Estimation of the variance function in structural break autoregressive models with nonstationary and explosive segments*

David I. Harvey, Stephen J. Leybourne and Yang Zu
School of Economics
University of Nottingham

June 8, 2022

Abstract

In this paper we consider estimating the innovation variance function when the conditional mean model is characterized by a structural break autoregressive model, which exhibits multiple unit root, explosive and stationary collapse segments, allowing for behaviour often seen in financial data where bubble and crash episodes are present. Estimating the variance function normally proceeds in two steps: estimating the conditional mean model, then using the residuals to estimate the variance function. In this paper, a nonparametric approach is proposed to estimate the complicated parametric conditional mean model in the first step. The approach turns out to provide a convenient solution to the problem and achieve robustness to any structural break features in the conditional mean model without the need of estimating them parametrically. In the second step, kernel-smoothed squares of the truncated first step residuals are shown to consistently estimate the variance function. In Monte Carlo simulations, we show that our proposed method performs very well in the presence of explosive and stationary collapse segments compared to the popular rolling standard deviation estimator that is commonly used in economics and finance. As an empirical illustration of our new approach, we apply the volatility estimator to recent Bitcoin data.

Keywords: structural break autoregressive model; nonstationary segments; explosive segments; nonparametric variance function estimation; truncation.

*The authors are grateful to two anonymous referees for comments that greatly improved the proposed estimation method and the presentation of the paper. The authors are also grateful to Christian Francq, Alain Hecq, Eiji Kurozumi, Degui Li, Jeroen Rombouts, Yongcheol Shin, Robert Taylor, Jean-Michel Zakoian, and participants of seminars at University of York, Maastricht University, University of Essex, the Center for Research in Economics and Statistics, and the 2019 NBER-NSF Time Series Conference, for helpful comments. The usual disclaimer applies. Correspondence to: David Harvey, School of Economics, University of Nottingham, NG7 2RD, UK. Email: dave.harvey@nottingham.ac.uk.

1 Introduction

In this paper, we consider estimating the variance function of the error in a Structural Break Autoregressive (SBAR) model, which allows for changes between unit root, explosive and stationary collapse behaviour in a time series. Recent progress in the literature on bubble testing (e.g. Phillips et al. (2011, 2015), among others) has provided convincing evidence for the existence of such behaviour in economic data during periods of economic exuberance and crises. The model we consider is of the following form:

$$y_t = \sum_{j=1}^{J} \{1 + \rho_{j1}^* \mathbb{I}([\tau_{j1}^* T] < t \leqslant [\tau_{j2}^* T]) + \rho_{j2}^* \mathbb{I}([\tau_{j2}^* T] < t \leqslant [\tau_{j3}^* T])\} y_{t-1} + u_t, \quad t = 2, \ldots, T \quad (1)$$

where $u_t = \sigma_t \varepsilon_t$ with ε_t a white noise error term with zero mean and unit variance, and we assume $y_1 = O_p(1)$. Here, $\mathbb{I}(.)$ denotes the indicator function and $[.]$ denotes the integer part. Our focus is then on the estimation of σ_t^2, which represents the variance of the SBAR error u_t at time t. The model in (1) has J episodes and we assume $0 < \tau_{11}^* < \tau_{12}^* < \tau_{13}^* < \tau_{21}^* < \ldots < \tau_{J3}^* < 1$. Within each episode j, y_t starts as a unit root process (a standard model for financial asset prices), but is subject to an explosive segment with parameter $\phi_{j1} := 1 + \rho_{j1}^*$ (with $\rho_{j1}^* > 0$) that begins after time $[\tau_{j1}^* T]$ and terminates at time $[\tau_{j2}^* T]$, with a stationary collapse segment with parameter $\phi_{j2} := 1 + \rho_{j2}^*$ (with $\rho_{j2}^* < 0$) running through to time $[\tau_{j3}^* T]$. After time $[\tau_{j3}^* T]$, the process returns to a unit root. The model also admits the possibility of no stationary collapse upon termination of an explosive segment, with the process returning directly to unit root behaviour if $\rho_{j2}^* = 0$. The number of episodes J, the locations of the break points τ_{11}^*, τ_{12}^*, τ_{13}^*, τ_{21}^*, ..., τ_{J3}^*, and the associated AR offset parameters ρ_{11}^*, ρ_{12}^*, ρ_{21}^*, ρ_{22}^*, ..., ρ_{J1}^*, ρ_{J2}^*, are all assumed unknown. When $J = 1$, the model is an extension of Phillips et al. (2011)'s instant crash, one episode bubble model to allow for a stationary crash segment; when $J > 1$, the model is an extension of Phillips et al. (2015)'s multiple episodes bubble model with instant crash to allow for stationary crash segments.

Estimation of the error variance function σ_t^2 relies on obtaining suitable prior estimates of the u_t. In each episode, the AR coefficients take a piecewise constant structure in each segment, with the majority of segments being nonstationary, either containing a unit root or being explosive. One approach would be to consider extending existing parametric estimation methods for the SBAR model along the lines of, inter alia, Davis et al. (2006), Chan et al. (2014), Bai and Perron (1998), Qu and Perron (2007) and Bai and Perron (2003). However, all these methods assume stationarity within each segment, and would need generalising to the nonstationary context involved in model (1). Moreover, such parametric approaches would also necessarily involve complicated steps of model selection (using information criteria or sequential testing) to determine the number of breaks and a follow-up estimation procedure for the timing of the breaks. When the number of episodes J is large and unknown, developing such a procedure suited for variance estimation in our SBAR model, being characterised by nonstationary segments, is far from straightforward.

In this paper we instead pursue a nonparametric approach to estimation of the SBAR model, in order to obtain residuals which can then be used to estimate the variance function. Our approach does not require estimation of either the number of breaks in the SBAR structure, or their timing, thus offering great flexibility in dealing with the conditional mean model. To be specific, we propose use of a local least squares [LLS] estimator to estimate the AR coefficients of the SBAR model, which we then use to obtain residuals. The resulting residuals, when truncated and squared, are then used to obtain a consistent estimator of the variance function.

The intuition underpinning our proposed method is as follows. At any given point in time, the LLS estimator only uses data local to that point; consequently, within the interior of a given segment with a constant AR coefficient, a consistent estimate for the corresponding AR coefficient (and thus the associated residuals) can be obtained. However, the problem with LLS estimation is that it will not be consistent around the time of the structural breaks, due to the fact that at some points the LLS estimator will necessarily use data from two different AR regimes. Consequently, some LLS residuals can be very large in (absolute) magnitude, and can render the subsequent variance estimation problematic. We therefore propose a truncation mechanism to mitigate against the problematic effects of such large residuals, and use the truncated residuals to estimate the variance function using kernel smoothing techniques. Under suitable conditions on the truncation parameter and the bandwidths, we are able to show both pointwise and uniform consistency of the variance function estimator.

While the bubble model we consider in (1) contains a natural ordering of regimes from unit root to explosive to stationary to unit root, the methodology and theoretical analysis of the paper actually applies automatically to models where the ordering of different autoregressive episodes is arbitrary; for example, it is not necessary for our consistency results for there to be a unit root regime prior to the first explosive regime. Our analysis also applies to the case where multiple stationary (or explosive) segments with different AR parameters occur in sequence. This potentially opens the way for applications of our new methodology in fields other than the present context of bubble modelling.

Nonparametric variance function estimation in structural break models has important applications in a number of theoretical and methodological situations. These include, for example, constructing tests for bubbles as considered in Kurozumi et al. (2020), and developing heteroskedasticity-robust real-time monitoring procedures for bubbles, as considered in Astill et al. (2020). The classical rolling standard deviation estimator used widely in macroeconomics and finance, e.g. Officer (1973), Bittlingmayer (1998), Blanchard and Simon (2001) and Mumtaz and Theodoridis (2017), is essentially the square root of a nonparametric variance function estimator which does not account for structural breaks in the conditional mean model. Our Monte Carlo simulation analysis shows that, when explosive and stationary collapse segments exist in the data, it can provide a very imprecise estimate of the true volatility, while the method we propose affords a convenient and superior method in such situations, with robustness to the structural breaks in the AR coefficients.

Nonparametric volatility measures have been widely studied in the statistics and econometrics literature. This includes Hall and Carroll (1989), Fan and Yao (1998) and Kristensen (2012), among others, in a regression model context, and Xu and Phillips (2008), Beare (2018) and Boswijk and Zu (2017) in autoregressive models. However, to the best of our knowledge, no extant papers considering variance estimation allow for structural breaks in the AR coefficients of the conditional mean AR model.

The structure of the paper is as follows. Section 2 presents assumptions for the model, motivates and defines the new estimator of the variance function and shows its pointwise and uniform consistency. Section 3 discusses how our approach can be extended to a model with a more general SBAR(k) structure. In section 4, we study the finite sample performance of the proposed new estimator and compare it with that of a non-truncated alternative version and the commonly used simple rolling standard deviation estimator, using Monte Carlo simulation. Section 5 presents an empirical illustration of our volatility estimation method using Bitcoin price data, while section 6 concludes. Proofs of the main results of the paper can be found in the Appendix, with proofs of the technical lemmas contained in the online Supplementary Appendix. In the remainder of the paper, we adopt the following notation: $\xrightarrow{d}$ denotes convergence in distribution, $\Rightarrow$ weak convergence to a stochastic process. For two sequences a_T and b_T, $a_T \vee b_T$ means taking the sequence with higher order.

2 Model assumptions and volatility estimator

We make the following assumptions regarding the innovation ε_t in model (1):

A1 $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = 1$ with $\mathcal{F}_t$ the natural filtration generated by $\{\varepsilon_t\}_{t\geq 1}$,

$\sup_t E(|\varepsilon_t|^p \,|\mathcal{F}_{t-1}) < \infty$ almost surely for some $p > 6$.

Assumption A1 requires that $\{\varepsilon_t\}$ is a conditionally homoskedastic martingale difference sequence with absolute moments existing up to order 6. The moment condition is justifiable in view of the recent empirical evidence in Francq and Zakoian (2020, 2021), who propose new tests for the existence of moments of financial returns and find the existence of the sixth moment in financial returns. Denoting $\sigma_t = \sigma(t/T)$, then $\sigma(t/T)$ represents both the conditional and the unconditional standard deviation of the SBAR error $\sigma_t\varepsilon_t$ in (1) given information up to time $t-1$. The considered model therefore permits time-variation in both conditional and unconditional variances. Although the conditional homoskedasticity assumption for the sequence $\{\varepsilon_t\}$ rules out GARCH effects in the standardised error, our estimator for the $\sigma(.)$ function will be able to absorb such effects (to a certain degree) in data. In some practical situations, it is certainly useful to consider an extension to allow for GARCH effects, i.e. $E(\varepsilon_t^2|\mathcal{F}_{t-1}) = \omega_t^2$ (with identification assumption $E(\omega_t^2) = 1$). In view of the Cavaliere et al. (2022) recent work on heteroskedastic fractional time series models, such an extension is possible using the relevant cumulant type assumptions for the dependence structure in $\{\varepsilon_t\}$ and the proof strategy therein, although we leave consideration of such an extension for future work.

As a convention, we refer to $\sigma(\tau)$ as the volatility function and $\sigma^2(\tau)$ as the variance function. We also refer to the domain of the $\sigma(.)$ function $[0, 1]$ as a *normalised* time scale, while $t = 1, 2, \ldots, T$ is the *observational* time scale. We assume the following regarding $\sigma(\tau)$:

A2 $\sigma(\tau)$ is a non-stochastic strictly positive, Lipschitz continuous function satisfying $\infty > \sigma_U > \sigma(\tau) > \sigma_L > 0$ for any $\tau \in [0, 1]$.

If the error sequence $\{u_t\}$ was observable, the variance function $\sigma^2(.)$ could be estimated by a kernel smoothing estimator of the form

$$\hat{\sigma}^2(\tau) = \sum_{i=1}^{T} w_{\tau,i} u_i^2 \tag{2}$$

where $w_{\tau,i} = K_h(i/T - \tau)/\sum_{i=1}^{T} K_h(i/T - \tau)$, $K_h(s) = K(s/h)/h$ with $K(.)$ a kernel function and h some bandwidth parameter. In practice, of course, $\{u_t\}$ is not observable and the nonparametric variance estimator is typically constructed by replacing the u_i in (2) with corresponding estimates $\hat{u}_i$. If the dimension of the parameter space is low, it may be plausible to develop a parametric approach to estimating the conditional mean model. However, given the large number of unknown parameters in our mean model (1), particularly given that J is unknown and can be large, such an approach is unappealing. Instead we propose a nonparametric LLS approach that will provide a simple and robust way of estimating the SBAR model, despite the mean model being parametric in nature. To be specific, letting ρ_t generically denote the AR offset parameter of the process at time t, we define a kernel-type LLS estimator for ρ_t as

$$\hat{\rho}_t = \arg\min_{\rho} \sum_{i=1}^{T} G_{h_1}((i-t)/T)(\Delta y_i - \rho y_{i-1})^2 \tag{3}$$

where $G_{h_1}(s) = G(s/h_1)/h_1$, $G(.)$ is a kernel function and h_1 is the bandwidth parameter. Solving analytically, we obtain for $t = 1, \ldots, T$,

$$\hat{\rho}_t = \left(\sum_{i=1}^{T} G_{h_1}((i-t)/T) y_{i-1}^2 \right)^{-1} \left(\sum_{i=1}^{T} G_{h_1}((i-t)/T) y_{i-1} \Delta y_i \right). \tag{4}$$

Given $\hat{\rho}_t$, we define the LLS AR residuals as $\hat{u}_t = \Delta y_t - \hat{\rho}_t y_{t-1}$. Substituting in the definition that $\Delta y_t = \rho_t y_{t-1} + u_t$, we have

$$\hat{u}_t - u_t = -(\hat{\rho}_t - \rho_t) y_{t-1}.$$

This implies that the magnitude of the error arising from using $\hat{u}_t$ instead of u_t depends on both the error of the LLS estimator $\hat{\rho}_t - \rho_t$ and the magnitude of the y_t process at time $t-1$.

LLS estimation is normally used to estimate time-varying parameters that evolve smoothly, whereas here we use it to estimate piecewise constant parameters, where the process can be allowed to have nonstationary and explosive segments. In our context, using the LLS estimator has the advantage of rendering model identification and break point estimation unnecessary, but a problem remains in that it may not be efficacious in time periods around the break points.

Intuitively, the LLS estimator should be consistent in the interior of a given segment, where the AR coefficient is constant. However, in the neighbourhoods on both sides of the regime change points (both of size h_1), consistency may not hold. In such cases, given that $\hat{u}_t - u_t = -(\hat{\rho}_t - \rho_t) y_{t-1}$, this discrepancy could be very large (in absolute value) when the level of y_{t-1} is very high. If such $\hat{u}_i$ were used in place of u_i in (2), a non-negligible error in this second step variance estimator may result.[1]

To deal with this issue, we propose using truncation-based residuals $\hat{u}_t \mathbb{I}(|\hat{u}_t| < \psi_T)$ in place of u_i in (2), where ψ_T is a truncation parameter which slowly diverges to infinity. The idea is that, in the interior of a given segment, since $\hat{\rho}_t$ is consistent, $\hat{u}_i = u_i - (\hat{\rho}_i - \rho_i) y_{i-1}$ will differ from u_i by

[1] In fact, it is clear from the later theoretical analysis of Theorems 1 and 2 that, under the conditions stated for these theorems, a variance estimator simply using the residuals $\hat{u}_i$ will be diverging and inconsisent.

only an $o_p(1)$ term, hence $\hat{u}_i$ will be $O_p(1)$, and the truncation will not be operative as $\psi_T \to \infty$. On the other hand, in the neighbourhoods of the break points, truncating $|\hat{u}_i|$ residuals that are larger than ψ_T to 0 will ensure that the truncated residuals have a magnitude controlled by ψ_T. Because inconsistent estimation of $\hat{\rho}_t$ can only occur for shrinking neighbourhoods of size h_1, careful selection of the order of the truncation parameter ψ_T and the kernel bandwidths ensures that consistency of the variance estimator can be established.

Our variance function estimator takes the following form, for $0 \le \tau \le 1$:

$$\hat{\sigma}^2(\tau) = \sum_{i=1}^{T} w_{\tau,i} \hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T) \tag{5}$$

where $w_{\tau,i} = K_{h_2}(i/T - \tau)/\sum_{i=1}^{T} K_{h_2}(i/T - \tau)$, $K_{h_2}(s) = K(s/h_2)/h_2$, and $K(.)$ is a kernel function and h_2 the associated bandwidth parameter. In this definition of the estimator, the squared residuals at the truncated points are replaced by zeros. As pointed out by a referee, this may induce a finite sample downward bias in the estimated variance. Alternatively, these truncated "observations" can be treated as missing data, without altering the asymptotic properties of the variance estimator. The appropriate modification would then be to define the weight as

$$w_{\tau,i} = \frac{K_{h_2}(i/T - \tau)}{\sum_{i=1}^{T} K_{h_2}(i/T - \tau)\mathbb{I}(|\hat{u}_i| < \psi_T)}. \tag{6}$$

We found the finite sample performance of the estimator in (6) to be markedly superior. Therefore, in our Monte Carlo simulations and empirical analysis, the weight definition in (6) is used throughout.

Next we study the asymptotic properties of the truncation-based kernel smoothing estimator for the variance function. We make the following additional assumptions:

B1 $G(.)$ is strictly positive on $[-1, 1]$, and $0 < \int_{-1}^{1} G(u)du = \gamma < \infty$.

B2 $h_1 \to 0$ as $T \to \infty$ with

$$\frac{\log(T)}{T^{1-\frac{4}{p-1}} h_1^{1-\frac{2}{p-1}}} \to 0 \tag{7}$$

with $p > 6$ as specified in Assumption A1.

C1 $K(.)$ is a bounded non-negative function defined on the real line, Lipschitz continuous, and satisfies $\int_{-\infty}^{\infty} K(u)du = 1$ and $\int_{-\infty}^{\infty} |K(u)u|du < \infty$.

C2 As $T \to \infty$, $\psi_T \to \infty$.

C3 As $T \to \infty$, $h_2 \to 0$. Defining $a = h_1/h_2$, a satisfies $a\psi_T^4 \to 0$.

