfloor - 1).$$

Proof of Lemma A3. Using Equation (A.4), we have

$$\sum_{j=\lfloor\tau_1 T\rfloor+1}^{t}|T^{-1/2}\phi^{-(j-1-\lfloor\tau_1 T\rfloor)}u_{j-1}| \leqslant \max_{\lfloor\tau_1 T\rfloor+1 \leqslant j \leqslant t}|T^{-1/2}\phi^{-(j-1-\lfloor\tau_1 T\rfloor)}u_{j-1}|\sum_{j=\lfloor\tau_1 T\rfloor+1}^{t}1$$
$$= O_p(t-\lfloor\tau_1 T\rfloor-1).$$

For the lower bound part of the proof, notice that

$$\sum_{j=\lfloor\tau_1 T\rfloor+1}^{t}|T^{-1/2}\phi^{-(j-1-\lfloor\tau_1 T\rfloor)}u_{j-1}| \geqslant \min_{\lfloor\tau_1 T\rfloor+1 \leqslant j \leqslant t}|T^{-1/2}\phi^{-(j-1-\lfloor\tau_1 T\rfloor)}u_{j-1}|\sum_{j=\lfloor\tau_1 T\rfloor+1}^{t}1.$$

From Equation (A.4), it is known that $|T^{-1/2}\phi^{-(j-1-\lfloor\tau_1 T\rfloor)}u_{j-1}| = O_p(1)$ for any $\lfloor\tau_1 T\rfloor + N + 1 \leqslant j \leqslant t$ and is nondegenerate to 0. Therefore, $O_p(t-\lfloor\tau_1 T\rfloor-1)$ is also a lower bound rate and the proof of the lemma is finished.

A.2 Proof of Lemma 1

First consider the decomposition

$$\begin{aligned}\hat{\sigma}_j^2-\sigma_j^2 &= \sum_{s=0}^{N}w_s e_{j-s}^2-\sigma_j^2\\ &= \sum_{s=0}^{N}w_s\sigma_{j-s}^2\varepsilon_{j-s}^2-\sigma_j^2\\ &= \sum_{s=0}^{N}w_s\sigma_{j-s}^2(\varepsilon_{j-s}^2-1)+\sum_{s=0}^{N}w_s\sigma_{j-s}^2-\sigma_j^2\\ &=: A_{1,j}+A_{2,j},\end{aligned} \tag{A.5}$$

where $A_{1,j}$ and $A_{2,j}$ are defined implicitly. By Lemma A1, we have $\max_j|A_{1,j}| = o_p(1)$. Next consider $A_{2,j}$,

$$\begin{aligned}\max_j|\sum_{s=0}^{N}w_s\sigma_{j-s}^2-\sigma_j^2| &= \max_j|\frac{\sum_{s=0}^{N}K\left(\frac{s}{N}\right)\sigma_{j-s}^2}{\sum_{s=0}^{N}K\left(\frac{s}{N}\right)}-\sigma_j^2|\\ &= \max_j|\frac{\sum_{s=0}^{N}K\left(\frac{s}{N}\right)(\sigma_{j-s}^2-\sigma_j^2)}{\sum_{s=0}^{N}K\left(\frac{s}{N}\right)}|\\ &= \max_j|\frac{\int_0^1 K(u)\left(\sigma^2\left(\frac{j-uN}{T}\right)-\sigma^2\left(\frac{j}{T}\right)\right)\mathrm{d}u+o(1)}{\int_0^1 K(u)\mathrm{d}u+o(1)}|,\end{aligned}$$

where we have used the approximation given by the convergence to the Riemann integral. The approximation error is clearly dependent on N and independent of j. Using the continuous differentiability of the $\sigma(.)$ function, we have that

$$\max_j \left| \frac{\int_0^1 K(u)\left(\sigma^2\left(\frac{j-uN}{T}\right) - \sigma^2\left(\frac{j}{T}\right)\right)\mathrm{d}u + o(1)}{\int_0^1 K(u)\mathrm{d}u + o(1)} \right| \leqslant C \max_j \frac{\int_0^1 |K(u)u|\mathrm{d}u}{|\int_0^1 K(u)\mathrm{d}u|}\left(\frac{N}{T}\right) \to 0,$$

by our assumption that $N/T \to 0$.