The following theorem details the large sample properties of the volatility function estimator $\hat{\sigma}^2(\tau)$:

Theorem 1. *Under conditions A1-A2, B1-B2, C1-C3, then for all* $\tau \in [0,1]$*,*

$$\hat{\sigma}^2(\tau) - \sigma^2(\tau) = o_p(1).$$

That is, $\hat{\sigma}^2(\tau)$ *is pointwise consistent for* $\sigma^2(\tau)$ *for all* $\tau \in [0,1)$*.*

In the next theorem, we show that the pointwise consistency result can be strengthened to a uniform consistency result, under a stronger bandwidth condition on h_2:

Theorem 2. *Under the conditions of Theorem 1, if in addition* $\log(T)/(T^{1-\frac{4}{p}}h_2) \to 0$*, we have*

$$\sup_{\tau \in [0,1]} \left|\hat{\sigma}^2(\tau) - \sigma^2(\tau)\right| = o_p(1).$$

Note that in Assumption B1 we impose $G(.)$ to be a two-sided kernel which is strictly positive over $[-1,1]$. For the result in Theorem 2, this is a requirement and hence a one-sided kernel is not permitted. Examples of a kernel for $G(.)$ that satisfies the requirements of Assumption B1 are the uniform kernel $G(u) = \mathbb{I}(-1 \leqslant u \leqslant 1)$ and the truncated Gaussian kernel $G(u) = (2\pi)^{-1/2}\exp(-u^2/2)\mathbb{I}(-1 \leqslant u \leqslant 1)$. In Assumption B2, (7) indicates the trade-off between the moment condition and the bandwidth restriction. If we take the smallest integer satisfying the $p > 6$ condition, i.e. $p = 7$, (7) becomes equivalent to $Th_1^2 \to \infty$ up to a logarithmic term. When ε_t is Gaussian so p can be arbitrarily large, (7) would be close to the commonly-used assumption $Th_1 \to \infty$. Assumptions on the second step kernel $K(.)$ are standard. The classical Gaussian kernel $K(u) = (2\pi)^{-1/2}\exp(-u^2/2)$ satisfies the conditions for $K(.)$ in Assumption C1. Assumption C2 only requires that ψ_T diverges, but technically Assumptions B2 and C3 also imply that ψ_T cannot diverge too fast, because $h_1 \to 0$ cannot be too fast (Assumptions B2 and C3), and $h_1 = o(h_1/h_2)$ (Assumption C3). A slowly diverging logarithmic rate for the truncation level ψ_T will satisfy Assumptions C2 and C3. For example, when $\psi_T = O(\log(T))$, and the rate difference a between h_1 and h_2 is some polynomial rate $T^{-\delta}$, $\delta > 0$, both Assumptions C2 and C3 will be satisfied. For the condition $\log(T)/(T^{1-\frac{4}{p}}h_2) \to 0$, when $p = 7$, it becomes equivalent to $Th_2^{7/3} \to \infty$ up to a logarithmic term; when ε_t is Gaussian and p can be arbitrarily large, it becomes $Th_2 \to \infty$ (up to a logarithmic term). Although the bandwidth conditions for h_1 and h_2 appear complicated, the conditions $Th_1^3 \to \infty$ and $Th_2^3 \to \infty$ would be sufficient for all possible values of p, and these are comparable with those in Fan and Yao (1998).

2.1 Bandwidth and truncation level selection

For given choices of $G(.)$ and $K(.)$, the associated bandwidths h_1 and h_2 can be selected to optimise the performance of the estimators. The bandwidths can be subjectively tuned to trade off the bias and variance of the nonparametric estimates, or chosen by data to minimise some criteria that are related to the prediction/estimation error. Here we outline a data-driven approach to choosing the bandwidths.

We propose using the so-called "leave-one-out" cross-validation (CV) method to select h_1 and h_2. To ensure the selected h_1 and h_2 satisfy the multiple bandwidth conditions (i.e. Assumptions B2 and C3, and potentially the extra condition for h_2 in Theorem 2), we follow the approach proposed in Patilea and Raissi (2012, 2014) to select h_1 and h_2 with suitable pre-specified rates. To be specific, we first set $h_1 = b_1\kappa_{1,T}$ and $h_2 = b_2\kappa_{2,T}$, with $\kappa_{1,T}$ and $\kappa_{2,T}$ chosen to be functions of T that satisfy the required rate conditions, and then use the leave-one-out CV method to select the constants b_1 and b_2 by optimizing the constants over fixed sets.

For the first step bandwidth $h_1 = b_1\kappa_{1,T}$, the CV criterion is defined as

$$CV_1(b_1) = \frac{1}{T}\sum_{t=1}^{T}(\Delta y_t - \hat{\rho}_{t,-}y_{t-1})^2$$

where $\hat{\rho}_{t,-}$ is the LLS estimator leaving out the observation at time t. The CV bandwidth is then selected by minimizing the CV criterion across $b_1 \in \mathcal{B}_1 := [b_1^{min}, b_1^{max}]$:

$$h_1 = \kappa_{1,T} \cdot \arg\min_{b_1 \in \mathcal{B}_1} CV_1(h).$$

The second step estimation for $\hat{\sigma}_t^2$ involves truncation. As discussed above, setting the truncation level as $\psi_T = c\log(T)$, where c is a positive constant, will be a valid choice. Practically, given that in the model (1) an initial unit root regime occurs prior to the first explosive episode, we propose setting c to be the sample standard deviation of Δy_t in this initial sub-sample, thereby acting as a simple measure for the average error volatility. Such a sub-sample is to be used because when explosive behaviour is present in the data, the sample standard deviation of Δy_t across the full sample will over-state the average error volatility. The advantage of using an average error volatility measure is that the truncation is then adapted to the (average) level of volatility in the specific dataset under study. In practical applications, although the onset of the initial putative explosive regime is unknown, it is usually straightforward to informally identify a sub-period of the data that is free of explosive behaviour. Denoting this average volatility estimate as $\breve{\sigma}$, the truncation level is then set as $\psi_T = \breve{\sigma}\log(T)$. Following truncation of the $\hat{u}_t$

series, leave-one-out cross-validation is then used to select the bandwidth h_2, which is defined as

$$h_2 = \kappa_{2,T} \cdot \arg\min_{b_2 \in \mathcal{B}_2} \sum_{t=1}^{T} (\hat{u}_t^2 \mathbb{I}(|\hat{u}_t| < \psi_T) - \hat{\sigma}_{t,-}^2)^2$$

where $\mathcal{B}_2 := [b_2^{min}, b_2^{max}]$ and $\hat{\sigma}_{t,-}^2$ denotes the variance estimator defined in (5) using observations excluding the squared truncated residuals $\hat{u}_t^2 \mathbb{I}(|\hat{u}_i| < \psi_T)$.

Notice that in our model specifications, second order dependence of the ε_t sequence is unspecified and it is possible that $Cov(\varepsilon_t^2, \varepsilon_s^2) \neq 0$. In practical applications, it is also the case that nontrivial dependence may exist in the squared residuals used in the second step estimator. Consequently, leave-one-out CV may not be the best choice for bandwidth selection in this case, given the findings of Chu and Marron (1991), and an alternative possibility would be to leave more observations out in computing the CV criterion function, i.e. using leave-p-out CV with $p > 1$.

3 Allowing for serial correlation in the errors of the SBAR model

Our variance function estimator can also be applied to an SBAR model which allows for higher order serial correlation in the errors. A $(k+1)$th order SBAR variant of our model can be obtained by augmenting (1) with lagged differences of Δy_t in each of the segments involved. Specifically:

$$\begin{aligned} y_t \;=\; & \sum_{j=1}^{J} [\{1 + \rho_{j1}^* \mathbb{I}([\tau_{j1}^* T] < t \leqslant [\tau_{j2}^* T]) + \rho_{j2}^* \mathbb{I}([\tau_{j2}^* T] < t \leqslant [\tau_{j3}^* T])\} y_{t-1} \\ & + \{\delta_{111} \Delta y_{t-1} + \cdots + \delta_{11k} \Delta y_{t-k}\} \mathbb{I}(t \leqslant [\tau_{11}^* T]) \\ & + \{\delta_{j21} \Delta y_{t-1} + \cdots + \delta_{j2k} \Delta y_{t-k}\} \mathbb{I}([\tau_{j1}^* T] < t \leqslant [\tau_{j2}^* T]) \\ & + \{\delta_{j31} \Delta y_{t-1} + \cdots + \delta_{j3k} \Delta y_{t-k}\} \mathbb{I}([\tau_{j2}^* T] < t \leqslant [\tau_{j3}^* T]) \\ & + \{\delta_{j41} \Delta y_{t-1} + \cdots + \delta_{j4k} \Delta y_{t-k}\} \mathbb{I}([\tau_{j3}^* T] < t \leqslant [\tau_{j+1,1}^* T])] + u_t \end{aligned} \tag{8}$$

where $\tau_{J+1,1}^* = 1$. Note that this model allows the lagged difference coefficients to be different across the four segments in each episode j, as well as across j. In this model, our LLS estimator can be applied analogously to estimate all the coefficients in this $(k+1)$th order SBAR model, simply by augmenting the LLS regression equation appropriately, i.e. replacing $\hat{\rho}_t$ in (3) with

$$\{\hat{\rho}_t, \hat{\delta}_1, ..., \hat{\delta}_k\} = \arg \min_{\{\rho_t, \delta_1, ..., \delta_k\}} \sum_{i=1}^{T} G_{h_1}((i-t)/T)(\Delta y_i - \rho y_{i-1} - \delta_1 \Delta y_{t-1} - \cdots - \delta_k \Delta y_{t-k})^2$$

which again has an analytical solution, a vector analogue of (4). The residuals from such LLS estimation can then be truncated and used to estimate the $\sigma(.)$ function as in the simpler case of the previous section. This extension further highlights the benefit of using a LLS nonparametric approach, as opposed to a high dimensional parametric estimation of (8).

4 Monte Carlo simulation

In this section, we use Monte Carlo simulation to compare the performance of our truncation-based variance estimator $\hat{\sigma}_t^2$ with two related estimators across a number of different specifications for the volatility and explosive episodes. The first estimator we compare with is the square of the classical rolling standard deviation estimator, which essentially applies the usual sample variance formula to Δy_t over rolling windows; we denote this estimator by $\hat{\sigma}_{t,rw}^2$. The second estimator used for comparison is our two-step estimator but *without* truncation (which we denote by $\hat{\sigma}_{t,nt}^2$); this is included to demonstrate the effect of truncation in our proposed estimation procedure.

We use the uniform kernel over $[-1, 1]$ for both the first step kernel function $G(.)$ and the second step kernel function $K(.)$. This choice of kernel functions is comparable with the rolling estimator, which applies equal weights to squared demeaned returns in a fixed window. The bandwidths used in this simulation study are $h_1 = T^{-1/3}/log(T)$ and $h_2 = T^{-1/4}$; it is straightforward to confirm that these bandwidths satisfy the imposed bandwidth conditions. We set the truncation level according to $\psi_T = \breve{\sigma} \log(T)$ with $\breve{\sigma}$ calculated using the standard deviation of Δy_t over the first 10% of the sample. The same bandwidth h_2 is then also used for the rolling estimator. We use 1000 Monte Carlo replications in all the simulations.

For simplicity, we simulate the DGP (1) with $J = 1$, so we suppress the dependence on j in the notation of this section. We use $T = 200, 500, 1000$, $y_1 = 5$, $\varepsilon_t \sim NIID(0, 1)$, and $\rho_1^* = \{0, 0.02, 0.04, 0.06\}$ for the explosive parameter magnitudes, along with $\rho_2^* = -\rho_1^*$, such that the stationary collapse offset is of the same magnitude as the explosive offset. We consider the break timings $(\tau_1^*, \tau_2^*, \tau_3^*) = \{(0.3, 0.7, 0.8),(0.4, 0.6, 0.7)\}$. In each case we consider the following volatility specifications for $\sigma(\tau)$, setting $\sigma^* = 0.03$:

(a) Constant volatility: $\sigma(\tau) = \sigma^* \;\; \forall \tau$.

(b) Early smooth upward shift: $\sigma(\tau) = \sigma^* + 2\sigma^* \frac{1}{1+\exp\{-20(\tau-0.3)\}}$.

(c) Mid-sample smooth upward shift: $\sigma(\tau) = \sigma^* + 2\sigma^* \frac{1}{1+\exp\{-20(\tau-0.5)\}}$.

(d) Late smooth upward shift: $\sigma(\tau) = \sigma^* + 2\sigma^* \frac{1}{1+\exp\{-20(\tau-0.8)\}}$.

(e) Double smooth shift: $\sigma(\tau) = \sigma^* + 2\sigma^* \frac{1}{1+\exp\{-20(\tau-0.3)\}} - 2\sigma^* \frac{1}{1+\exp\{-20(\tau-0.7)\}}$.

(f) Upward trend: $\sigma(\tau) = \sigma^* + 2\sigma^* \tau$.

In cases (b) to (d), the volatility undergoes a (logistic) smooth transition from σ^* to $3\sigma^*$ at different points in time. In case (e), the volatility transitions from σ^* to $3\sigma^*$ early in the sample before returning to σ^* at a later point in the sample period.

We compare the performance of $\hat{\sigma}_t^2$, $\hat{\sigma}_{t,rw}^2$ and $\hat{\sigma}_{t,nt}^2$ using two measures: Root Mean Integrated Squared Error (RMISE) and Mean Integrated Absolute Percentage Error (MIAPE). Specifically, and with a slight abuse of notation, letting $\hat{\sigma}^2(\tau)$ denote an arbitrary estimator for $\sigma^2(\tau)$, these measures are defined as follows:

$$
\begin{aligned}
\text{RMISE} &= \sqrt{E \int_0^1 (\hat{\sigma}(\tau) - \sigma(\tau))^2 \mathrm{d}\tau}, \\
\text{MIAPE} &= E \int_0^1 \left| \frac{\hat{\sigma}(\tau) - \sigma(\tau)}{\sigma(\tau)} \right| \mathrm{d}\tau.
\end{aligned}
$$

The integration and the expectation above cannot be evaluated exactly, so in our simulations the integrals are approximated by T discretised points, with the expectations based on the means across replications.

We report the results for both measures in Tables 1-4 for different combinations of break timings and error summary measures. From Tables 1-4, we observe that in the case of no explosivity ($\rho_1^* = 0$), all the estimators perform similarly. In the cases where an explosive (and stationary collapse) segment exists, the error measures for our proposed estimator $\hat{\sigma}_t^2$ change relatively little as the explosive magnitude increases, while the rolling estimator $\hat{\sigma}_{t,rw}^2$ and the "no truncation" version of our estimator $\hat{\sigma}_{t,nt}^2$ have errors that increase substantially with both $\rho_1^* = -\rho_2^*$ and T, with particularly high error measures observed for the large explosive magnitude settings. As might be expected, the $\hat{\sigma}_{t,rw}^2$ estimator is overall the least accurate of the three in the presence of an explosive segment, followed by $\hat{\sigma}_{t,nt}^2$, while our proposed $\hat{\sigma}_t^2$ estimator offers by far the best performance, with truncation clearly performing a very important role in improving the accuracy of the variance function estimation. We also observe that the error measures associated with $\hat{\sigma}_t^2$ are, without exception, monotonically decreasing in T (in contrast

Volatility	$\rho_1^* = -\rho_2^*$	$T = 200$ $\hat{\sigma}_t^2$	$\hat{\sigma}_{t,nt}^2$	$\hat{\sigma}_{t,rw}^2$	$T = 500$ $\hat{\sigma}_t^2$	$\hat{\sigma}_{t,nt}^2$	$\hat{\sigma}_{t,rw}^2$	$T = 1000$ $\hat{\sigma}_t^2$	$\hat{\sigma}_{t,nt}^2$	$\hat{\sigma}_{t,rw}^2$
Model (a)	0.0000	0.0026	0.0026	0.0023	0.0018	0.0018	0.0016	0.0013	0.0013	0.0012
	0.0200	0.0122	0.0617	0.1804	0.0036	0.6397	1.5687	0.0014	27.9658	61.2222
	0.0400	0.0058	0.7288	1.5514	0.0035	63.3387	> 100	0.0027	> 100	> 100
	0.0600	0.0075	5.0935	9.3166	0.0056	> 100	> 100	0.0045	> 100	> 100
Model (b)	0.0000	0.0178	0.0106	0.0104	0.0114	0.0073	0.0071	0.0078	0.0054	0.0053
	0.0200	0.0166	0.0383	0.1459	0.0111	0.6051	1.5335	0.0079	28.1420	61.6480
	0.0400	0.0177	0.6892	1.5111	0.0116	63.5299	> 100	0.0080	> 100	> 100
	0.0600	0.0181	5.0623	9.2941	0.0117	> 100	> 100	0.0082	> 100	> 100
Model (c)	0.0000	0.0154	0.0093	0.0090	0.0101	0.0066	0.0065	0.0071	0.0051	0.0050
	0.0200	0.0143	0.0399	0.1504	0.0103	0.6051	1.5314	0.0075	27.9413	61.2074
	0.0400	0.0164	0.6924	1.5139	0.0115	63.3075	> 100	0.0082	> 100	> 100
	0.0600	0.0177	5.0558	9.2788	0.0124	> 100	> 100	0.0091	> 100	> 100
Model (d)	0.0000	0.0121	0.0102	0.0098	0.0085	0.0073	0.0071	0.0063	0.0055	0.0054
	0.0200	0.0115	0.0525	0.1684	0.0076	0.6278	1.5555	0.0057	27.9560	61.2113
	0.0400	0.0114	0.7134	1.5359	0.0082	63.3266	> 100	0.0062	> 100	> 100
	0.0600	0.0129	5.0774	9.3006	0.0093	> 100	> 100	0.0072	> 100	> 100
Model (e)	0.0000	0.0157	0.0135	0.0137	0.0107	0.0092	0.0092	0.0078	0.0068	0.0068
	0.0200	0.0176	0.0537	0.1647	0.0109	0.6253	1.5553	0.0075	28.1605	61.6673
	0.0400	0.0154	0.7122	1.5344	0.0102	63.5506	> 100	0.0077	> 100	> 100
	0.0600	0.0150	5.0861	9.3180	0.0104	> 100	> 100	0.0079	> 100	> 100
Model (f)	0.0000	0.0095	0.0066	0.0060	0.0058	0.0045	0.0041	0.0039	0.0034	0.0031
	0.0200	0.0078	0.0425	0.1543	0.0051	0.6148	1.5439	0.0039	28.1063	61.5582
	0.0400	0.0095	0.6994	1.5214	0.0061	63.5088	> 100	0.0042	> 100	> 100
	0.0600	0.0110	5.0682	9.2955	0.0069	> 100	> 100	0.0049	> 100	> 100

Note: $\hat{\sigma}_t^2$, $\hat{\sigma}_{t,nt}^2$ and $\hat{\sigma}_{t,rw}^2$ denote the truncation-based variance estimator, the estimator without truncation, and the classical rolling window estimator, respectively.

Table 1: RMISE for variance estimates: $(\tau_1^*, \tau_2^*, \tau_3^*) = (0.3, 0.7, 0.8)$

to those for $\hat{\sigma}_{t,nt}^2$ and $\hat{\sigma}_{t,rw}^2$), in line with what we would expect from our theoretical consistency results derived in section 2. It is encouraging to see that the attractive behaviour of the new $\hat{\sigma}_t^2$ estimator is maintained across all the different volatility specifications employed and the different SBAR settings considered for the explosive and stationary segment magnitudes and timings. These findings are also confirmed by both the RMISE and MIAPE error measures.

5 Empirical illustration: Bitcoin data 2016-2019

In this section, we compare the empirical performance of our estimator with the untruncated estimator and the rolling standard deviation estimator, using a dataset for the logarithms of Bitcoin prices as y_t. The rolling standard deviation, at various fixed lengths of intervals (e.g. 30-day, 60-day, 120-day, etc.), is available in databases such as the Bloomberg Terminal and Thompson Reuters Datastream.

Bitcoin is a digital asset designed to work as a medium of exchange that uses cryptography. It has long been recognised as a speculative asset among financial economists. Bitcoin gained much media exposure in late 2017 because of the rapid increase and decrease in its price. Its unit price increased dramatically, reaching a historical high on December 17th, 2017, after which

Volatility	$\rho_1^* = -\rho_2^*$	$T=200$ $\hat\sigma_t^2$	$\hat\sigma_{t,nt}^2$	$\hat\sigma_{t,rw}^2$	$T=500$ $\hat\sigma_t^2$	$\hat\sigma_{t,nt}^2$	$\hat\sigma_{t,rw}^2$	$T=1000$ $\hat\sigma_t^2$	$\hat\sigma_{t,nt}^2$	$\hat\sigma_{t,rw}^2$
Model (a)	0.0000	0.0702	0.0701	0.0604	0.0474	0.0474	0.0419	0.0353	0.0353	0.0315
	0.0200	0.2971	1.5600	5.1065	0.0869	14.5626	39.6193	0.0367	>100	>100
	0.0400	0.1634	18.3270	42.5372	0.0927	>100	>100	0.0695	>100	>100
	0.0600	0.1971	>100	>100	0.1412	>100	>100	0.1075	>100	>100
Model (b)	0.0000	0.2251	0.1599	0.1630	0.1488	0.1056	0.1073	0.1036	0.0770	0.0777
	0.0200	0.2155	0.4166	1.5721	0.1478	4.6226	13.0841	0.1058	>100	>100
	0.0400	0.2395	5.8927	14.5309	0.1602	>100	>100	0.1115	>100	>100
	0.0600	0.2520	42.5400	85.2695	0.1668	>100	>100	0.1161	>100	>100
Model (c)	0.0000	0.1820	0.1399	0.1439	0.1273	0.0972	0.0985	0.0924	0.0735	0.0744
	0.0200	0.1803	0.4631	2.0429	0.1356	4.9954	15.3439	0.1018	>100	>100
	0.0400	0.2407	6.7788	17.9248	0.1772	>100	>100	0.1304	>100	>100
	0.0600	0.2875	48.7844	>100	0.2137	>100	>100	0.1602	>100	>100
Model (d)	0.0000	0.1487	0.1486	0.1527	0.1086	0.1048	0.1067	0.0820	0.0784	0.0797
	0.0200	0.1956	0.9880	3.6392	0.1076	10.1845	28.6784	0.0760	>100	>100
	0.0400	0.2054	12.5623	30.6302	0.1426	>100	>100	0.1097	>100	>100
	0.0600	0.2603	88.6872	>100	0.1916	>100	>100	0.1484	>100	>100
Model (e)	0.0000	0.2508	0.2543	0.2741	0.1753	0.1673	0.1775	0.1291	0.1212	0.1276
	0.0200	0.3388	0.9912	2.9853	0.1994	8.7940	23.5196	0.1247	>100	>100
	0.0400	0.2781	11.4496	25.9884	0.1754	>100	>100	0.1308	>100	>100
	0.0600	0.2605	80.1094	>100	0.1826	>100	>100	0.1360	>100	>100
Model (f)	0.0000	0.1119	0.0861	0.0852	0.0694	0.0588	0.0584	0.0483	0.0441	0.0436
	0.0200	0.0991	0.4806	1.9777	0.0654	5.8051	16.5738	0.0491	>100	>100
	0.0400	0.1279	7.3803	18.1855	0.0839	>100	>100	0.0582	>100	>100
	0.0600	0.1586	53.2021	>100	0.1037	>100	>100	0.0740	>100	>100

Note: $\hat\sigma_t^2$, $\hat\sigma_{t,nt}^2$ and $\hat\sigma_{t,rw}^2$ denote the truncation-based variance estimator, the estimator without truncation, and the classical rolling window estimator, respectively.

Table 2: MIAPE for variance estimates: $(\tau_1^*, \tau_2^*, \tau_3^*) = (0.3, 0.7, 0.8)$

Volatility	$\rho_1^* = -\rho_2^*$	$T=200$ $\hat\sigma_t^2$	$\hat\sigma_{t,nt}^2$	$\hat\sigma_{t,rw}^2$	$T=500$ $\hat\sigma_t^2$	$\hat\sigma_{t,nt}^2$	$\hat\sigma_{t,rw}^2$	$T=1000$ $\hat\sigma_t^2$	$\hat\sigma_{t,nt}^2$	$\hat\sigma_{t,rw}^2$
Model (a)	0.0000	0.0026	0.0026	0.0023	0.0018	0.0018	0.0016	0.0013	0.0013	0.0012
	0.0200	0.0102	0.0194	0.0682	0.0079	0.0710	0.1953	0.0063	0.5140	1.1450
	0.0400	0.0120	0.1292	0.2953	0.0058	1.2052	2.2215	0.0030	51.3026	81.2238
	0.0600	0.0103	0.4541	0.8540	0.0062	12.2071	18.6075	0.0047	>100	>100
Model (b)	0.0000	0.0178	0.0106	0.0104	0.0114	0.0073	0.0071	0.0078	0.0054	0.0053
	0.0200	0.0167	0.0113	0.0452	0.0105	0.0481	0.1645	0.0072	0.4933	1.1344
	0.0400	0.0167	0.0982	0.2588	0.0108	1.1782	2.2005	0.0075	52.2679	82.7706
	0.0600	0.0171	0.4171	0.8161	0.0109	12.2718	18.7245	0.0075	>100	>100
Model (c)	0.0000	0.0154	0.0093	0.0090	0.0101	0.0066	0.0065	0.0071	0.0051	0.0050
	0.0200	0.0146	0.0145	0.0521	0.0093	0.0533	0.1712	0.0063	0.4896	1.1194
	0.0400	0.0145	0.1065	0.2680	0.0101	1.1775	2.1924	0.0077	51.3372	81.2934
	0.0600	0.0153	0.4257	0.8235	0.0110	12.1818	18.5828	0.0085	>100	>100
Model (d)	0.0000	0.0121	0.0102	0.0098	0.0085	0.0073	0.0071	0.0063	0.0055	0.0054
	0.0200	0.0148	0.0208	0.0654	0.0114	0.0694	0.1917	0.0091	0.5116	1.1417
	0.0400	0.0149	0.1239	0.2882	0.0105	1.2012	2.2165	0.0073	51.3001	81.2209
	0.0600	0.0150	0.4469	0.8457	0.0108	12.2030	18.6026	0.0080	>100	>100
Model (e)	0.0000	0.0157	0.0135	0.0137	0.0107	0.0092	0.0092	0.0078	0.0068	0.0068
	0.0200	0.0152	0.0175	0.0536	0.0106	0.0547	0.1739	0.0082	0.5014	1.1440
	0.0400	0.0155	0.1099	0.2725	0.0102	1.1891	2.2126	0.0072	52.2747	82.7769
	0.0600	0.0151	0.4314	0.8316	0.0097	12.2829	18.7364	0.0070	>100	>100
Model (f)	0.0000	0.0095	0.0066	0.0060	0.0058	0.0045	0.0041	0.0039	0.0034	0.0031
	0.0200	0.0087	0.0122	0.0521	0.0056	0.0559	0.1754	0.0043	0.5001	1.1361
	0.0400	0.0086	0.1085	0.2709	0.0056	1.1880	2.2074	0.0040	51.8282	82.0673
	0.0600	0.0092	0.4297	0.8283	0.0061	12.2434	18.6739	0.0045	>100	>100

Note: $\hat\sigma_t^2$, $\hat\sigma_{t,nt}^2$ and $\hat\sigma_{t,rw}^2$ denote the truncation-based variance estimator, the estimator without truncation, and the classical rolling window estimator, respectively.

Table 3: RMISE for variance estimates: $(\tau_1^*, \tau_2^*, \tau_3^*) = (0.4, 0.6, 0.7)$

Volatility	$\rho_1^* = -\rho_2^*$	$T = 200$			$T = 500$			$T = 1000$		
		$\hat{\sigma}_t^2$	$\hat{\sigma}_{t,nt}^2$	$\hat{\sigma}_{t,rw}^2$	$\hat{\sigma}_t^2$	$\hat{\sigma}_{t,nt}^2$	$\hat{\sigma}_{t,rw}^2$	$\hat{\sigma}_t^2$	$\hat{\sigma}_{t,nt}^2$	$\hat{\sigma}_{t,rw}^2$
Model (a)	0.0000	0.0702	0.0701	0.0604	0.0474	0.0474	0.0419	0.0353	0.0353	0.0315
	0.0200	0.2642	0.5093	1.8923	0.1994	1.6030	4.7974	0.1438	10.4224	25.3197
	0.0400	0.3245	3.3386	8.1399	0.1575	26.9090	53.1514	0.0755	> 100	> 100
	0.0600	0.2851	11.6579	23.2564	0.1563	> 100	> 100	0.1104	> 100	> 100
Model (b)	0.0000	0.2251	0.1599	0.1630	0.1488	0.1056	0.1073	0.1036	0.0770	0.0777
	0.0200	0.2130	0.1668	0.5432	0.1387	0.4197	1.4192	0.0956	3.3264	8.3430
	0.0400	0.2147	0.9515	2.6421	0.1430	8.8386	17.8896	0.1015	> 100	> 100
	0.0600	0.2202	3.7391	7.9741	0.1456	91.3034	> 100	0.1015	> 100	> 100
Model (c)	0.0000	0.1820	0.1399	0.1439	0.1273	0.0972	0.0985	0.0924	0.0735	0.0744
	0.0200	0.1874	0.2186	0.8152	0.1240	0.5873	2.0424	0.0869	4.2206	10.9465
	0.0400	0.2031	1.3906	3.8222	0.1472	11.7341	24.2712	0.1143	> 100	> 100
	0.0600	0.2281	5.2921	11.2538	0.1706	> 100	> 100	0.1328	> 100	> 100
Model (d)	0.0000	0.1487	0.1486	0.1527	0.1086	0.1048	0.1067	0.0820	0.0784	0.0797
	0.0200	0.2788	0.4531	1.5290	0.2068	1.3701	4.0704	0.1499	9.2523	22.2931
	0.0400	0.3047	2.6422	6.6014	0.2015	23.2374	45.6425	0.1334	> 100	> 100
	0.0600	0.3206	9.4154	18.9759	0.2266	> 100	> 100	0.1655	> 100	> 100
Model (e)	0.0000	0.2508	0.2543	0.2741	0.1753	0.1673	0.1775	0.1291	0.1212	0.1276
	0.0200	0.2626	0.3317	0.9235	0.1901	0.6740	2.1076	0.1439	4.3972	11.1099
	0.0400	0.2760	1.6299	4.1058	0.1761	12.1990	24.8113	0.1200	> 100	> 100
	0.0600	0.2562	5.8556	12.0245	0.1624	> 100	> 100	0.1168	> 100	> 100
Model (f)	0.0000	0.1119	0.0861	0.0852	0.0694	0.0588	0.0584	0.0483	0.0441	0.0436
	0.0200	0.1059	0.1664	0.7069	0.0687	0.5927	1.9933	0.0533	4.5926	11.4387
	0.0400	0.1101	1.2975	3.5191	0.0705	12.0997	24.2961	0.0520	> 100	> 100
	0.0600	0.1219	5.0564	10.5327	0.0811	> 100	> 100	0.0621	> 100	> 100

Note: $\hat{\sigma}_t^2$, $\hat{\sigma}_{t,nt}^2$ and $\hat{\sigma}_{t,rw}^2$ denote the truncation-based variance estimator, the estimator without truncation, and the classical rolling window estimator, respectively.

Table 4: MIAPE for variance estimates: $(\tau_1^*, \tau_2^*, \tau_3^*) = (0.4, 0.6, 0.7)$

the price underwent a rapid depreciation, losing a third of its value by December 30th. Harvey et al. (2020) studied the daily level price data from late 2017 to early 2018 using a sign-based approach to bubble testing and dating, and confirmed the existence of explosive behaviour from November 13th 2017 to December 7th 2017, a period before the highest value was reached. In this illustration, we focus on the log Bitcoin price over a longer period of daily closing prices, from January 1st 2016 to April 2nd 2019, obtained from Yahoo Finance. Figure 1 provides a plot of the log Bitcoin price.

We use the same uniform kernels as in the Monte Carlo simulations for our proposed estimator $\hat{\sigma}_t^2$ from (5), the non-truncated estimator $\hat{\sigma}_{t,nt}^2$ and the rolling standard deviation estimator $\hat{\sigma}_{t,rw}^2$ discussed in section 4. The truncation level is again chosen as described in section 2.1 with $\breve{\sigma}$ calculated using the standard deviation of Δy_t over the first 10% of the sample. The computed actual truncation level is roughly 7 times the standard deviation $\breve{\sigma}$. For the bandwidths, we implemented the leave-one-out CV methods discussed in section 2.1 with $h_1 = b_1 T^{-1/3}/\log(T)$ and $h_2 = b_2 T^{-1/4}$, and the optimisation for both b_1 and b_2 is performed over a grid from 1/5 to 5 with increment 0.1 (i.e. $\mathcal{B}_1 = \mathcal{B}_2 = \{0.2, 0.3, ..., 4.9, 5\}$). The selected h_1 corresponds to performing the first step local least squares using a window of 40 evenly split lead and lag ob-

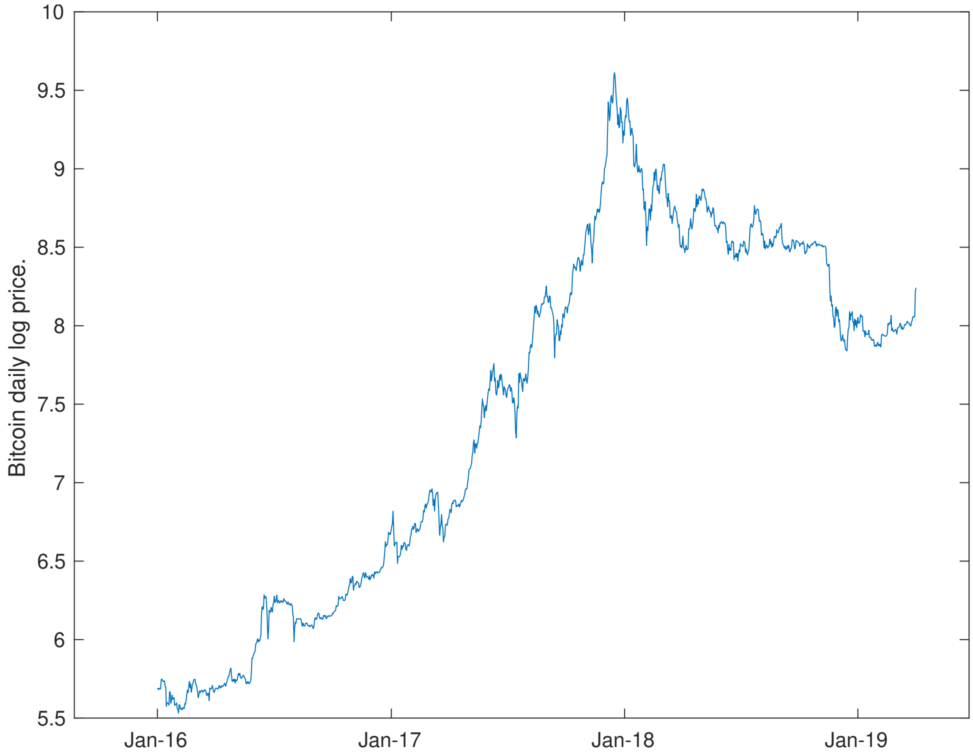

Figure 1: Bitcoin log price.

servations, and the selected h_2 corresponds to performing the second step variance estimation using a window of 120 evenly split lead and lag observations. To account for the possibility of non-trivial dependence in the truncated squared residuals sequence, we also implemented leave-p-out CV for the selection of h_2; we found very similar bandwidths were obtained as those selected by leave-one-out CV, hence we report results only for the leave-one-out case.