Taken together these results establish that $\max_j|\hat{\sigma}_j^2 - \sigma_j^2| = o_p(1)$.

A.3 Proof of Lemma 2

Using the decomposition in Equation (A.5), we have

$$\hat{\sigma}_j^2 - \sigma_j^2 = A_{1,j} + A_{2,j}.$$

The object of interest can therefore be written as

$$|(\hat{\sigma}_{j-1}^2 - \sigma_{j-1}^2) - (\hat{\sigma}_j^2 - \sigma_j^2)| = |(A_{1,j-1} - A_{1,j}) + (A_{2,j-1} - A_{2,j})|.$$

Consider first the difference $A_{1,j} - A_{1,j-1}$,

$$\begin{aligned}
&A_{1,j} - A_{1,j-1} \\
&= \textstyle\sum_{s=0}^N w_s\sigma_{j-s}^2(\varepsilon_{j-s}^2 - 1) - \sum_{s=0}^N w_s\sigma_{j-1-s}^2(\varepsilon_{j-1-s}^2 - 1) \\
&= \frac{\sum_{s=0}^N K\left(\frac{s}{N}\right)\sigma_{j-s}^2(\varepsilon_{j-s}^2 - 1) - \sum_{s=0}^N K\left(\frac{s}{N}\right)\sigma_{j-1-s}^2(\varepsilon_{j-1-s}^2 - 1)}{\sum_{s=0}^N K\left(\frac{s}{N}\right)} \\
&= \frac{\sum_{s=0}^N K\left(\frac{s}{N}\right)\sigma_{j-s}^2(\varepsilon_{j-s}^2 - 1) - \sum_{s=1}^{N+1} K\left(\frac{s-1}{N}\right)\sigma_{j-s}^2(\varepsilon_{j-s}^2 - 1)}{\sum_{s=0}^N K\left(\frac{s}{N}\right)} \\
&= \frac{\sum_{s=1}^N \left(K\left(\frac{s}{N}\right) - K\left(\frac{s-1}{N}\right)\right)\sigma_{j-s}^2(\varepsilon_{j-s}^2 - 1) + K(0)\sigma_j^2(\varepsilon_j^2 - 1) - K(1)\sigma_{j-1-N}^2(\varepsilon_{j-1-N}^2 - 1)}{\sum_{s=0}^N K\left(\frac{s}{N}\right)} \\
&= \frac{\sum_{s=1}^N \left(K\left(\frac{s}{N}\right) - K\left(\frac{s-1}{N}\right)\right)\sigma_{j-s}^2(\varepsilon_{j-s}^2 - 1)}{\sum_{s=0}^N K\left(\frac{s}{N}\right)},
\end{aligned}$$

where we have used the fact that $K(0) = K(1) = 0$. Because K is continuously differentiable over (0, 1), we can employ the mean value theorem to show that the foregoing expression becomes

$$\frac{1}{N}\frac{\sum_{s=1}^{N} K'(\tau_s)\sigma^2_{j-s}(\varepsilon^2_{j-s}-1)}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)},$$

where $\tau_s \in ((s-1)/N, s/N)$. Using the same strategy as used in Lemma A1, coupled with the absolute integrability assumption placed on the characteristic function of $K'(.)$ in Assumption 1, we can obtain that $\max_j|\sum_{s=1}^{N} K'(\tau_s)\sigma^2_{j-s}(\varepsilon^2_{j-s}-1)| = O_p(\sqrt{N})$, and thus that

$$\max_j |A_{1,j} - A_{1,j-1}| = \max_j \left| \frac{\frac{1}{N}\sum_{s=1}^{N} K'(\tau_s)\sigma^2_{j-s}(\varepsilon^2_{j-s}-1)}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)} \right| = O_p\left(\frac{\sqrt{N}}{N^2}\right) = O_p(N^{-3/2}),$$

which is clearly $o_p(1/T)$ under our bandwidth assumption $N^{3/2}/T \to \infty$.