In Figure 2, we provide plots of the volatility estimators. We plot the square root of each estimator in the figures, since this can be interpreted as estimated volatility. We observe that the rolling estimator $\hat{\sigma}_{t,rw}$ generally produces the highest volatility estimates, while our proposed estimator $\hat{\sigma}_t$ produces the lowest; the "no truncation" version of our estimator $\hat{\sigma}_{t,nt}$ produces volatility estimates that lie inbetween $\hat{\sigma}_t$ and $\hat{\sigma}_{t,rw}$ ($\hat{\sigma}_{t,nt}$ coincides with $\hat{\sigma}_t$ for much of the sample period). This observation is in line with our intuition and simulation results regarding the differential behaviour among the estimators. Considering the volatility estimators over different sub-periods of the data, we find that the most obvious differences occur over the predominantly explosive phase of the data leading up to January 2018, during which time there are also a number of relatively smaller rapid expansions and contractions. In this period, the $\hat{\sigma}_{t,rw}$ and $\hat{\sigma}_{t,nt}$ estimates are relatively similar, while the $\hat{\sigma}_t$ estimates are substantially lower. This feature is largely common across the first step window widths considered, and it is not surprising that the greatest differences arise due to the truncation element of $\hat{\sigma}_t$, which has most effect around the times of explosive behaviour in the data. In the earlier period of less dramatic

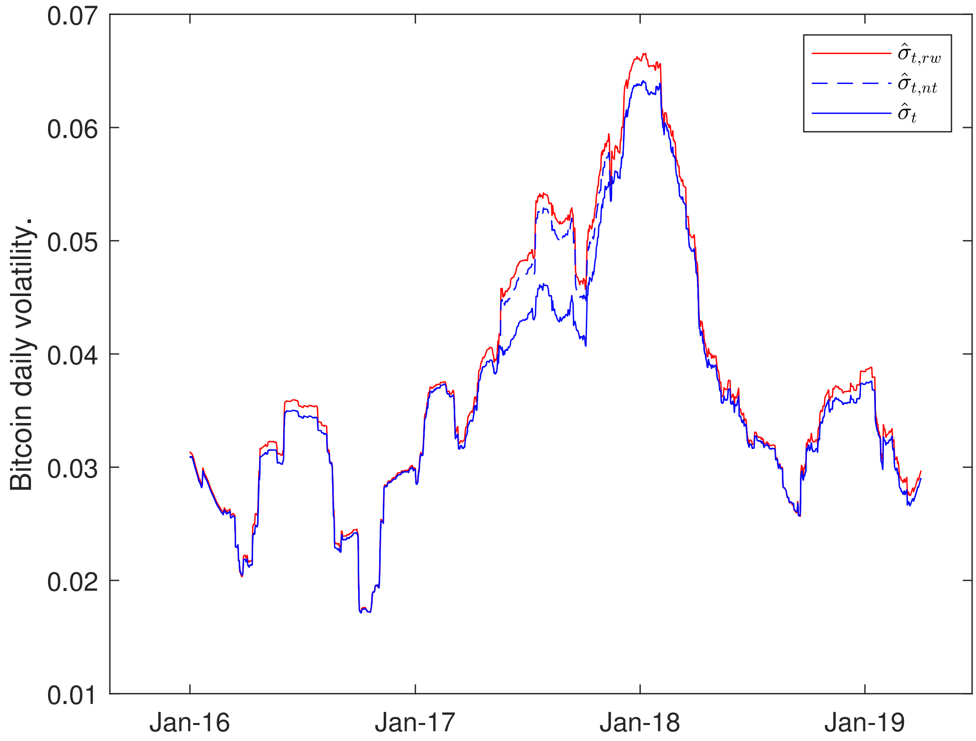

Figure 2: Volatility estimates: $\hat{\sigma}_t^2$, $\hat{\sigma}_{t,nt}^2$ and $\hat{\sigma}_{t,rw}^2$ denote the truncation-based variance estimator, the estimator without truncation, and the classical rolling window estimator, respectively.

price increase (up to mid 2017) and also in the later period following the numerical peak of the series in January 2018, it is interesting to observe that the correction made by truncation is less apparent, with the $\hat{\sigma}_t$ and $\hat{\sigma}_{t,nt}$ estimators producing really quite similar results, while both $\hat{\sigma}_t$ and $\hat{\sigma}_{t,nt}$ produce volatility estimates a little lower than $\hat{\sigma}_{t,rw}$, especially in the cases of the shorter first step bandwidths. In these periods, it seems that the truncation plays less of a role, while the LLS approach involved in $\hat{\sigma}_t$ and $\hat{\sigma}_{t,nt}$ results in reduced volatility estimates compared to the simple rolling approach, suggesting that properly accounting for the AR dynamics can lead to a less inflated (and arguably more reliable) estimator of the true volatility.

6 Conclusion

In this paper we have proposed a new two-step truncation-based variance estimator for time series data containing possibly multiple unit root, explosive and stationary collapse segments at unknown times. The estimator consists of a first step LLS estimation, and a second step smoothing of truncated squared residuals obtained from the first step nonparametric estimation. We have derived uniform consistency results for the estimator in the context of possible explosive (and subsequent stationary collapse) behaviour. Our simulation results confirm the superiority of the new estimator relative to competing methods in terms of accuracy in the presence of explosive/stationary collapse behaviour, as might be observed during an asset price bubble/crash in financial data. In an empirical illustration employing recent Bitcoin data, we also demonstrate the potential value of the estimator in a practical application. While our approach focuses on estimating the variance function while treating the autoregressive dynamics separately, in practice changes in variance and changes in the autoregressive parameter may be interlinked. It would be interesting to entertain the possibility of an extended (and inevitably more complex) model allowing for such endogenous interactions, and we consider this as an avenue for potential future research.

References

Astill, S., Harvey, D.I., Leybourne, S.J., Taylor, A.M.R. and Zu, Y. (2020). CUSUM-based monitoring for explosive financial bubbles in the presence of time-varying volatility. *Journal of Time Series Analysis*, 39, 863–891.

Bai, J. and Perron, P., (1998). Estimating and testing linear models with multiple structural changes. *Econometrica*, 47–78.

Bai, J. and Perron, P., (2003). Computation and analysis of multiple structural change models. *Journal of applied econometrics*, 18, 1–22.

Beare, B.K. (2018). Unit root testing with unstable volatility. *Journal of Time Series Analysis*, 39, 816–835.

Bittlingmayer, G. (1998). Output, stock volatility, and political uncertainty in a natural experiment: Germany, 1880-1940. *Journal of Finance*, 53, 2243–2257.

Blanchard, O. and Simon, J. (2001). The long and large decline in US output volatility. *Brookings Papers on Economic Activity*, 2001, 135–174.

Boswijk, H.P. and Zu, Y. (2018). Adaptive wild bootstrap testing for a unit root with nonstationary volatility. *Econometrics Journal*, 21, 87–113.

Cavaliere, G. (2005). Unit root tests under time-varying variances. *Econometric Reviews*, 23, 259–292.

Cavaliere, G., Nielsen M. and Taylor A.M.R. (2022). Adaptive inference in heteroscedastic fractional time series models. *Journal of Business & Economic Statistics*, 40, 50–65.

Chan, N.H., Yau, C.Y. and Zhang, R.M. (2014). Group LASSO for structural break time series. *Journal of the American Statistical Association*, 109, 590–599.

Chu, C.K. and Marron, J.S. (1991). Comparison of two bandwidth selectors with dependent errors. *The Annals of Statistics*, 19, 1906-1918.

Davis, R.A., Lee, T.C.M. and Rodriguez-Yam, G.A. (2006). Structural break estimation for nonstationary time series models. *Journal of the American Statistical Association*, 101, 223–239.

Fan, J. and Yao, Q. (1998). Efficient estimation of conditional variance functions in stochastic regression. *Biometrika*, 85, 645–660.

Francq, C. and Zakoian, J.M. (2020). Testing the existence of moments for GARCH processes. *Journal of Econometrics*, 227, 47–64.

Francq, C. and Zakoian, J.M. (2021). Testing the existence of moments and estimating the tail index of augmented GARCH processes. Working paper, University of Lille.

Hall, P. and Carroll, R.J. (1989). Variance function estimation in regression: the effect of estimating the mean. *Journal of the Royal Statistical Society, Series B (Methodological)*, 51, 3–14.

Harvey, D. I., Leybourne, S.J. and Zu, Y. (2020). Sign-based unit root tests for explosive financial bubbles in the presence of deterministically time-varying volatility. *Econometric Theory*, 36, 122–169.

Kurozumi, E., Skrobotov, A. and Tsarev, A., (2020). Time-transformed test for the explosive bubbles under non-stationary volatility. Working paper, Russian Presidential Academy of National Economy and Public Administration.

Kristensen, D., (2012). Nonparametric detection and estimation of structural change. *Econometrics Journal*, 15, 420–461.

Li, D., Phillips, P.C.B. and Gao, J. (2016). Uniform consistency of nonstationary kernel-weighted sample covariances for nonparametric regression. *Econometric Theory*, 32, 655–685.

Mumtaz, H. and Theodoridis, K. (2017). Common and country specific economic uncertainty. *Journal of International Economics*, 105, 205–216.

Officer, R.R. (1973). The variability of the market factor of the New York Stock Exchange. *Journal of Business*, 46, 434–453.

Patilea, V. and Raissi, H., (2012). Adaptive estimation of vector autoregressive models with time-varying variance: application to testing linear causality in mean. *Journal of Statistical Planning and Inference*, 142, 2891–2912.

Patilea, V. and Raissi, H., (2014). Testing second-order dynamics for autoregressive processes in presence of time-varying variance. *Journal of the American Statistical Association*, 109, 1099–1111.

de la Peña, V. H. (1999). A general class of exponential inequalities for martingales and ratios. *Annals of Probability*, 27, 537–564.

Phillips, P.C.B., Li, D., and Gao, J. (2017). Estimating smooth structural change in cointegration models. *Journal of Econometrics*, 196, 180–195.

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

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

Qu, Z. and Perron, P., (2007). Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75, 459–502.

Shiryaev, A.N. (1996). *Probability - Graduate Texts in Mathematics vol. 95.* New York: Springer-Verlag.

Xu, K.L. and Phillips, P.C.B. (2008). Adaptive estimation of autoregressive models with time-varying variances. *Journal of Econometrics*, 142, 265–280.

Appendix: Proofs of theorems

Technical lemmas

To begin, we state a number of technical lemmas that will be used in the proofs of the theorems. Proofs of the technical lemmas are contained in the online Supplementary Appendix.

Lemma 1. *Under conditions A1-A3 and B1-B2, if $\rho_1^* > 0$ and $\rho_2^* < 0$,*

$$\begin{aligned}
\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1} |\hat{\rho}_t - \rho_t| &= O_p\left(\sqrt{\frac{\log(T)}{T^2 h_1}}\right) \\
\max_{t_1^*+Th_1+1\leqslant t\leqslant t_2^*-Th_1} |\phi_1^{(t+Th_1-1-t_1^*)}(\hat{\rho}_t - \rho_t)| &= O_p\left(T^{1/p-1/2}\log^2(T)\right) \\
\max_{t_2^*+Th_1+1\leqslant t\leqslant t_3^*-Th_1} |\phi_2^{(t-Th_1-1-t_2^*)}(\hat{\rho}_t - \rho_t)| &= O_p\left(\phi_1^{-(t_2^*-t_1^*)}T^{1/p-1/2}\log^2(T)\right) \\
\max_{t_3^*+Th_1+1\leqslant t\leqslant T-Th_1} |\hat{\rho}_t - \rho_t| &= O_p\left(\sqrt{\frac{\log(T)}{(a_T \vee T^{1/2})^2 Th_1}}\right).
\end{aligned}$$

In Lemma 1, the uniform rate of convergence in the pre-explosive unit root regime is the same as that in Li et al. (2016) and Phillips et al. (2017). The uniform rates derived in the explosive, stationary and the post-stationary unit root regimes are new.

Lemma 2. *Under the same condition as Lemma 1, when $Th_1 + 1 \leqslant t \leqslant t_1^* - Th_1$,*

$$\frac{1}{Th_1}\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \to \gamma$$

which by assumption satisfies $0 < \gamma < \infty$.

When $t_1^* + Th_1 + 1 \leqslant t \leqslant t_2^* - Th_1$, for $\phi_1 > 1$, $a > 0$

$$\frac{1}{\phi_1^{a(t+Th_1-1-t_1^*)}}\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)\phi_1^{a(i-1-t_1^*)} = O(1)$$

and in the limit it is strictly positive and nondegenerate to 0.

When $t_2^* + Th_1 + 1 \leqslant t \leqslant t_3^* - Th_1$, for $1 > \phi_2 > 0$, $a > 0$

$$\frac{1}{\phi_2^{a(t-Th_1-1-t_2^*)}}\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)\phi_2^{a(i-1-t_2^*)} = O(1)$$

and in the limit it is strictly positive and nondegenerate to 0.

When $t_3^* + Th_1 + 1 \leqslant t \leqslant T - Th_1$,

$$\frac{1}{Th_1}\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \to \gamma.$$

Lemma 3. *Uniformly for* $0 \leqslant \tau \leqslant \tau_1^*$,

$$(1/\sqrt{T})y_{[\tau T]} \Rightarrow U(\tau),$$

where $U(s) := \int_0^s \sigma(r)\mathrm{d}W(r)$.

Uniformly for $t_1^* < t \leqslant t_2^*$,

$$T^{-1/2}\phi_1^{-(t-t_1^*)} y_t \xrightarrow{d} U(\tau_1^*).$$

Uniformly for $t_2^* < t \leqslant t_3^*$,

$$T^{-1/2}\phi_1^{-(t_2^*-t_1^*)}\phi_2^{-(t-t_2^*)} y_t \xrightarrow{d} U(\tau_1^*).$$

Uniformly for $\tau_3^* < \tau \leqslant 1$, *for* y^* *defined in A3,*

$$(a_T \vee T^{1/2})^{-1} y_{[\tau T]} \Rightarrow \begin{cases} U(\tau) - U(\tau_3^*) & a_T/T^{1/2} \to 0, \\ y^* + U(\tau) - U(\tau_3^*) & a_T/T^{1/2} \to 1, \\ y^* & a_T/T^{1/2} \to \infty. \end{cases}$$

Lemma 4. *Assume the same conditions as Lemma 1.*

When $Th_1 + 1 \leqslant t \leqslant t_1^* - Th_1$,

$$T^{-2}h_1^{-1}\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1}^2 = O_p(1)$$

and it is also strictly positive and nondegenerate to 0 with probability 1.

When $t_1^* + Th_1 + 1 < t \leqslant t_2^* - Th_1$,

$$T^{-1}\phi_1^{-2(t+Th_1-1-t_1^*)}\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1}^2 = O_p(1)$$

and it is also strictly positive and nondegenerate to 0 with probability 1.

When $t_2^* + Th_1 + 1 < t \leqslant t_3^* - Th_1$,

$$T^{-1}\phi_1^{-2(t_2^*-t_1^*)}\phi_2^{-2(t-Th_1-1-t_2^*)}\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1}^2 = O_p(1)$$

and it is also strictly positive and nondegenerate to 0 with probability 1.

When $t_3^* + Th_1 + 1 < t \leqslant T - Th_1$,

$$(a_T \vee T^{1/2})^{-2} T^{-1} h_1^{-1} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1}^2 = O_p(1)$$

and it is also strictly positive and nondegenerate to 0 with probability 1.

Lemma 5. *Assume the same conditions as Lemma 1.*

When $Th_1 + 1 \leqslant t \leqslant t_1^* - Th_1$ *we have*

$$\max_{Th_1+1 \leqslant t \leqslant t_1^* - Th_1} \left| \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1} u_i \right| = O_p\left(\sqrt{T^2 h_1 \log(T)}\right). \tag{9}$$

When $t_1^* + Th_1 + 1 < t \leqslant t_2^* - Th_1$,

$$\max_{t_1^*+Th_1+1 < t \leqslant t_2^* - Th_1} \left| \phi_1^{-(t+Th_1-1-t_1^*)} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1} u_i \right| = O_p\left(T^{1/p+1/2} \log^2(T)\right). \tag{10}$$

When $t_2^* + Th_1 + 1 < t \leqslant t_3^* - Th_1$,

$$\max_{t_2^*+Th_1+1 < t \leqslant t_3^* - Th_1} \left| \phi_2^{-(t-Th_1-1-t_2^*)} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1} u_i \right| = O_p\left(\phi_1^{(t_2^*-t_1^*)} T^{1/p+1/2} \log^2(T)\right). \tag{11}$$

When $t_3^* + Th_1 + 1 < t \leqslant T - Th_1$,

$$\max_{t_3^*+Th_1+1 < t \leqslant T - Th_1} \left| \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1} u_i \right| = O_p\left(\sqrt{(a_T \vee T^{1/2})^2 Th_1 \log(T)}\right). \tag{12}$$

Lemma 6. *Under the same assumptions as Theorem 2, we have*

$$\max_{1 \leqslant s \leqslant N} \left| \sum_{i=1}^{T} K\left(\frac{i/T - \tau_s}{h_2}\right) (u_i^2 - \sigma_i^2) \right| = O_p\left(\sqrt{Th_2 \log(T)}\right). \tag{13}$$

Proof of Theorem 1

First we give some preliminaries. It proves convenient to characterise the level of the process at the end of the first stationary collapse segment in the form $y_{[\tau_{13}^* T]} = y^* a_T$, where y^* is a random variable bounded in probability and a_T is a deterministic sequence. Since the collapse segment is a mean-reverting stationary AR process, $y_{[\tau_{13}^* T]}$ will either be $O_p(1)$ or divergent; it follows then that, without loss of generality, $a_T = 1$ or $a_T \to \infty$, respectively. The proof of the theorem then proceeds in three main steps. We first prove the theorem in the case where $J = 1$ and $a_T \to \infty$. Next we extend the proof to the case of $J > 1$, now with $y_{[\tau^*_{j3}T]} = y^{j*}a_{j,T}$, $j = 1, .., J$, where the y^{j*} again denote random variables bounded in probability, and where the determinstic sequences $a_{j,T} \to \infty$. Finally, we further extend the proof to the case of $a_{j,T} = 1$.

(i) Proof for Theorem 1 when $J = 1$ and $a_T \to \infty$

Since $J = 1$ in this part of the proof, we suppress the dependence on j in all our notation in this part.

Define $\tilde{\sigma}^2(\tau_s) = \sum_{i=1}^T w_{\tau_s,i}u_i^2$ and $\bar{\sigma}^2(\tau_s) = \sum_{i=1}^T w_{\tau_s,i}\sigma_i^2$. We first make the following decomposition:

$$\hat{\sigma}^2(\tau_s) - \sigma^2(\tau_s) = (\hat{\sigma}^2(\tau_s) - \tilde{\sigma}^2(\tau_s)) + (\tilde{\sigma}^2(\tau_s) - \bar{\sigma}^2(\tau_s)) + (\bar{\sigma}^2(\tau_s) - \sigma^2(\tau_s)) \equiv A + B + C.$$

For A, we make the following decomposition

$$\begin{aligned}
A &= \hat{\sigma}^2(\tau_s) - \tilde{\sigma}^2(\tau_s) \\
&= \sum_{i=1}^{T} w_{\tau_s,i}(\hat{u}_i^2\mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2) \\
&= \left(\sum_{i=1}^{Th_1} + \sum_{i=Th_1+1}^{t_1^*-Th_1} + \sum_{i=t_1^*-Th_1+1}^{t_1^*+Th_1} + \sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} + \sum_{i=t_2^*-Th_1+1}^{t_2^*+Th_1}\right) w_{\tau_s,i}(\hat{u}_i^2\mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2) \\
&\quad + \left(\sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} + \sum_{i=t_3^*-Th_1+1}^{t_3^*+Th_1} + \sum_{i=t_3^*+Th_1+1}^{T-Th_1} + \sum_{i=T-Th_1+1}^{T}\right) w_{\tau_s,i}(\hat{u}_i^2\mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2) \\
&\equiv A_0 + A_1 + A_1' + A_2 + A_2' + A_3 + A_3' + A_4 + A_4' \qquad (14)
\end{aligned}$$

where we decompose the sum into nine terms. In these terms, A_1, A_2, A_3, A_4 correspond to the interiors of the respective unit root, explosive, stationary and unit root segments; A_1', A_2', A_3' correspond to the shrinking neighbourhoods around the change points between segments; while A_0 and A_4' are the two boundaries of the sample.