Next notice that

$$\begin{aligned} A_{2,j} &= \textstyle\sum_{s=0}^{N} w_s \sigma^2_{j-s} - \sigma^2_j \\ &= \frac{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)\sigma^2_{j-s}}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)} - \sigma^2_j \\ &= \frac{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)(\sigma^2_{j-s} - \sigma^2_j)}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}, \end{aligned}$$

so that the difference $A_{2,j} - A_{2,j-1}$ can be written as

$$\begin{aligned} A_{2,j} - A_{2,j-1} &= \frac{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)(\sigma^2_{j-s} - \sigma^2_j) - \sum_{s=0}^{N} K\left(\frac{s}{N}\right)(\sigma^2_{j-1-s} - \sigma^2_{j-1})}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)} \\ &= \frac{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)((\sigma^2_{j-s} - \sigma^2_{j-1-s}) - (\sigma^2_j - \sigma^2_{j-1}))}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)} \\ &= \frac{\frac{1}{T}\sum_{s=0}^{N} K\left(\frac{s}{N}\right)((\sigma^2)'(\tau_{j-1-s}) - (\sigma^2)'(\tau_{j-1}))}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}, \end{aligned}$$

where $\tau_{j-1-s} \in ((j-1-s)/T, (j-s)/T), \tau_{j-1} \in ((j-1)/T, j/T)$, where we have used the mean value theorem based on the differentiability of the $\sigma^2(.)$ function. By the Lipschitz assumption made on the $(\sigma^2)'(.)$ function, we have that

$$\max_j |A_{2,j} - A_{2,j-1}| \leqslant \frac{\frac{1}{T}\sum_{s=0}^{N} K\left(\frac{s}{N}\right) \max_j |(\sigma^2)'(\tau_{j-1-s}) - (\sigma^2)'(\tau_{j-1})|}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}$$

$$\leqslant \frac{C\frac{1}{T}\sum_{s=0}^{N} K\left(\frac{s}{N}\right)\frac{s}{T}}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}$$

$$= \frac{C\frac{1}{T^2}\sum_{s=0}^{N} K\left(\frac{s}{N}\right)s}{\sum_{s=0}^{N} K\left(\frac{s}{N}\right)}.$$

Using the Riemann integral approximation, we have that $\frac{1}{N}\sum_{s=0}^{N} K\left(\frac{s}{N}\right) = \int K + o(1)$, and that $\frac{1}{N^2}\sum_{s=0}^{N} K\left(\frac{s}{N}\right)s = O(1)$ because of our assumption that $\int |K(x)x|\mathrm{d}x < \infty$. We thus have that

$$\max_j |A_{2,j} - A_{2,j-1}| = O\left(\frac{1}{T^2}\frac{N^2}{N}\right) = O\left(\frac{N}{T^2}\right),$$

which is clearly $o(1/T)$ because of our assumption that $N/T \to 0$.

Finally, because

$$\max_j |(\hat{\sigma}_{j-1}^2 - \sigma_{j-1}^2) - (\hat{\sigma}_j^2 - \sigma_j^2)| \leqslant \max_j |A_{1,j} - A_{1,j-1}| + \max_j |A_{2,j} - A_{2,j-1}|,$$

the stated result then follows straightforwardly.

A.4 Proof of Theorem 1

Notice first that

$$\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\frac{e_j}{\hat{\sigma}_j} = \frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\frac{\sigma_j\varepsilon_j}{\hat{\sigma}_j}.$$

Now,

$$\begin{aligned}\frac{1}{\sqrt{T}}\sum_{j=T+1}^{\lfloor Tr\rfloor}\left(\frac{\sigma_j}{\hat{\sigma}_j}\right)\varepsilon_j &= \frac{1}{\sqrt{T}}\sum_{j=1}^{\lfloor Tr\rfloor}\left(\frac{\sigma_j}{\hat{\sigma}_j}\right)\varepsilon_j - \frac{1}{\sqrt{T}}\sum_{j=1}^{T}\left(\frac{\sigma_j}{\hat{\sigma}_j}\right)\varepsilon_j \\ &= \frac{\sigma_{\lfloor Tr\rfloor}}{\hat{\sigma}_{\lfloor Tr\rfloor}}\left(\frac{1}{\sqrt{T}}\sum_{j=1}^{\lfloor Tr\rfloor}\varepsilon_j\right) - \frac{\sigma_T}{\hat{\sigma}_T}\left(\frac{1}{\sqrt{T}}\sum_{j=1}^{T}\varepsilon_j\right) \\ &\quad + \sum_{j=T+1}^{\lfloor Tr\rfloor}\left(\frac{\sigma_j}{\hat{\sigma}_j} - \frac{\sigma_{j-1}}{\hat{\sigma}_{j-1}}\right)\left(\frac{1}{\sqrt{T}}\sum_{s=1}^{j-1}\varepsilon_s\right),\end{aligned} \tag{A.6}$$