We first consider A_1'. By construction we have $E(\hat{u}_i^2\mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2) = O(\psi_T^2)$ for all i. Using this, together with Xu and Phillips (2008)'s Lemma A (d) that $\max_{1\leqslant i\leqslant T, 0\leqslant s\leqslant 1} w_{\tau_s,i} =$

$1/(Th_2)$, we have

$$
\begin{aligned}
|A_1'| &= \left| \sum_{i=t_1^*-Th_1+1}^{t_1^*+Th_1} w_{\tau_s,i}(\hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2) \right| \\
&\leqslant \left| \left(\sum_{i=t_1^*-Th_1+1}^{t_1^*+Th_1} w_{\tau_s,i}^2 \right)^{1/2} \left(\sum_{i=t_1^*-Th_1+1}^{t_1^*+Th_1} (\hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2)^2 \right)^{1/2} \right| \\
&\leqslant \left| \left(\max_{s,i} w_{\tau_s,i} \sum_{i=t_1^*-Th_1+1}^{t_1^*+Th_1} w_{\tau_s,i} \right)^{1/2} \left(\sum_{i=t_1^*-Th_1+1}^{t_1^*+Th_1} (\hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2)^2 \right)^{1/2} \right| \\
&= O_p\left(\left(\frac{1}{Th_2} \right)^{1/2} \times (Th_1 \psi_T^4)^{1/2} \right) = O_p\left(\left(\frac{h_1 \psi_T^4}{h_2} \right)^{1/2} \right)
\end{aligned}
$$

where we use the Cauchy-Schwarz inequality in the second step and the order of $\sum_{i=t_1^*-Th_1+1}^{t_1^*+Th_1}(\hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2)^2$ is easily implied from evaluating its expectation. Using the same argument for the neighbourhoods of the change points and the sample boundaries, we have

$$
|A_0 + A_1' + A_2' + A_3' + A_4'| = O_p\left(\left(\frac{h_1 \psi_T^4}{h_2} \right)^{1/2} \right). \tag{15}
$$

With the truncation mechanism, it is seen that the orders of these terms are controlled by ψ_T, which is chosen by the researcher. Without the truncation mechanism, these terms could be much larger and even prevent consistency of the variance estimator.

Next we consider A_1. We have

$$
\begin{aligned}
& \sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}(\hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2) \\
= & \sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}(((\rho_i - \hat{\rho}_i) y_{i-1} + u_i)^2 \mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2) \\
= & \sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}(\rho_i - \hat{\rho}_i)^2 y_{i-1}^2 \mathbb{I}(|\hat{u}_i| < \psi_T) + 2 \sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}(\rho_i - \hat{\rho}_i) y_{i-1} u_i \mathbb{I}(|\hat{u}_i| < \psi_T) \\
& + \sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}(u_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T) - u_i^2) \\
\equiv & A_{11} + A_{12} + A_{13}.
\end{aligned}
$$

For A_{11},

$$
\begin{aligned}
|A_{11}| &\leqslant \left|\sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i)^2 y_{i-1}^2\right| \\
&\leqslant \max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1}(\rho_i-\hat{\rho}_i)^2 \max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1} y_{i-1}^2 \left|\sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}\right| \\
&= O_p\left(\frac{\log(T)}{T^2h_1}\times T\right) = O_p\left(\frac{\log(T)}{Th_1}\right) \qquad (16)
\end{aligned}
$$

using the result of Lemma 1, Lemma 3 and noting that $0<\max_{0\leqslant s\leqslant 1}\sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}\leq 1$. For A_{12}, we have

$$
\begin{aligned}
|A_{12}| &\leqslant \left|\sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i) y_{i-1}u_i\right| \\
&\leqslant \max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1}|\rho_i-\hat{\rho}_i| \max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1} |y_{i-1}| \sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}|u_i| \\
&= O_p\left(\sqrt{\frac{\log(T)}{T^2h_1}\times T}\right) = O_p\left(\sqrt{\frac{\log(T)}{Th_1}}\right), \qquad (17)
\end{aligned}
$$

where $\sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}|u_i| = O_p(1)$ follows from evaluating its expectation. For A_{13},

$$
\begin{aligned}
E|A_{13}| &= \sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i} E(u_i^2\mathbb{I}(|\hat{u}_i|\geqslant\psi_T)) \\
&\leqslant \sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}(Eu_i^4)^{1/2}(P(|\hat{u}_i|\geqslant\psi_T))^{1/2} \\
&\to 0, \qquad (18)
\end{aligned}
$$

because when $Th_1+1\leqslant i\leqslant t_1^*-Th_1$, $\hat{u}_i=\Delta y_i-\hat{\rho}_i y_{i-1}=(\rho_i-\hat{\rho}_i)y_{i-1}+u_i=O_p(1)$ for all i in this range by Lemma 1.

Next we consider A_2. Similar to A_1 we have

$$
\begin{aligned}
A_2 &= \sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i)^2 y_{i-1}^2\mathbb{I}(|\hat{u}_i|<\psi_T) + 2\sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i)y_{i-1}u_i\mathbb{I}(|\hat{u}_i|<\psi_T) \\
&\quad + \sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} w_{\tau_s,i}(u_i^2\mathbb{I}(|\hat{u}_i|<\psi_T)-u_i^2) \\
&\equiv A_{21}+A_{22}+A_{23}.
\end{aligned}
$$

For A_{21},

$$
\begin{aligned}
A_{21} &\leqslant \left|\sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i)^2 y_{i-1}^2\right| \\
&\leqslant T\phi_1^{-2Th_1}\sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} w_{\tau_s,i}|\phi_1^{2(i+Th_1-1-t_1^*)}(\rho_i-\hat{\rho}_i)^2||T^{-1}\phi_1^{-2(i-1-t_1^*)}y_{i-1}^2| \\
&\leqslant T\phi_1^{-2Th_1}\max_{t_1^*+Th_1+1\leqslant i\leqslant t_2^*-Th_1}|\phi_1^{2(i+Th_1-1-t_1^*)}(\rho_i-\hat{\rho}_i)^2| \\
&\quad \max_{t_1^*+Th_1+1\leqslant i\leqslant t_2^*-Th_1}|T^{-1}\phi_1^{-2(i-1-t_1^*)}y_{i-1}^2|\sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} w_{\tau_s,i} \\
&= O_p(\phi_1^{-2Th_1}T^{2/p}\log^4(T)),
\end{aligned}
\tag{19}
$$

where we have used the results of Lemma 3 and Lemma 1. For A_{22},

$$
\begin{aligned}
A_{22} &\leqslant \left|\sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i)y_{i-1}u_i\right| \\
&\leqslant T^{1/2}\phi_1^{-Th_1}\sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} w_{\tau_s,i}|\phi_1^{(i+Th_1-1-t_1^*)}(\rho_i-\hat{\rho}_i)||T^{-1/2}\phi_1^{-(i-1-t_1^*)}y_{i-1}||u_i| \\
&\leqslant T^{1/2}\phi_1^{-Th_1}\max_{t_1^*+Th_1+1\leqslant i\leqslant t_2^*-Th_1}|\phi_1^{(i+Th_1-1-t_1^*)}(\rho_i-\hat{\rho}_i)||T^{-1/2}\phi_1^{-(i-1-t_1^*)}y_{i-1}|\sum_{i=t_1^*+Th_1+1}^{t_2^*-Th_1} w_{\tau_s,i}|u_i| \\
&= O_p(\phi_1^{-Th_1}T^{1/p}\log^2(T)).
\end{aligned}
\tag{20}
$$

Using the same argument as in deriving (18), we have $A_{23}=o_p(1)$.

For A_3 we again consider the dominant terms

$$
\begin{aligned}
A_3 &= \sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i)^2 y_{i-1}^2\mathbb{I}(|\hat{u}_i|<\psi_T)+2\sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i)y_{i-1}u_i\mathbb{I}(|\hat{u}_i|<\psi_T) \\
&\quad +\sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} w_{\tau_s,i}(u_i^2\mathbb{I}(|\hat{u}_i|<\psi_T)-u_i^2) \\
&\equiv A_{31}+A_{32}+A_{33}.
\end{aligned}
$$

A_{31} can be analyzed as follows

$$\begin{aligned}
A_{31} &\leqslant \left|\sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i)^2 y_{i-1}^2\right| \\
&\leqslant \phi_2^{2Th_1-2} T \phi_1^{2(t_2^*-t_1^*)} \\
&\quad \sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} w_{\tau_s,i}\phi_2^{2(i-Th_1-t_2^*)}(\rho_i-\hat{\rho}_i)^2 |T^{-1}\phi_1^{-2(t_2^*-t_1^*)}\phi_2^{-2(i-1-t_2^*)} y_{i-1}^2| \\
&\leqslant \phi_2^{2Th_1-2} T \phi_1^{2(t_2^*-t_1^*)} \max_{t_2^*+Th_1+1\leqslant i\leqslant t_3^*-Th_1} \phi_2^{2(i-Th_1-t_2^*)}(\rho_i-\hat{\rho}_i)^2 \\
&\quad \max_{t_2^*+Th_1+1\leqslant i\leqslant t_3^*-Th_1} |T^{-1}\phi_1^{-2(t_2^*-t_1^*)}\phi_2^{-2(i-1-t_2^*)} y_{i-1}^2| \sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} w_{\tau_s,i} \\
&= O_p(\phi_2^{2Th_1} T^{2/p}\log^4(T)) \qquad (21)
\end{aligned}$$

where we have used the results of Lemma 3 and Lemma 1. For A_{32},

$$\begin{aligned}
A_{32} &\leqslant \left|\sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} w_{\tau_s,i}(\rho_i-\hat{\rho}_i) y_{i-1}u_i\right| \\
&\leqslant \phi_2^{Th_1-1} T^{1/2} \phi_1^{(t_2^*-t_1^*)} \\
&\quad \sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} w_{\tau_s,i}\phi_2^{(i-Th_1-t_2^*)}|\rho_i-\hat{\rho}_i| |T^{-1/2}\phi_1^{-(t_2^*-t_1^*)}\phi_2^{-(i-1-t_2^*)} y_{i-1}||u_i| \\
&\leqslant \phi_2^{Th_1-1} T^{1/2} \phi_1^{(t_2^*-t_1^*)} \max_{t_2^*+Th_1+1\leqslant i\leqslant t_3^*-Th_1 i} \phi_2^{(i-Th_1-t_2^*)}|\rho_i-\hat{\rho}_i| \\
&\quad \max_{t_2^*+Th_1+1\leqslant i\leqslant t_3^*-Th_1} |T^{-1/2}\phi_1^{-(t_2^*-t_1^*)}\phi_2^{-(i-1-t_2^*)} y_{i-1}| \sum_{i=t_2^*+Th_1+1}^{t_3^*-Th_1} w_{\tau_s,i}|u_i| \\
&= O_p(\phi_2^{Th_1} T^{1/p}\log^2(T)). \qquad (22)
\end{aligned}$$

Using the same argument as in deriving (18), we have $A_{33} = o_p(1)$.

A_4 can be decomposed in the same way as A_1 and each term will respectively have the same order as A_{11}, A_{12} and A_{13}. Therefore, collecting the results of (16), (17), (18) (and also the derived order for A_{23} and A_{33}), (19), (20), (21) and (22), we have

$$\begin{aligned}
A &= O_p\left(\frac{\log(T)}{Th_1}\right) + O_p\left(\sqrt{\frac{\log(T)}{Th_1}}\right) + o_p(1) \\
&\quad + O_p(T\phi_1^{-2Th_1} T^{2/p-1}\log^4(T)) + O_p(T^{1/2}\phi_1^{-Th_1} T^{1/p-1/2}\log^2(T)) \\
&\quad + O_p(\phi_2^{2Th_1} T^{2/p}\log^4(T)) + O_p(\phi_2^{Th_1} T^{1/p}\log^2(T)).
\end{aligned}$$

Under Assumptions B2 and C3, $A = o_p(1)$. Specifically, we note that because $\log(T)$ and T are separately dominated by some polynomial rate of Th_1, the above terms involving powers of

Th_1, i.e. the terms $\phi_1^{-2Th_1}$, $\phi_1^{-Th_1}$, $\phi_2^{2Th_1}$ and $\phi_2^{Th_1}$ will converge to 0 at a speed faster than any combined polynomial diverging rate of $\log(T)$ and T, hence the last four terms are $o_p(1)$.

For the term $B = \tilde{\sigma}^2(\tau_s) - \bar{\sigma}^2(\tau_s)$, notice that $\bar{\sigma}^2(\tau_s) - \tilde{\sigma}^2(\tau_s) = \sum_{i=1}^T w_{\tau_s,i}(u_i^2 - \sigma_i^2)$ and $\{u_i^2 - \sigma_i^2\}$ is a m.d.s. indexed by i. Using Burkholder's inequality and then the Markov inequality, it can be shown that $\frac{1}{Th_2}\sum_{i=1}^T K\left(\frac{i/T-\tau_s}{h_2}\right)(u_i^2 - \sigma_i^2) = O_p\left(\sqrt{1/(Th_2)}\right)$. Also since that $0 < \frac{1}{Th_2}\sum_{i=1}^T K\left(\frac{i/T-\tau_s}{h_2}\right) < \infty$ we find $|\bar{\sigma}^2(\tau_s) - \tilde{\sigma}^2(\tau_s)| = O_p\left(\sqrt{1/(Th_2)}\right)$ under assumption C3.

For the term $C = \bar{\sigma}^2(\tau_s) - \sigma^2(\tau_s)$, we notice that

$$|C| = \left| \frac{\frac{1}{Th_2}\sum_{i=1}^T K(\frac{i/T-\tau_s}{h_2})(\sigma^2(i/T) - \sigma^2(\tau_s))}{\frac{1}{Th_2}\sum_{i=1}^T K(\frac{i/T-\tau_s}{h_2})} \right|.$$

The numerator satisfies

$$\begin{aligned}
& \left| \frac{1}{Th_2}\sum_{i=1}^T K(\frac{i/T-\tau_s}{h_2})(\sigma^2(i/T) - \sigma^2(\tau_s)) \right| \\
= & \left| \int_0^1 \frac{1}{h_2} K\left(\frac{u-\tau_s}{h_2}\right)(\sigma^2(u) - \sigma^2(\tau_s))\mathrm{d}u(1 + o(1/T)) \right| \\
= & \left| \int_{\frac{-\tau_s}{h_2}}^{\frac{1-\tau_s}{h_2}} K(z)(\sigma^2(\tau_s + h_2 z) - \sigma^2(\tau_s))\mathrm{d}z(1 + o(1)) \right| \\
\leqslant & \ L^\beta h_2 \int_{\frac{-\tau_s}{h_2}}^{\frac{1-\tau_s}{h_2}} |K(z)z|\mathrm{d}z(1 + o(1))
\end{aligned}$$

by the Lipschitz continuity of the $\sigma^2(.)$ function. Notice now the integral $\int_{\frac{-\tau_s}{h_2}}^{\frac{1-\tau_s}{h_2}} |K(z)z|\mathrm{d}z \leqslant \int_{-\infty}^{\infty} |K(z)z|\mathrm{d}z < \infty$, by Assumption C1. The numerator thus has order $O(h_2)$. The denominator clearly satisfies $0 < \frac{1}{Th_2}\sum_{i=1}^T K(\frac{i/T-\tau_s}{h_2}) < \infty$ for all $1 \leqslant s \leqslant N$. Thus we have $C = O(h_2)$, which is clearly $o(1)$ under the assumption that $h_2 \to 0$.

In summary, we have shown that $\max_s |\hat{\sigma}^2(\tau_s) - \sigma^2(\tau_s)| = o_p(1)$ and our proof is complete.

(ii) Extending to $J > 1$, $a_{j,T} \to \infty$

When $J > 1$, later episodes will be similar to the first episode, with the only difference being that the initial values of later episodes will be the end of the previous episode's stationary collapse segment. As in the single episode case, for episode j, write the level of the process at the end of a stationary collapse segment in the form $y_{[\tau_{j,3}^* T]} = y^{j,*}a_{j,T}$, where $y^{j,*}$ is a random variable bounded in probability and $a_{j,T}$ is a deterministic sequence. Then for any episode $j > 1$, the initial value of episode j is essentially $y_{[\tau_{j-1,3}^* T]} = y^{j-1,*}a_{j-1,T}$. The analysis of any episode j can then be conducted in exactly the same way as before.

Under the same condition as Lemma 1, when $j > 1$, if $\rho_{j,1}^* > 0$ and $\rho_{j,2}^* < 0$, it can be shown that

$$\max_{t_{j-1,3}^*+Th_1+1\leqslant t\leqslant t_{j,1}^*-Th_1} |\hat{\rho}_{j,t}-\rho_{j,t}| = O_p\left(\sqrt{\frac{\log(T)}{(a_{j-1,T}\vee T^{1/2})^2 Th_1}}\right)$$

$$\max_{t_{j,1}^*+Th_1+1\leqslant t\leqslant t_{j,2}^*-Th_1} |\phi_{j,1}^{(t+Th_1-1-t_{j,1}^*)}(\hat{\rho}_{j,t}-\rho_{j,t})| = O_p(T^{1/p}(a_{j-1,T}\vee T^{1/2})^{-1}\log^2(T))$$

$$\max_{t_{j,2}^*+Th_1+1\leqslant t\leqslant t_{j,3}^*-Th_1} |\phi_{j,2}^{(t-Th_1-1-t_{j,2}^*)}(\hat{\rho}_{j,t}-\rho_{j,t})| = O_p(\phi_1^{-(t_2^*-t_1^*)}T^{1/p}(a_{j-1,T}\vee T^{1/2})^{-1}\log^2(T)).$$

Notice that no more conditions than those required for Lemma 1 are needed for the above results in the general episode j case. This is because larger initial values actually make the estimation problem easier (in the sense of a higher rate of convergence), so there is no need to impose additional conditions.

(iii) Extending to $J>1$ and $a_{j,T}=1$

In the above, we prove our results when $a_T\to\infty$. From the proof, it can be seen that when $a_{j-1,T}=1$, the rate of convergence of the LLS estimator in the subsequent unit root segment ($O_p\left(\sqrt{\frac{\log(T)}{(a_{j-1,T}\vee T^{1/2})^2 Th_1}}\right)$) is still valid. However, $a_{j-1,T}=1$ makes analysis of the rate of convergence for the LLS estimator in the stationary collapse regimes more complicated. For episode $j>1$, when $a_{j,T}=1$, the process will have already collapsed to an $O_p(1)$ level at some time in the preceding stationary collapse segment, i.e. for some t satisfying $t_{j,2}^*+1\leqslant t\leqslant t_{j,3}^*$. Denote this point in time as t_j^{mr}, where the superscript signifies mean reversion. It can be anticipated that the rate of convergence of the LLS estimator will be different before, after and around the time of t_j^{mr}. When the LLS estimator uses all the observations before t_j^{mr}, its rate of convergence will be the same as that derived in the previous section for the stationary collapse segment. When the LLS estimator uses all the observations after t_j^{mr} but before $t_{j,3}$, then the same proof strategy of deriving the uniform rate can be applied and it can be shown that $\max_{t_{j,1}^*+Th_1+1\leqslant t\leqslant t_{j,2}^*-Th_1}|\hat{\rho}_{j,t}-\rho_{j,t}|=O_p\left(\sqrt{\log(T)/(Th_1)}\right)$ and the associated residuals will still deliver a consistent variance function estimator. When the LLS estimator uses observations from both sides of t_j^{mr}, the rate of convergence of the LLS estimator will be a complicated quantity; however, such a rate does not require an explicit derivation because the length of the interval involved has order Th_1. As in our analysis for the term A_1', the effect of this interval on the final estimator will vanish under our bandwidth and truncation parameter assumptions. Therefore, our result also holds for the case $a_{j,T}=1$.

Proof of Theorem 2

As in the proof of Theorem 1, we again proceed using the same three steps. We first prove Theorem 2 when $J = 1$ and $a_T \to \infty$.

Again in this part of the proof, we suppress the dependence on j in all our notation. In what follows, $\max_t$ without specifying the range of t means the maximum is taken over $1 \leqslant t \leqslant T$; a $\max_s$ without specifying the range of s means the maximum is taken over $1 \leqslant s \leqslant N$. N will be defined later in the proof. Let $\mathbb{C}$ denote a generic positive number, the value of which changes with the context where it is employed. In a summation of the form $\sum_{i=Th_1+1}^{T-Th_1}$, Th_1 may not be an integer, thus the above summation is a short-hand notation for the summation over the integers in the range $[Th_1+1, T-Th_1]$.

First partition the interval $[0,1]$ into N equilength subintervals $\mathbf{I}_s = [(s-1)/N, s/N]$ for $s = 1, \ldots, N$. Let τ_s be the center of $\mathbf{I}_s$. Clearly the intervals cover $[0,1]$. We then have the following decomposition:

$$
\begin{aligned}
& \sup_{\tau\in[0,1]} |\hat{\sigma}^2(\tau) - \sigma^2(\tau)| \\
= \; & \max_{1\leqslant s\leqslant N} \sup_{\tau\in\mathbf{I}_s} |\hat{\sigma}^2(\tau) - \sigma^2(\tau)| \\
= \; & \max_{1\leqslant s\leqslant N} \sup_{\tau\in\mathbf{I}_s} |\hat{\sigma}^2(\tau) - \hat{\sigma}^2(\tau_s) + \hat{\sigma}^2(\tau_s) - \sigma^2(\tau_s) + \sigma^2(\tau_s) - \sigma^2(\tau)| \\
\leqslant \; & \max_{1\leqslant s\leqslant N} \sup_{\tau\in\mathbf{I}_s} |\hat{\sigma}^2(\tau) - \hat{\sigma}^2(\tau_s)| + \max_{1\leqslant s\leqslant N} |\hat{\sigma}^2(\tau_s) - \sigma^2(\tau_s)| + \max_{1\leqslant s\leqslant N} \sup_{\tau\in\mathbf{I}_s} |\sigma^2(\tau_s) - \sigma^2(\tau)|.
\end{aligned}
$$

First consider $\hat{\sigma}^2(\tau) - \hat{\sigma}^2(\tau_s)$. By definition

$$
\begin{aligned}
& \hat{\sigma}^2(\tau) - \hat{\sigma}^2(\tau_s) \\
= \; & \frac{\sum_{i=1}^T K_{h_2}(i/T - \tau)\,\hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T)}{\sum_{i=1}^T K_{h_2}(i/T - \tau)} - \frac{\sum_{i=1}^T K_{h_2}(i/T - \tau_s)\,\hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T)}{\sum_{i=1}^T K_{h_2}(i/T - \tau_s)} \\
= \; & \frac{\sum_{i=1}^T \left(K_{h_2}(i/T - \tau) - K_{h_2}(i/T - \tau_s)\right)\hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T)}{\sum_{i=1}^T K_{h_2}(i/T - \tau)} \\
& - \hat{\sigma}^2(\tau_s) \frac{\sum_{i=1}^T \left(K_{h_2}(i/T - \tau) - K_{h_2}(i/T - \tau_s)\right)}{\sum_{i=1}^T K_{h_2}(i/T - \tau)} \\
\leqslant \; & \frac{\sum_{i=1}^T \left| K\left(\frac{i/T-\tau}{h_2}\right) - K\left(\frac{i/T-\tau_s}{h_2}\right)\right| \hat{u}_i^2 \mathbb{I}(|\hat{u}_i| < \psi_T)}{\sum_{i=1}^T K(\frac{i/T-\tau}{h_2})} \\
& + \hat{\sigma}^2(\tau_s) \frac{\sum_{i=1}^T \left| K\left(\frac{i/T-\tau}{h_2}\right) - K\left(\frac{i/T-\tau_s}{h_2}\right)\right|}{\sum_{i=1}^T K\left(\frac{i/T-\tau}{h_2}\right)}.
\end{aligned}
$$

Using the Lipschitz continuity assumption of $K(.)$ over $\mathbb{R}$, we have

$$\begin{aligned}
& \max_{1\leqslant s\leqslant N}\sup_{\tau\in\mathbf{I}_s}|\hat{\sigma}^2(\tau)-\hat{\sigma}^2(\tau_s)| \\
\leqslant\ & \frac{\mathbb{C}\sum_{i=1}^T \hat{u}_i^2\mathbb{I}(|\hat{u}_i|<\psi_T)}{Nh_2\sum_{i=1}^T K(\frac{i/T-\tau}{h_2})}+\hat{\sigma}^2(\tau_s)\frac{\mathbb{C}\sum_{i=1}^T 1}{Nh_2\sum_{i=1}^T K\left(\frac{i/T-\tau}{h_2}\right)} \\
=\ & O_p\left(\frac{\psi_T^2}{Nh_2^2}\right)+O_p\left(\frac{1}{Nh_2^2}\right)=O_p\left(\frac{\psi_T^2}{Nh_2^2}\right)
\end{aligned}$$

where we have used $\max_\tau \hat{\sigma}^2(\tau)=O_p(1)$ and that $(1/T)\sum_{i=2}^T K_{h_2}(i/T-\tau)=\int_{-\infty}^{\infty}K(u)du+o(1)>0$; $(1/T)\sum_{i=1}^T \hat{u}_i^2\mathbb{I}(|\hat{u}_i|<\psi_T)=O_p(\psi_T^2)$ follows from evaluating its expectation. Now taking $N=(Th_2)^2$, the above becomes $O_p(\psi_T^2/(T^2h_2^4))$. Then we consider the third term $\max_{1\leqslant s\leqslant N}\sup_{\tau\in\mathbf{I}_s}|\sigma^2(\tau_s)-\sigma^2(\tau)|$. By the Lipschitz continuity of $\sigma(.)$,

$$\max_{1\leqslant s\leqslant N}\sup_{\tau\in\mathbf{I}_s}|\sigma^2(\tau_s)-\sigma^2(\tau)|\leq \mathbb{C}\left(\frac{1}{N}\right)$$

which is clearly $o(1)$. Then to show the result of the theorem, it remains to show that the second term satisfies $\max_{1\leqslant s\leqslant N}|\hat{\sigma}^2(\tau_s)-\sigma^2(\tau_s)|=o_p(1)$, which we do in the following.

Using the notations $\tilde{\sigma}^2(\tau_s)$, $\bar{\sigma}^2(\tau_s)$ and A, B and C defined in the proof of Theorem 1, we have the decomposition:

$$\max_{1\leqslant s\leqslant N}|\hat{\sigma}^2(\tau_s)-\sigma^2(\tau_s)|\leqslant \max_s|A|+\max_s|B|+\max_s|C|.$$

The proof proceeds by showing that each of the three terms are $o_p(1)$. For A, we can make the same decomposition as in (14). With reference to the derivation for the pointwise rate for A_1' in the proof of Theorem 1, it is seen that the derived rate is also uniform in s. This is also true for other boundary and break time neighbourhood terms, hence we have

$$\max_s|A_0+A_1'+A_2'+A_3'+A_4'|=o_p(1).$$

Considering the pointwise rates derived in Theorem 1 for the other sub terms of A, we note that the rates for the terms A_{11}, A_{13}, A_{21}, A_{23}, A_{31}, A_{33} are already uniform in s. The pointwise rates for the terms A_{12}, A_{22} and A_{32} are not uniform because the order of the terms such as $\max_s\sum w_{\tau_s,i}|u_i|$ cannot be calculated by evaluating their expectations as in Theorem 1. We compute the order of these terms in the following.

For A_{12}, we have

$$\begin{aligned}
&\max_s |A_{12}| \\
\leqslant & \max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1} |\rho_i - \hat{\rho}_i| \max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1} |y_{i-1}| \max_s \sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}|u_i| \\
\leqslant & \max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1} |\rho_i - \hat{\rho}_i| \max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1} |y_{i-1}| \max_s \left(\sum_{i=Th_1+1}^{t_1^*-Th_1} w_{\tau_s,i}^2\right)^{1/2} \left(\sum_{i=Th_1+1}^{t_1^*-Th_1} u_i^2\right)^{1/2} \\
= & O_p\left(\sqrt{\frac{\log(T)}{T^2 h_1} \times T \times \frac{1}{Th_2} \times T}\right) = O_p\left(\sqrt{\frac{\log(T)}{Th_1 h_2}}\right)
\end{aligned}$$

which is $o_p(1)$ under our bandwidth assumption C3.

Similarly, we can show

$$\max_s |A_{22}| = O_p(\phi_1^{-Th_1} T^{1/p} \log^2(T) h_2^{-1/2})$$

and

$$\max_s |A_{32}| = O_p(\phi_2^{Th_1} T^{1/p} \log^2(T) h_2^{-1/2})$$

which are both $o_p(1)$ in view of Assumption C3.

The term A_4 can be analyzed in the same way as A_1 and we have $\max_s |A_4| = o_p(1)$. In summary, we have shown $\max_{1\leqslant s\leqslant N} |A| = o_p(1)$.

For the term $\max_s |B|$, notice that $\bar{\sigma}^2(\tau_s) - \tilde{\sigma}^2(\tau_s) = \sum_{i=1}^T w_{\tau_s,i}(u_i^2 - \sigma_i^2)$ and $\{u_i^2 - \sigma_i^2\}$ is a m.d.s. indexed by i. By Lemma 6 and the fact that $0 < \frac{1}{Th_2}\sum_{i=1}^T K\left(\frac{i/T-\tau_s}{h_2}\right) < \infty$ we find

$$\max_s |B| = O_p\left(\sqrt{\frac{\log(T)}{Th_2}}\right) = o_p(1)$$

under assumption C3.

For the term $\max_s |C|$, we note that the pointwise rate derived in Theorem 1 is also uniform in s. Thus we have $\max_s |C| = O(h_2^\beta)$, which is clearly $o(1)$ under the assumption that $h_2 \to 0$.

In summary, we have shown that $\max_s |\hat{\sigma}^2(\tau_s) - \sigma^2(\tau_s)| = o_p(1)$ and our proof is complete for the case $J = 1$ and $a_{j,T} \to \infty$.

The proof can be extended to the case $J > 1$, $a_{j,T} \to \infty$ by exactly the same argument used in the proof of Theorem 1. To extend the proof to the case $J > 1$, $a_{j,T} = 1$, note again that the effect of the length Th_1 neighbourhood around the mean-reverting time point t_j^{mr} on the uniform behaviour of the variance estimator can also be studied in the same way as the effect of the A_1' term; we have already shown above that $\max_s |A_1'| = o_p(1)$.

Supplementary Appendix to

"Estimation of the variance function in structural break autoregressive models with nonstationary and explosive segments"

David I. Harvey, Stephen J. Leybourne and Yang Zu
School of Economics, University of Nottingham

Proofs of technical lemmas

Proof of Lemma 1

First notice that since the G function is only non-zero over $[-1,1]$, in general for $Th_1+1 \leqslant t \leqslant T-Th_1$ we have,

$$\begin{aligned}\hat{\rho}_t &= \left(\sum_{i=1}^{T} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}^2\right)^{-1}\left(\sum_{i=1}^{T} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}\Delta y_i\right) \\ &= \left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}^2\right)^{-1}\left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}^2\rho_i\right) \\ &\quad + \left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}^2\right)^{-1}\left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}u_i\right).\end{aligned}$$

Then notice that the $\rho(.)$ function is piecewise constant in each of the four regimes. For example, when $Th_1+1 \leqslant t \leqslant t_1^*-Th_1$, the indices in the range $t-Th_1 \leqslant i \leqslant t+Th_1$ will satisfy $1 \leqslant i \leqslant t_1^*$ and live in the first regime, where by assumption $\rho_i = 0$ always holds. Thus the above becomes

$$\hat{\rho}_t = 0 + \left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}^2\right)^{-1}\left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}u_i\right).$$

In the same way, when $t_1^*+Th_1+1 < t \leqslant t_2^*-Th_1$, the above estimator will only use data from the second (explosive) regime, where $\rho_i = \rho_1^*$ always holds. Thus the above becomes

$$\hat{\rho}_t = \rho_1^* + \left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}^2\right)^{-1}\left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}u_i\right).$$

We also easily have when $t_2^*+Th_1+1 < t \leqslant t_3^*-Th_1$,

$$\hat{\rho}_t = \rho_2^* + \left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}^2\right)^{-1}\left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right) y_{i-1}u_i\right)$$

and when $t_3^* + Th_1 + 1 < t \leqslant T - Th_1$,

$$\hat{\rho}_t = 0 + \left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1} \left(\frac{i-t}{T} \right) y_{i-1}^2 \right)^{-1} \left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1} \left(\frac{i-t}{T} \right) y_{i-1} u_i \right).$$

Using the definition of the $\rho(.)$ function, the above results in the four regimes can be written compactly as

$$\hat{\rho}_t - \rho_t = \left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1} \left(\frac{i-t}{T} \right) y_{i-1}^2 \right)^{-1} \left(\sum_{i=t-Th_1}^{t+Th_1} G_{h_1} \left(\frac{i-t}{T} \right) y_{i-1} u_i \right)$$

in all the four considered "interior" ranges of this theorem. With this representation, the claimed results of the theorem follow straightforwardly from the results for the corresponding regimes in Lemma 4 and Lemma 5.

Proof of Lemma 2

When $Th_1 + 1 \leqslant t \leqslant t_1^* - Th_1$, the claimed result can be proved by standard argument such as that in proof of Lemma A (c) of Xu and Phillips (2008).

When $t_1^* + Th_1 + 1 \leqslant t \leqslant t_2^* - Th_1$,

$$\frac{1}{\phi_1^{a(t+Th_1-1-t_1^*)}} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \phi_1^{a(i-1-t_1^*)} = \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \phi_1^{a(i-t-Th_1)}.$$

By the positivity of the kernel function, $\phi_1 > 1$, $a > 0$ the above is clearly strictly positive and nondegenerate to 0. On the other hand, since $G(.)$ is bounded, we have

$$\begin{aligned} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \phi_1^{a(i-t-Th_1)} &\leqslant \mathbb{C} \sum_{i=t-Th_1}^{t+Th_1} \phi_1^{a(i-t-Th_1)} \\ &= O(1) \end{aligned}$$

where $\sum_{i=t-Th_1}^{t+Th_1} \phi_1^{a(i-t-Th_1)} < \infty$ because $i-t-Th_1 \leqslant 0$ for all i in the range of the summation.

When $t_2^* + Th_1 + 1 \leqslant t \leqslant t_3^* - Th_1$,

$$\begin{aligned} \frac{1}{\phi_2^{a(t-Th_1-1-t_2^*)}} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \phi_2^{a(i-1-t_2^*)} &= \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \phi_2^{a(i-t+Th_1)} \\ &\leqslant \mathbb{C} \sum_{i=t-Th_1}^{t+Th_1} \phi_2^{a(i-t+Th_1)} \\ &= O(1) \end{aligned}$$

by noticing that G is bounded, $i - t + Th_1 > 0$ for all i in the above range of summation and $1 > \phi_2 > 0$. It is also nondegenerate to 0 by the same argument as used for the previous regime.

When $t_3^* + Th_1 + 1 \leqslant t \leqslant T - Th_1$, the claim of the lemma can be shown using the same argument for the regime $1 \leqslant t \leqslant t_1^* - Th_1$.

Proof of Lemma 3

When $0 \leqslant \tau \leqslant \tau_1^*$, that is when $1 \leqslant t \leqslant t_1^*$, since $(1/\sqrt{T})y_{t-1} = (1/\sqrt{T})\sum_{i=2}^{t-1}\sigma_i\varepsilon_i + o_p(1)$, the claimed result follows easily from, e.g. Lemma 1 of Cavaliere (2005).