where we have used the summation by parts formula in the second step. For the first term, observe first that

$$\hat{\sigma}_j - \sigma_j = \frac{\hat{\sigma}_j^2 - \sigma_j^2}{\hat{\sigma}_j + \sigma_j}.$$

Now since $\max_j(\hat{\sigma}_j + \sigma_j)^{-1} = O_p(1)$, using the definition of $\hat{\sigma}_j$ and Lemma 1, it is straightforwardly seen that

$$\max_j |\hat{\sigma}_j - \sigma_j| = o_p(1).$$

Next, by the FCLT for martingale difference sequences and an application of Slutsky's theorem, we have that the first term in Equation (A.6) satisfies

$$\frac{\sigma_{\lfloor Tr \rfloor}}{\hat{\sigma}_{\lfloor Tr \rfloor}} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor Tr \rfloor} \varepsilon_j \right) \Rightarrow W(r).$$

By the same argument, the second term in Equation (A.6) satisfies

$$\frac{\sigma_T}{\hat{\sigma}_T} \left(\frac{1}{\sqrt{T}} \sum_{j=1}^{T} \varepsilon_j \right) \Rightarrow W(1).$$

The stated result in Theorem 1 then follows if the third term in Equation (A.6) can be shown to be of $o_p(1)$. This we will now demonstrate. To that end, notice that

$$\begin{aligned} &\left| \sum_{j=T+1}^{\lfloor Tr \rfloor} \left(\frac{\sigma_j}{\hat{\sigma}_j} - \frac{\sigma_{j-1}}{\hat{\sigma}_{j-1}} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s \right) \right| = \left| \sum_{j=T+1}^{\lfloor Tr \rfloor} \frac{\frac{\sigma_j^2}{\hat{\sigma}_j^2} - \frac{\sigma_{j-1}^2}{\hat{\sigma}_{j-1}^2}}{\frac{\sigma_j}{\hat{\sigma}_j} + \frac{\sigma_{j-1}}{\hat{\sigma}_{j-1}}} \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s \right) \right| \\ &\leqslant \max_j \left| \frac{1}{\frac{\sigma_j}{\hat{\sigma}_j} + \frac{\sigma_{j-1}}{\hat{\sigma}_{j-1}}} \right| \left| \sum_{j=T+1}^{\lfloor Tr \rfloor} \left(\frac{\sigma_j^2}{\hat{\sigma}_j^2} - \frac{\sigma_{j-1}^2}{\hat{\sigma}_{j-1}^2} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s \right) \right|. \end{aligned}$$

In large samples, $\frac{\sigma_j}{\hat{\sigma}_j} + \frac{\sigma_{j-1}}{\hat{\sigma}_{j-1}}$ is clearly positive and bounded away from 0, and so it follows that $\max_j \left| \frac{1}{\frac{\sigma_j}{\hat{\sigma}_j} + \frac{\sigma_{j-1}}{\hat{\sigma}_{j-1}}} \right| = O_p(1)$. Let us now look at the other term involved:

$$\begin{aligned} &\left| \sum_{j=T+1}^{\lfloor Tr \rfloor} \left(\frac{\sigma_j^2}{\hat{\sigma}_j^2} - \frac{\sigma_{j-1}^2}{\hat{\sigma}_{j-1}^2} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s \right) \right| \\ &= \left| \sum_{j=T+1}^{\lfloor Tr \rfloor} \left(\frac{\sigma_j^2 \hat{\sigma}_{j-1}^2 - \sigma_{j-1}^2 \hat{\sigma}_j^2}{\hat{\sigma}_j^2 \hat{\sigma}_{j-1}^2} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s \right) \right| \\ &= \left| \sum_{j=T+1}^{\lfloor Tr \rfloor} \left(\frac{\sigma_j^2 (\hat{\sigma}_{j-1}^2 - \sigma_{j-1}^2) - \sigma_{j-1}^2 (\hat{\sigma}_j^2 - \sigma_j^2)}{\hat{\sigma}_j^2 \hat{\sigma}_{j-1}^2} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s \right) \right| \\ &= \left| \sum_{j=T+1}^{\lfloor Tr \rfloor} \left(\frac{\sigma_j^2 [(\hat{\sigma}_{j-1}^2 - \sigma_{j-1}^2) - (\hat{\sigma}_j^2 - \sigma_j^2)] + (\sigma_j^2 - \sigma_{j-1}^2)(\hat{\sigma}_j^2 - \sigma_j^2)}{\hat{\sigma}_j^2 \hat{\sigma}_{j-1}^2} \right) \left(\frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s \right) \right| \\ &\leqslant \left(\max_j \left| \frac{\sigma_j^2 [(\hat{\sigma}_{j-1}^2 - \sigma_{j-1}^2) - (\hat{\sigma}_j^2 - \sigma_j^2)]}{\hat{\sigma}_j^2 \hat{\sigma}_{j-1}^2} \right| + \max_j \left| \frac{(\sigma_j^2 - \sigma_{j-1}^2)(\hat{\sigma}_j^2 - \sigma_j^2)}{\hat{\sigma}_j^2 \hat{\sigma}_{j-1}^2} \right| \right) \sum_{j=1}^{\lfloor Tr \rfloor} \left| \frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s \right|. \end{aligned} \tag{A.7}$$

From Lemma 2, we have that

$$\max_j |(\hat{\sigma}_{j-1}^2 - \sigma_{j-1}^2) - (\hat{\sigma}_j^2 - \sigma_j^2)| = o_p\left(\frac{1}{T}\right).$$

From Lemma 1 and the path continuity of the $\sigma^2(.)$ function, we also have that

$$\max_j |(\sigma_j^2 - \sigma_{j-1}^2)(\hat{\sigma}_j^2 - \sigma_j^2)| = O(1/T)o_p(1) = o_p\left(\frac{1}{T}\right).$$

Combining these results with the following, each of which follows straightforwardly from results given previously,

$$\max_j \left|\frac{1}{\hat{\sigma}_j^2}\right| = O_p(1), \quad \max_j \left|\frac{\sigma_{j-1}^2}{\hat{\sigma}_j^2 \hat{\sigma}_{j-1}^2}\right| = O_p(1), \max_j \left|\frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s\right| = O_p(1),$$

we have that

$$\left|\sum_{j=T+1}^{\lfloor Tr \rfloor} \left(\frac{\sigma_j^2}{\hat{\sigma}_j^2} - \frac{\sigma_{j-1}^2}{\hat{\sigma}_{j-1}^2}\right)\left(\frac{1}{\sqrt{T}} \sum_{s=1}^{j-1} \varepsilon_s\right)\right| = o_p(1),$$

as required to establish the stated result in Theorem 1.

A.5 Proof of Theorem 2

Under H_1, and when $\lfloor \tau_1 T \rfloor + 1 \leqslant t \leqslant \lfloor \tau_2 T \rfloor$,

$$SV_T^t = \sum_{j=T+1}^{t} \frac{\Delta y_j}{\hat{\sigma}_{j,N}} = \left(\sum_{j=T+1}^{\lfloor \tau_1 T \rfloor} + \sum_{j=\lfloor \tau_1 T \rfloor + 1}^{t}\right) \frac{\Delta y_j}{\hat{\sigma}_{j,N}} \equiv A_T + B_t.$$

We first look at B_t. When $\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant \lfloor \tau_2 T \rfloor$, we have

$$\Delta y_j = \delta u_{j-1} + e_j.$$

Substituting this into B_t we have

$$B_t = \delta \sum_{j=\lfloor \tau_1 T \rfloor + 1}^{t} \frac{u_{j-1}}{\hat{\sigma}_{j,N}} + \sum_{j=\lfloor \tau_1 T \rfloor + 1}^{t} \frac{e_j}{\hat{\sigma}_{j,N}} \equiv B_{t1} + B_{t2}. \tag{A.8}$$