When $t_1^* < t \leqslant t_2^*$, by repeated backward substitution we have

$$y_t = u_t + \phi_1 u_{t-1} + \ldots + \phi_1^{t-t_1^*-1}u_{t_1^*+1} + \phi_1^{t-t_1^*}y_{t_1^*}$$

where $y_{t_1^*}$ is the last observation in the unit root regime (and also serves as the initial value for the explosive regime). Defining

$$\begin{aligned} A_t &= \phi_1^{t-t_1^*}y_{t_1^*}, \\ B_t &= u_t + \phi_1 u_{t-1} + \ldots + \phi_1^{t-t_1^*-1}u_{t_1^*+1} \end{aligned}$$

and we have $y_t = A_t + B_t$. For the term B_t, notice that it is a martingale, by Burkholder's inequality (cf. Shiryaev (1996, p.499)),

$$\begin{aligned} E\left|B_t\right|^2 &= E\left|u_t + \phi_1 u_{t-1} + \ldots + \phi_1^{t-t_1^*-1}u_{t_1^*+1}\right|^2 \\ &\sim E(u_t^2 + \phi_1^2 u_{t-1}^2 + \ldots + \phi_1^{2(t-t_1^*-1)}u_{t_1^*+1}^2) \\ &\leqslant \mathbb{C}\frac{\phi_1^{2(t-t_1^*)}-1}{\phi_1^2-1} = O(\phi_1^{2(t-t_1^*)}). \end{aligned}$$

Thus it follows easily that $\phi_1^{-(t-t_1^*)}B_t = O_p(1)$ for any $t_1^* < t \leqslant t_2^*$. Furthermore, notice that by Doob's maximal inequality for martingales, we also easily have $\max_{t_1^*<t\leqslant t_2^*}\phi_1^{-(t-t_1^*)}|B_t| = O_p(1)$. For the term A_t:

$$T^{-1/2}\phi_1^{-(t-t_1^*)}A_t = T^{-1/2}y_{t_1^*} \xrightarrow{d} U(\tau_1^*),$$

noticing that t_1^* is in the unit root regime. In total, we have uniformly for $t_1^* < t \leqslant t_2^*$,

$$T^{-1/2}\phi_1^{-(t-t_1^*)}y_t = T^{-1/2}\phi_1^{-(t-t_1^*)}(A_t + B_t) \xrightarrow{d} U(\tau_1^*).$$

When $t_2^* < t \leqslant t_3^*$, by repeated backward substitution we have

$$y_t = \left(u_t + \phi_2 u_{t-1} + \ldots + \phi_2^{t-t_2^*-1}u_{t_2^*+1}\right) + \phi_2^{t-t_2^*}y_{t_2^*} := B_t' + A_t',$$

where B_t' and A_t' are defined implicitly. Since $0 < \phi_2 < 1$, we easily have $B_t' = O_p(1)$ for any $t_2^* < t \leqslant t_3^*$. Clearly B_t' is a martingale, again by Doob's maximal inequality for martingales, we also have $\max_{t_2^*<t\leqslant t_3^*}|B_t'| = O_p(1)$. Now by assumption that the process crashes into $y_t^* a_T$ with $a_T \to \infty$ at the end of this regime, the only possible term having this order is A_t'. This also implies that A_t' is the dominating term and $y_t = A_t'(1+o_p(1))$. Then using the weak convergence result from the previous explosive regime that $T^{-1/2}\phi_1^{-(t_2^*-t_1^*)}y_{t_2^*} \xrightarrow{d} U(\tau_1^*)$, we have uniformly for $t_2^* < t \leqslant t_3^*$,

$$T^{-1/2}\phi_1^{-(t_2^*-t_1^*)}\phi_2^{-(t-t_2^*)}y_t = T^{-1/2}\phi_1^{-(t_2^*-t_1^*)}y_{t_2^*}(1+o_p(1)) \xrightarrow{d} U(\tau_1^*).$$

When $\tau_3^* < \tau \leqslant 1$, that is $t_3^* < t = [\tau T] \leqslant T$, again by repeated backward substitution

$$y_t = y_{t_3^*} + u_{t_3^*+1} + u_{t_3^*+2} + \ldots + u_t.$$

The process is a random walk process starting with a term $y_{t_3^*} = O_p(a_T)$, while $\sum_{i=t_3^*+1}^{t} u_i$ is a sum of m.d.s., which in total has order $O_p(T^{1/2})$. Again by Doob's inequality, this rate is also uniform for $t_3^* < t \leqslant T$. Thus the asymptotic behavior of y_t in this regime depends on the relative size of the initial value $y_{t_3^*}$ and $\sum_{i=t_3^*+1}^{t} u_i$. If $a_T/T^{1/2} \to 1$, the two parts have the same order and

$$\frac{1}{\sqrt{T}} y_{[\tau T]} = \frac{1}{\sqrt{T}} y_{t_3^*} + \frac{1}{\sqrt{T}} (u_{t_3^*+1} + u_{t_3^*+2} + \ldots + u_{[\tau T]}) \Rightarrow y^* + (U(\tau) - U(\tau_3^*)).$$

If $a_T/T^{1/2} \to 0$, then $\sum_{i=t_3^*+1}^{t} u_i$ dominates and

$$\frac{1}{\sqrt{T}} y_{[\tau T]} = \frac{1}{\sqrt{T}} y_{t_3^*} + \frac{1}{\sqrt{T}} (u_{t_3^*+1} + u_{t_3^*+2} + \ldots + u_{[\tau T]}) \Rightarrow (U(\tau) - U(\tau_3^*)).$$

If $a_T/T^{1/2} \to \infty$, then the initial value part dominates and

$$\frac{1}{a_T} y_{[\tau T]} = \frac{1}{a_T} y_{t_3^*} + \frac{1}{a_T} (u_{t_3^*+1} + u_{t_3^*+2} + \ldots + u_{[\tau T]}) \Rightarrow y^*.$$

The proof for the lemma is thus finished.

Proof of Lemma 4

When $Th_1 + 1 \leqslant t \leqslant t_1^* - Th_1$, we have

$$\begin{aligned} T^{-2} h_1^{-1} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{T}\right) y_{i-1}^2 &\leqslant T^{-2} h_1^{-1} \max_{1 \leqslant i \leqslant t_1^*} y_{i-1}^2 \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{T}\right) \\ &= O_p(1), \end{aligned}$$

because $\max_{1 \leqslant i \leqslant t_1^*} y_{i-1}^2 = O_p(T)$ follows easily by applying the functional continuous mapping theorem to the result of Lemma 3 in the corresponding regime. Also from Lemma 2 we have $\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{T}\right) = O(Th_1)$.

When $t_1^* + Th_1 + 1 < t \leqslant t_2^* - Th_1$, we have

$$\begin{aligned} &\frac{1}{T \phi_1^{2(t+Th_1-1-t_1^*)}} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1}^2 \\ \leqslant\ &\frac{1}{\phi_1^{2(t+Th_1-1-t_1^*)}} \max_{t_1^*+1 < i \leqslant t_2^*} |T^{-1} \phi_1^{-2(i-1-t_1^*)} y_{i-1}^2| \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \phi_1^{2(i-1-t_1^*)}. \end{aligned}$$

By the results of Lemma 3 and Lemma 2 in the corresponding regime, using an argument similar to the above first regime, the claimed results for the explosive regime follow easily.

When $t_2^* + Th_1 + 1 < t \leqslant t_3^* - Th_1$, we have

$$
\begin{aligned}
& \frac{1}{T\phi_1^{2(t_2^*-t_1^*)}\phi_2^{2(t-Th_1-1-t_2^*)}} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1}^2 \\
\leqslant & \frac{1}{\phi_2^{2(t-Th_1-1-t_2^*)}} \max_{t_2^*+1<i\leqslant t_3^*} |T^{-1}\phi_1^{-2(t_2^*-t_1^*)}\phi_2^{-2(i-1-t_2^*)} y_{i-1}^2| \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \phi_2^{2(i-1-t_2^*)}.
\end{aligned}
$$

Similarly, the claimed result for this regime follows from the result of Lemma 3 and Lemma 2 in the corresponding regime by the same argument used above.

When $t_3^* + Th_1 + 1 < t \leqslant T - Th_1$,

$$
\begin{aligned}
& (a_T \vee T^{1/2})^{-2} T^{-1} h_1^{-1} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1}^2 \\
\leqslant & \max_{t_3^*+1\leqslant i\leqslant T} \left| \frac{1}{a_T \vee T^{1/2}} y_{i-1} \right|^2 T^{-1} h_1^{-1} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right).
\end{aligned}
$$

Again, the claimed result follows from the result of Lemma 3 and Lemma 2 in the corresponding regime by the same argument used above.

Proof of Lemma 5

The proof uses a standard argument (i.e. the truncation argument and the exponential inequality) as in Li et al. (2016) and Phillips et al. (2017) to derive the stated uniform rates of convergence.

When $Th_1 + 1 \leqslant t \leqslant t_1^* - Th_1$ denote

$$
B_{t,T} = \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1} u_i.
$$

Notice that $\{y_{i-1}u_i\}$ is a m.d.s. indexed by i, with respect to the natural filtration generated by the $\{u_i\}$ sequence. Define

$$
b_i = y_{i-1}u_i, \quad \bar{b}_i = y_{i-1}u_i \mathbb{I}(|y_{i-1}| \leqslant c_1)\mathbb{I}(|u_i| \leqslant c_2)
$$

and

$$
\tilde{b}_i = b_i - \bar{b}_i.
$$

Further, define

$$
\bar{B}_{t,T} = \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) (\bar{b}_i - E(\bar{b}_i|\mathcal{F}_{i-1}))
$$

and

$$
\tilde{B}_{t,T} = \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) (\tilde{b}_i - E(\tilde{b}_i|\mathcal{F}_{i-1}))
$$

where both $\bar{B}_{t,T}$ and $\tilde{B}_{t,T}$ are martingales by construction. Also note that

$$\begin{aligned}
\bar{B}_{t,T} + \tilde{B}_{t,T} &= \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)(\bar{b}_i + \tilde{b}_i) - \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) E(\bar{b}_i + \tilde{b}_i|\mathcal{F}_{i-1}) \\
&= \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) b_i - \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) E(b_i|\mathcal{F}_{i-1}) \\
&= \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) b_i = B_{t,T}
\end{aligned}$$

where we use the fact that $b_i = y_{i-1}u_i$ is a m.d.s. indexed by i. We have thus decomposed $B_{t,T}$ into two terms, where each term is by definition a sum of m.d.s., with $\bar{B}_{t,T}$ having bounded support.

Next, we show

$$\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1} |\bar{B}_{t,T}| = O_p\left(\sqrt{T^2 h_1 \log(T)}\right), \tag{S.1}$$

$$\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1} |\tilde{B}_{t,T}| = o_p\left(\sqrt{T^2 h_1 \log(T)}\right) \tag{S.2}$$

which imply (9).

We first prove (S.1). $\bar{B}_{t,T}$ is by construction a sum of m.d.s. with bounded support. We first compute the previsible quadratic variation,

$$\begin{aligned}
\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1} V_{t,T}^2 &:= \max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1} \sum_{i=t-Th_1}^{t+Th_1} E\left(G^2\left(\frac{i-t}{Th_1}\right)(\bar{b}_i - E(\bar{b}_i|\mathcal{F}_{i-1}))^2|\mathcal{F}_{i-1}\right) \\
&\leqslant \max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1} \sum_{i=t-Th_1}^{t+Th_1} G^2\left(\frac{i-t}{Th_1}\right) E(\bar{b}_i^2|\mathcal{F}_{i-1}) \\
&= \max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1} \sum_{i=t-Th_1}^{t+Th_1} G^2\left(\frac{i-t}{Th_1}\right) y_{i-1}^2 \mathbb{I}(|y_{i-1}| \leqslant c_1)\sigma_i^2 E(\mathbb{I}(|u_i| \leqslant c_2)|\mathcal{F}_{i-1}) \\
&\leqslant \max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1} \sum_{i=t-Th_1}^{t+Th_1} G^2\left(\frac{i-t}{Th_1}\right) y_{i-1}^2\sigma_i^2 \\
&\leqslant \max_i y_{i-1}^2\sigma_i^2 \max_{i,t} G\left(\frac{i-t}{Th_1}\right) \max_t \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \\
&= O_p\left(\left(\sqrt{T}\right)^2\right) O(Th_1) = O_p(T^2 h_1).
\end{aligned}$$

In the above, we have used the shorthand notation $\max_i$, $\max_t$ and $\max_{i,t}$ when the range of the maximal taken is implicit in the context. Now

$$\begin{aligned}
& P\left(\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}|\bar{B}_{t,T}|>x\right)\\
\leqslant\ & P\left(\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}|\bar{B}_{t,T}|>x,\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}V_{t,T}^2\leqslant y\right)+P\left(\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}V_{t,T}^2>y\right)\\
\leqslant\ & \sum_{t=1}^{T}P\left(|\bar{B}_{t,T}|>x,\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}V_{t,T}^2\leqslant y\right)+P\left(\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}V_{t,T}^2>y\right)\\
\leqslant\ & \sum_{t=1}^{T}P\left(|\bar{B}_{t,T}|>x,V_{t,T}^2\leqslant y\right)+P\left(\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}V_{t,T}^2>y\right).
\end{aligned}$$

By the exponential inequality for m.d.s. in Theorem 1.2A of de la Peña (1999),

$$\sum_{t=1}^{T}P(|\bar{B}_{t,T}|>x,V_{t,T}^2\leqslant y)\leqslant\sum_{t=1}^{T}\exp\left(-\frac{x^2}{2(y+c_1c_2x)}\right).$$

Then choosing $x=\mathbb{C}Th_1^{1/2}\log^{1/2}(T)$, $y=\mathbb{C}T^2h_1$, $c_1=(T^2h_1)^{\frac{1}{2}}\left(T^{\frac{3}{2}}h_1\right)^{-\frac{1}{p-1}}T^{-\frac{1}{2(p-1)}}\log^{-\frac{1}{2}}(T)$, $c_2=\left(T^{\frac{3}{2}}h_1\right)^{\frac{1}{p-1}}T^{\frac{1}{2(p-1)}}$, then the above becomes

$$\begin{aligned}
\sum_{t=1}^{T}\exp\left(-\frac{x^2}{2(y+c_1c_2x)}\right) &= \sum_{t=1}^{T}\exp\left(-\frac{\mathbb{C}^2T^2h_1\log(T)}{2(\mathbb{C}T^2h_1+\mathbb{C}T^2h_1)}\right)\\
&= \sum_{t=1}^{T}\exp\left(-\frac{\mathbb{C}\log(T)}{4}\right)\\
&= T\times T^{-\mathbb{C}/4}\to 0,
\end{aligned}$$

if we choose $\mathbb{C}$ large enough. On the other hand, in view of $\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}V_{t,T}^2=O_p(T^2h_1)$ derived above and $y=\mathbb{C}T^2h_1$, $P(\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}V_{t,T}^2>y)$ can be made arbitrarily small by choosing $\mathbb{C}$ large enough. In total we have

$$P\left(\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}|\bar{B}_{t,T}|>\mathbb{C}Th_1^{1/2}\log^{1/2}(T)\right)\to 0$$

thus (S.1) is proved.

For (S.2), notice that

$$\begin{aligned}
|\tilde{B}_{t,T}| &= \left|\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)(\tilde{b}_i-E(\tilde{b}_i|\mathcal{F}_{i-1}))\right|\\
&\leqslant \sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)|\tilde{b}_i|+\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)|E(\tilde{b}_i|\mathcal{F}_{i-1})|\\
&= B_1+B_2
\end{aligned}$$

where B_1 and B_2 are defined implicitly. First consider B_1, and notice that

$$\begin{aligned}
|\tilde{b}_i| &= |y_{i-1}u_i\mathbb{I}(|y_{i-1}|>c_1\text{ or }|u_i|>c_2)|\\
&\leqslant |y_{i-1}u_i\mathbb{I}(|y_{i-1}|>c_1)|+|y_{i-1}u_i\mathbb{I}(|u_i|>c_2)|
\end{aligned}$$

so that

$$\begin{aligned}
P\left(\max_{1\leqslant i\leqslant t_1^*}|\tilde{b}_i|>0\right) &\leqslant P\left(\max_{1\leqslant i\leqslant t_1^*}|y_{i-1}u_i\mathbb{I}(|y_{i-1}|>c_1)|>0\right)+P\left(\max_{1\leqslant i\leqslant t_1^*}|y_{i-1}u_i\mathbb{I}(|u_i|>c_2)|>0\right)\\
&\leqslant P\left(\max_{1\leqslant i\leqslant t_1^*}|y_i|>c_1\right)+P\left(\max_{1\leqslant i\leqslant t_1^*}|u_i|>c_2\right)\\
&\leqslant P\left(\max_{1\leqslant i\leqslant t_1^*}|y_i|>c_1\right)+\sum_{i=1}^{T}P\left(|u_i|>c_2\right)\\
&\leqslant P\left(\max_{1\leqslant i\leqslant t_1^*}|T^{-1/2}y_i|>T^{-1/2}c_1\right)+\mathbb{C}\left(\frac{TE|u_i|^p}{c_2^p}\right)
\end{aligned}$$

where in the second step, we have used the fact that the events $\{\max_{1\leqslant i\leqslant t_1^*}|y_{i-1}u_i\mathbb{I}(|y_{i-1}|>c_1)|>0\}$ and $\{\max_{1\leqslant i\leqslant t_1^*}|y_{i-1}u_i\mathbb{I}(|u_i|>c_2)|>0\}$ respectively imply $\{\max_{1\leqslant i\leqslant t_1^*}|y_i|>c_1\}$ and $\{\max_{1\leqslant i\leqslant t_1^*}|u_i|>c_2\}$. Using the definition of c_1

$$T^{-1/2}c_1=\left(\frac{\log(T)}{T^{1-\frac{4}{p-1}}h_1^{1-\frac{2}{p-1}}}\right)^{-\frac{1}{2}}\to\infty$$

by assumption B2. Then using the fact that $\max_{1\leqslant i\leqslant t_1^*}|T^{-1/2}y_i|=O_p(1)$, we have

$$P(\max_{1\leqslant i\leqslant t_1^*}|T^{-1/2}y_i|>T^{-1/2}c_1)\to 0. \tag{S.3}$$

Now using the definition of c_2 and noticing the finiteness of $E|u_i|^p$, we have

$$\frac{TE|u_i|^p}{c_2^p}=O\left(\left(\frac{1}{T^{1+\frac{1}{p}}h_1}\right)^{\frac{p}{p-1}}\right)=o(1).$$

Thus in total we have $P(\max_{1\leqslant i\leqslant T}|\tilde{b}_i|>0)=o(1)$. It then follows that

$$\begin{aligned}
\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}|B_1| &= \max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}\left|\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)|\tilde{b}_i|\right|\\
&\leqslant \max_{1\leqslant i\leqslant T}|\tilde{b}_i|\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)\\
&= o_p(Th_1)=o_p\left(\sqrt{T^2h_1\log(T)}\right)
\end{aligned} \tag{S.4}$$

where we have used $\max_t\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)=O(Th_1)$.

Now, for B_2,

$$\begin{aligned}\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)|E(\tilde{b}_i|\mathcal{F}_{i-1})| &\leqslant \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) E(|y_{i-1}u_i\mathbb{I}(|y_{i-1}|>c_1)||\mathcal{F}_{i-1}) \\ &\quad + \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) E\left(|y_{i-1}u_i\mathbb{I}(|u_i|>c_2)||\mathcal{F}_{i-1}\right) \\ &\leqslant \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)|y_{i-1}|\mathbb{I}(|y_{i-1}|>c_1)E(|u_i||\mathcal{F}_{i-1}) \\ &\quad + \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)|y_{i-1}|E\left(|u_i\mathbb{I}(|u_i|>c_2)||\mathcal{F}_{i-1}\right) \\ &= B_{21}+B_{22}\end{aligned}$$

where B_{21} and B_{22} are defined implicitly. For B_{22},

$$\begin{aligned}\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}|B_{22}| &= \max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}\left|\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)|y_{i-1}|E(|u_i\mathbb{I}(|u_i|>c_2)||\mathcal{F}_{i-1})\right| \\ &\leqslant \max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}\left|\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)|y_{i-1}|\frac{E(|u_i|^p|\mathcal{F}_{i-1})}{c_2^{p-1}}\right| \\ &\leqslant \mathbb{C}\frac{1}{c_2^{p-1}}\max_{1\leqslant i\leqslant t_1^*}|y_{i-1}|\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}\left|\sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)\right| \\ &= \frac{1}{c_2^{p-1}}O_p\left(\sqrt{T}\right)O(Th_1) = O_p\left(\frac{1}{c_2^{p-1}}T^{3/2}h_1\right),\end{aligned}$$

Using the definition of c_2, the above becomes

$$\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}|B_{22}| = O_p\left(T^{-1/2}\right) = o_p(1) = o_p\left(\sqrt{T^2h_1\log(T)}\right). \tag{S.5}$$

For B_{21}, first notice that

$$P\left(\max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1}|y_{i-1}|\mathbb{I}(|y_{i-1}|>c_1)>0\right) \leqslant P\left(\max_{Th_1+1\leqslant i\leqslant t_1^*-Th_1}|y_{i-1}|>c_1\right)$$

which is $o(1)$ from (S.3). Thus

$$\begin{aligned}&\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}|B_{21}| \\ &= \max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}\left|\sum_{i=t-Th_1}^{t+Th_1} G_{h_1}\left(\frac{i-t}{T}\right)|y_{i-1}|\mathbb{I}(|y_{i-1}|>c_1)E(|u_i||\mathcal{F}_{i-1})\right| \\ &\leqslant \max_{1\leqslant i\leqslant t_1^*}|y_{i-1}|\mathbb{I}(|y_{i-1}|>c_1)\max_{1\leqslant i\leqslant t_1^*}E(|u_i||\mathcal{F}_{i-1})\max_{Th_1+1\leqslant t\leqslant t_1^*-Th_1}\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right) \\ &= o_p(Th_1) = o_p\left(\sqrt{T^2h_1\log(T)}\right).\end{aligned} \tag{S.6}$$

Combining results (S.4), (S.5) and (S.6), we have proved (S.2).