For B_{t1}, notice with $\phi = 1 + \delta$,

$$|B_{t1}| = \delta N^{1/2} \left| \sum_{j=\lfloor \tau_1 T \rfloor + 1}^{t} \frac{T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} u_{j-1}}{N^{1/2} T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} \hat{\sigma}_{j,N}} \right|$$

$$\leqslant \delta \max_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant t} \left| \frac{1}{N^{1/2} T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} \hat{\sigma}_{j,N}} \right| N^{1/2} \sum_{j=\lfloor \tau_1 T \rfloor + 1}^{t} |T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} u_{j-1}|$$

$$= \frac{\delta}{\min_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant t} |N^{1/2} T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} \hat{\sigma}_{j,N}|} N^{1/2} \sum_{j=\lfloor \tau_1 T \rfloor + 1}^{t} |T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} u_{j-1}|$$

$$= O_p(N^{1/2}(t - \lfloor \tau_1 T \rfloor - 1)),$$

where we have used the result $\min_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant t} |N^{1/2} T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} \hat{\sigma}_{j,N}|$ is nondegenerate to 0 in Lemma A2 and the result of Lemma A3. Using the same argument, we can show that B_{t2} is dominated by B_{t1} in order, so the order of B_t is determined by B_{t1}.

Next we derive a lower bound for the divergence rate of B_{t1}. Notice $|B_{t1}|$ also satisfies

$$|B_{t1}| \geqslant \delta \min_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant t} \left| \frac{1}{N^{1/2} T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} \hat{\sigma}_{j,N}} \right| N^{1/2} \sum_{j=\lfloor \tau_1 T \rfloor + 1}^{t} |T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} u_{j-1}|$$

$$= \frac{\delta}{\max_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant t} |N^{1/2} T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} \hat{\sigma}_{j,N}|} N^{1/2} \sum_{j=\lfloor \tau_1 T \rfloor + 1}^{t} |T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} u_{j-1}|.$$

From Lemma A2, we have $\max_{\lfloor \tau_1 T \rfloor + 1 \leqslant j \leqslant t} |N^{1/2} T^{-1/2} \phi^{-(j-1-\lfloor \tau_1 T \rfloor)} \hat{\sigma}_{j,N}| = O_p(1)$, and using the result of Lemma A3, it follows that B_t diverges at a rate at least as fast as $O_p(N^{1/2}(t - \lfloor \tau_1 T \rfloor - 1))$.

Now, since $A_T = O_p(T^{1/2})$ and clearly does not grow with t, B_t dominates under the alternative, and the derived divergence rate $N^{1/2}(t - \lfloor \tau_1 T \rfloor - 1)$ is clearly higher than the boundary function $c_t \sqrt{t}$ and the claim of the proposition follows.

References

Astill, S., D. I. Harvey, S. J. Leybourne, and A. M. R. Taylor. 2017. Tests for an End-of-Sample Bubble in Financial Time Series. *Econometric Reviews* 36: 651–666.

Astill, S., D. I. Harvey, S. J. Leybourne, R. Sollis, and A. M. R. Taylor. 2018. Real-Time Monitoring for Explosive Financial Bubbles. *Journal of Time Series Analysis* 39: 863–891.

Beare, B. K. 2018. Unit Root Testing with Unstable Volatility. *Journal of Time Series Analysis* 39: 816–835.

Boswijk, H. P., and Y. Zu. 2018. Adaptive Wild Bootstrap Testing for a Unit Root with Nonstationary Volatility. *The Econometrics Journal* 2: 87–113.

Boswijk, H. P., and Y. Zu. 2021. Adaptive Testing for Cointegration with Nonstationary Volatility. *Journal of Business and Economic Statistics*, Forthcoming.

Calvo-Gonzalez, O., R. Shankar, and R. Trezzi. 2010. "Are Commodity Prices More Volatile Now?" World Bank Policy Research Working Paper 5460.

Cavaliere, G. 2004. Unit Root Tests under Time-Varying Variances. *Econometric Reviews* 23: 259–292.

Chu, C.-S., J. M. Stinchcombe, and H. White. 1996. Monitoring Structural Change. *Econometrica* 64: 1045–1065.

Diba, B. T., and H. I. Grossman. 1988. Explosive Rational Bubbles in Stock Prices? *American Economic Review* 78: 520–530.

Evans, G. W. 1991. Pitfalls in Testing for Explosive Bubbles in Asset Prices. *American Economic Review* 81: 922–930.

Fan, J., and Q. Yao. 2003. *Nonlinear Time Series: Nonparametric and Parametric Methods*. New York: Springer.

Hall, P., and W. R. Schucany. 1989. A Local Cross-Validation Algorithm. *Statistics & Probability Letters* 8: 109–117.

Hansen, B. E. 1995. Regression with Nonstationary Volatility. *Econometrica* 63: 1113–1132.

Härdle, W., P. Hall, and J. S. Marron. 1988. How Far Are the Automatically Chosen Regression Smoothing Parameters from Their Optimum? *Journal of the American Statistical Association* 83: 86–95.

Harris, D., and H. Kew. 2017. Adaptive Long Memory Testing under Heteroskedasticity. *Econometric Theory* 33: 755–778.

Harris, D., H. Kew, and A. M. R. Taylor. 2020. Level Shift Estimation in the Presence of Nonstationary Volatility with an Application to the Unit Root Testing Problem. *Journal of Econometrics* 219: 354–388.

Harvey, D. I., S. J. Leybourne, and R. Sollis. 2015. Recursive Right-Tailed Unit Root Tests for an Explosive Asset Price Bubble. *Journal of Financial Econometrics* 13: 166–187.

Harvey, D. I., S. J. Leybourne, R. Sollis, and A. M. R. Taylor. 2016. Tests for Explosive Financial Bubbles in the Presence of Nonstationary Volatility. *Journal of Empirical Finance* 38: 548–574.

Harvey, D. I., S. J. Leybourne, and Y. Zu. 2019. Testing Explosive Bubbles with Time-Varying Volatility. *Econometric Reviews* 38: 1131–1151.

Harvey, D. I., S. J. Leybourne, and Y. Zu. 2020. Sign-Based Unit Root Tests for Explosive Financial Bubbles in the Presence of Nonstationary Volatility. *Econometric Theory* 36: 122–169.

Homm, U., and J. Breitung. 2012. Testing for Speculative Bubbles in Stock Markets: A Comparison of Alternative Methods. *Journal of Financial Econometrics* 10: 198–231.

McMillan, D. G., and M. E. Wohar. 2011. Structural Breaks in Volatility: The Case of UK Sector Returns. *Applied Financial Economics* 21: 578–595.

Pagan, A., and A. Ullah. 1999. *Nonparametric Econometrics*. Cambridge: Cambridge University Press.

Phillips, P. C. B., and S.-P. Shi. 2018. Financial Bubble Implosion and Reverse Regression. *Econometric Theory* 34: 705–753.

Phillips, P. C. B., S.-P. Shi, and J. Yu. 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., Y. Wu, and J. Yu. 2011. Explosive Behavior in the 1990s Nasdaq: When Did Exuberance Escalate Stock Values? *International Economic Review* 52: 201–226.

Phillips, P. C. B., and K.-L. Xu. 2006. Inference in Autoregression under Heteroskedasticity. *Journal of Time Series Analysis* 27: 289–308.

Rapach, D. E., J. K. Strauss, and M. E. Wohar. 2008. "Forecasting Stock Return Volatility in the Presence of Structural Breaks." In D. E. Rapach, and M. E. Wohar (eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Vol. 3 of Frontiers Emerald Series Frontiers of Economics and Globalization. Bingley: Emerald. pp. 381–416.

Schwarz, G. 1978. Estimating the Dimension of a Model. *The Annals of Statistics* 6: 461–464.

Shiryaev, A. N. 1996. *Probability—Graduate Texts in Mathematics*. Vol. 95. New York: Springer.

Vivian, A., and M. E. Wohar. 2012. Commodity Volatility Breaks. *Journal of International Financial Markets, Institutions and Money* 22: 395–422.

Xu, K.-L., and P. C. B. Phillips. 2008. Adaptive Estimation of Autoregressive Models with Time-Varying Variances. *Journal of Econometrics* 142: 265–280.