When $t_1^* + Th_1 + 1 < t \leqslant t_2^* - Th_1$, during the explosive regime, we use exactly the same strategy of proof as in the previous regime, although the constants used and the rates derived will be different. Define

$$B'_{t,T} = \phi_1^{-(t+Th_1-1-t_1^*)} \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) y_{i-1}u_i.$$

Define

$$b'_{i,t} = \phi_1^{-(t+Th_1-1-t_1^*)} y_{i-1}u_i$$

$$\bar{b}'_{i,t} = \phi_1^{-(t+Th_1-1-t_1^*)} y_{i-1}u_i \mathbb{I}(|\phi_1^{-(i-1-t_1^*)} y_{i-1}| \leqslant c_1)\mathbb{I}(\phi_1^{(i-t-Th_1)}|u_i| \leqslant c_2),$$

and

$$\tilde{b}'_i = b'_i - \bar{b}'_i.$$

Also define

$$\bar{B}'_{t,T} = \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)(\bar{b}'_i - E(\bar{b}'_i|\mathcal{F}_{i-1}))$$

and

$$\tilde{B}'_{t,T} = \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right)(\tilde{b}'_i - E(\tilde{b}'_i|\mathcal{F}_{i-1})).$$

Because $\phi_1^{-(t+Th_1-1-t_1^*)} y_{i-1}u_i$ is a m.d.s. indexed by i, we again have

$$\bar{B}'_{t,T} + \tilde{B}'_{t,T} = B'_{t,T}.$$

We have thus decomposed $B'_{t,T}$ into two terms, where both terms are by construction martingales, with $\bar{B}'_{t,T}$ having bounded support.

Next, we show

$$\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} |\bar{B}'_{t,T}| = O_p\left(T^{1/p+1/2}\log^2(T)\right), \tag{S.7}$$

$$\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} |\tilde{B}'_{t,T}| = o_p\left(T^{1/p+1/2}\log^2(T)\right) \tag{S.8}$$

and the claimed result of the lemma follows easily.

We first prove (S.7). $\bar{B}'_{t,T}$ is clearly a sum of m.d.s. with bounded support. We first compute the previsible quadratic variation. Similar to the previous regime

$$\begin{aligned}
& \max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} V_T'^2 \\
\leqslant & \max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} \phi_1^{-2(t+Th_1-1-t_1^*)} \sum_{i=t-Th_1}^{t+Th_1} G^2\left(\frac{i-t}{Th_1}\right) y_{i-1}^2\sigma_i^2 \\
\leqslant & \max_i |\phi_1^{-2(i-1-t_1^*)} y_{i-1}^2| \max_i \sigma_i^2 \max_{i,t} G\left(\frac{i-t}{Th_1}\right) \max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} \sum_{i=t-Th_1}^{t+Th_1} \phi_1^{2(i-t-Th_1)} G\left(\frac{i-t}{Th_1}\right) \\
= & O_p(T)
\end{aligned}$$

where we use the result of Lemma 3. Now using a similar argument as before,

$$P\left(\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}|\bar{B}'_{t,T}|>x\right)\leqslant\sum_{t=1}^{T}P\left(|\bar{B}'_{t,T}|>x,V_t'^2\leqslant y\right)+P\left(\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}V_t'^2>y\right).$$

By Theorem 1.2A of de la Peña (1999), we have

$$\sum_{t=1}^{T}P(|\bar{B}'_{t,T}|>x,V_t'^2\leqslant y)\quad\leqslant\quad\sum_{t=1}^{T}\exp\left(-\frac{x^2}{2(y+c_1c_2x)}\right).$$

Then choosing $x=\mathbb{C}T^{1/p+1/2}\log^2(T),y=\mathbb{C}T,c_1=T^{1/2}\log^{1/2}(T),c_2=T^{1/p}\log^{1/2}(T)$, the above becomes

$$\begin{aligned}\sum_{t=1}^{T}\exp\left(-\frac{x^2}{2(y+c_1c_2x)}\right)&=\sum_{t=1}^{T}\exp\left(-\frac{\mathbb{C}^2T^{1+2/p}\log^4(T)}{2(\mathbb{C}T+\mathbb{C}T^{1+2/p}\log^3(T)}\right)\\&\sim\sum_{t=1}^{T}\exp\left(-\frac{\mathbb{C}\log(T)}{2}\right)\\&=T^{1-\mathbb{C}/2}\to 0\end{aligned}$$

if we choose $\mathbb{C}$ large enough. On the other hand, in view of the facts that $\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}V_t'^2=O_p(T)$ and $y=\mathbb{C}T$, $P(\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}V_t'^2>y)$ can be made arbitrarily small by choosing $\mathbb{C}$ large enough. In total we have

$$P\left(\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}|\bar{B}'_{t,T}|>\mathbb{C}T^{1/p+1/2}\log^2(T)\right)\to 0$$

and we have (S.7) proved.

For (S.8), notice that

$$\begin{aligned}|\tilde{B}'_{t,T}|&=\left|\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)(\tilde{b}'_i-E(\tilde{b}'_i|\mathcal{F}_{i-1}))\right|\\&\leqslant\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)|\tilde{b}'_i|+\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)|E(\tilde{b}'_i|\mathcal{F}_{i-1})|\\&=B'_1+B'_2.\end{aligned}$$

First look at B'_1. Notice that by definition

$$\begin{aligned}B'_1&=\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)|\tilde{b}'_i|\\&\leqslant\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)|\phi_1^{-(t+Th_1-1-t_1^*)}y_{i-1}u_i\mathbb{I}(|\phi_1^{-(i-1-t_1^*)}y_{i-1}|>c_1)|\\&\quad+\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)|\phi_1^{-(t+Th_1-1-t_1^*)}y_{i-1}u_i\mathbb{I}(|\phi_1^{(i-t-Th_1)}u_i|>c_2)|\\&=B'_{11}+B'_{12}.\end{aligned}$$

For B'_{11}, notice that

$$P\left(\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}|B'_{11}|>0\right)\leqslant P\left(\max_{t_1^*+1\leqslant i\leqslant t_2^*}|\phi_1^{-(i-1-t_1^*)}y_{i-1}|>c_1\right)\to 0,$$

by the definition of c_1. We thus have $\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}|B'_{11}|=o_p(1)$. B'_{12} can be analyzed in a similar way as B'_{11} and notice that

$$\begin{aligned}
P\left(\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}|B'_{12}|>0\right) &\leqslant P\left(\max_{t_1^*+1\leqslant i\leqslant t_2^*}|u_i|>c_2\right)\\
&\leqslant \sum_{i=t_1^*+1}^{t_2^*}P(|u_i|>c_2)\\
&\leqslant \sum_{i=t_1^*+1}^{t_2^*}\frac{E|u_i|^p}{c_2^p}\to 0,
\end{aligned}$$

using the definition of c_2. We thus have

$$\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}|B'_{12}|=o_p(1).$$

In total, we then have

$$\max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1}|B'_1|=o_p\left(T^{1/2+1/p}\log^2(T)\right). \tag{S.9}$$

Now, for B'_2,

$$\begin{aligned}
&\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)|E(\tilde{b}_i|\mathcal{F}_{i-1})|\\
\leqslant&\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)\left|\phi_1^{-(i-1-t_1^*)}y_{i-1}\right|\phi_1^{(i-t-Th_1)}\mathbb{I}(|\phi_1^{-(i-1-t_1^*)}y_{i-1}|>c_1)E(|u_i||\mathcal{F}_{i-1})\\
&+\sum_{i=t-Th_1}^{t+Th_1}G\left(\frac{i-t}{Th_1}\right)\left|\phi_1^{-(i-1-t_1^*)}y_{i-1}\right|\phi_1^{(i-t-Th_1)}E(|u_i\mathbb{I}(|\phi_1^{(i-t_2^*)}u_i|>c_2)||\mathcal{F}_{i-1})\\
=&\ B'_{21}+B'_{22}.
\end{aligned}$$

First look at B'_{22},

$$\begin{aligned}
& \max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} |B'_{22}| \\
= & \max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} \left| \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) |\phi_1^{-(i-1-t_1^*)} y_{i-1}| \phi_1^{(i-t-Th_1)} \right. \\
& \left. E(|u_i \mathbb{I}(|\phi_1^{(i-t-Th_1)} u_i| > c_2)||\mathcal{F}_{i-1}) \right| \\
\leqslant & \ \mathbb{C} \max_{t_1^*+1\leqslant i\leqslant t_2^*} |\phi_1^{-(i-1-t_1^*)} y_{i-1}| \frac{E(|u_i|^p|\mathcal{F}_{i-1})}{c_2^{p-1}} \\
& \max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} \left| \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \phi_1^{(i-t-Th_1)} \right| \\
= & \ O_p\left(\sqrt{T}\right) \times o_p(1) \times O_p(1) = o_p(T^{1/2+1/p} \log^2(T)), \qquad \text{(S.10)}
\end{aligned}$$

where we have used the definition of c_2. Then we look at B'_{21}. First notice that

$$\begin{aligned}
& P\left(\max_{t_1^*+1\leqslant i\leqslant t_2^*} |\phi_1^{-(i-1-t_1^*)} y_{i-1}| \mathbb{I}(|\phi_1^{-(i-1-t_1^*)} y_{i-1}| > c_1) > 0 \right) \\
\leqslant & \ P\left(\max_{t_1^*+1\leqslant i\leqslant t_2^*} |\phi_1^{-(i-1-t_1^*)} y_{i-1}| > c_1 \right) \to 0.
\end{aligned}$$

Thus we have

$$\begin{aligned}
& \max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} |B'_{21}| \\
= & \max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} \left| \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \left|\phi_1^{-(i-1-t_1^*)} y_{i-1}\right| \mathbb{I}\left(\left|\phi_1^{-(i-1-t_1^*)} y_{i-1}\right| > c_1\right) \right. \\
& \left. \phi_1^{(i-t-Th_1)} E(|u_i||\mathcal{F}_{i-1}) \right| \\
\leqslant & \max_{t_1^*+1\leqslant i\leqslant t_2^*} \left|\phi_1^{-(i-1-t_1^*)} y_{i-1}\right| \mathbb{I}\left(\left|\phi_1^{-(i-1-t_1^*)} y_{i-1}\right| > c_1\right) \\
& \max_{t_1^*+1\leqslant i\leqslant t_2^*} E(|u_i||\mathcal{F}_{i-1}) \max_{t_1^*+Th_1<t\leqslant t_2^*-Th_1} \left| \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) \phi_1^{(i-t-Th_1)} \right| \\
= & \ o_p(1) \times C \times O(1) = o_p\left(T^{1/2+1/p} \log^2(T)\right). \qquad \text{(S.11)}
\end{aligned}$$

Combining the results in (S.9), (S.10) and (S.11) we have (S.8) proved.

When $t_2^* + Th_1 + 1 < t \leqslant t_3^* - Th_1$, again define

$$B''_{t,T} = \sum_{i=t-Th_1}^{t+Th_1} G\left(\frac{i-t}{Th_1}\right) b''_i$$

with

$$b''_i = \phi_2^{-(t-Th_1-1-t_2^*)} y_{i-1} u_i.$$

The claimed uniform rate of convergence can be derived in a similar fashion as in the explosive regime, but with the truncated m.d.s. defined as

$$\bar{b}_i'' = \phi_2^{-(t-Th_1-1-t_2^*)} y_{i-1} u_i \mathbb{I}(|\phi_2^{-(i-1-t_2^*)} y_{i-1}| \leqslant c_1) \mathbb{I}(\phi_2^{i-t+Th_1} |u_i| \leqslant c_2)$$

and its complement as

$$\tilde{b}_i'' = b_i'' - \bar{b}_i''$$

and with the following choice of constants $x = \mathbb{C} T^{1/2+1/p} \phi_1^{(t_2^*-t_1^*)} \log^2(T)$, $y = \mathbb{C} T \phi_1^{2(t_2^*-t_1^*)}$, $c_1 = T^{1/2} \phi_1^{(t_2^*-t_1^*)} \log^{1/2}(T)$, $c_2 = T^{1/p} \log^{1/2}(T)$.

When $t_3^* + Th_1 + 1 < t \leqslant T - Th_1$, the process is a random walk process that restarts from $y^* a_T$. When $a_T/T^{1/2} \to 0$ or $a_T/T^{1/2} \to 1$, the derivation and results are exactly the same as in the $Th_1 + 1 \leqslant t \leqslant t_1^* - Th_1$ case. When $a_T/T^{1/2} \to \infty$, the derivation is still the same as in the first random walk regime except with the following choice of constants $x = \mathbb{C} a_T (Th_1)^{1/2} \log^{1/2}(T)$, $y = \mathbb{C} a_T^2 Th_1$, $c_1 = a_T (Th_1)^{\frac{1}{2}} \left(T^{\frac{3}{2}} h_1\right)^{-\frac{1}{p-1}} T^{-\frac{1}{2(p-1)}} \log^{-\frac{1}{2}}(T)$, $c_2 = \left(T^{\frac{3}{2}} h_1\right)^{\frac{1}{p-1}} T^{\frac{1}{2(p-1)}}$. No further bandwidth assumptions are needed.

Therefore, the detailed proofs for the crash regime and the last random walk regime are omitted to avoid repetition.

Proof of Lemma 6

First denote $z_i = u_i^2 - \sigma_i^2$, and denote the object of interest as

$$Z_{s,T} = \sum_{i=1}^{T} K\left(\frac{i/T - \tau_s}{h_2}\right) z_{i-1}.$$

Notice that $\{z_i\}$ is a m.d.s. indexed by i, with respect to the natural filtration generated by the $\{u_i\}$ sequence. The proof uses a truncation technique similar to that used in the proof of Lemma 5. Define

$$\bar{z}_i = z_i \mathbb{I}(|z_i| \leqslant c)$$

and

$$\tilde{z}_i = z_i - \bar{z}_i.$$

Further, define

$$\bar{Z}_{s,T} = \sum_{i=1}^{T} K\left(\frac{i/T - \tau_s}{h_2}\right) (\bar{z}_i - E(\bar{z}_i | \mathcal{F}_{i-1}))$$

and

$$\tilde{Z}_{s,T} = \sum_{i=1}^{T} K\left(\frac{i/T - \tau_s}{h_2}\right) (\tilde{z}_i - E(\tilde{z}_i | \mathcal{F}_{i-1})).$$

Because z_i is a m.d.s. indexed by i, as in the proof of Lemma 5 we easily have

$$\bar{Z}_{s,T} + \tilde{Z}_{s,T} = Z_{s,T}.$$

We have thus decomposed $Z_{s,T}$ into two terms, where both terms are by definition martingales, with $\bar{Z}_{s,T}$ having bounded support.

Next, we show

$$\max_{1\leqslant s\leqslant N}|\bar{Z}_{s,T}| = O_p\left(\sqrt{Th_2\log(T)}\right), \tag{S.12}$$

$$\max_{1\leqslant s\leqslant N}|\tilde{Z}_{s,T}| = o_p\left(\sqrt{Th_2\log(T)}\right). \tag{S.13}$$

which will imply the result of the lemma.

We first prove (S.12). $\bar{Z}_{s,T}$ is clearly a sum of m.d.s. with bounded support. In the following, we again use $\max_s$ as the shorthand notation for $\max_{1\leqslant s\leqslant N}$. We first compute the previsible quadratic variation,

$$\begin{aligned}
\max_s V_{s,T}^2 &:= \max_s \sum_{i=1}^T E\left(K^2\left(\frac{i/T-\tau_s}{h_2}\right)(\bar{z}_i - E(\bar{z}_i|\mathcal{F}_{i-1}))^2|\mathcal{F}_{i-1}\right)\\
&\leqslant \max_s \sum_{i=1}^T K^2\left(\frac{i/T-\tau_s}{h_2}\right)E(\bar{z}_i^2|\mathcal{F}_{i-1})\\
&\leqslant \mathbb{C}\max_s \sum_{i=1}^T K^2\left(\frac{i/T-\tau_s}{h_2}\right)\sigma_i^4 E(\mathbb{I}(|z_i|\leqslant c)|\mathcal{F}_{i-1})\\
&\leqslant \mathbb{C}\max_s \sum_{i=1}^T K^2\left(\frac{i/T-\tau_s}{h_2}\right)\sigma_i^4.\\
&\leqslant \mathbb{C}\max_{1\leqslant i\leqslant T}\sigma_i^4 \max_{1\leqslant i\leqslant T,1\leqslant s\leqslant N} K\left(\frac{i/T-\tau_s}{h_2}\right)\max_s\left|\sum_{i=1}^T K\left(\frac{i/T-\tau_s}{h_2}\right)\right|\\
&= O_p(Th_2).
\end{aligned}$$

Now using the same argument as in the proof of Lemma 5,

$$P\left(\max_s|\bar{Z}_{s,T}|>x\right)\leqslant\sum_{s=1}^N P(|\bar{Z}_{s,T}|>x, V_{s,T}^2\leqslant y)+P(\max_s V_{s,T}^2>y).$$

By Theorem 1.2A of de la Peña (1999), we have for any s,

$$P(|\bar{Z}_{s,T}|>x, V_{s,T}^2\leqslant y)\leqslant\exp\left(-\frac{x^2}{2(y+cx)}\right).$$

Then, choosing $x=\mathbb{C}\,(Th_2)^{1/2}\log^{1/2}(T), y=\mathbb{C}Th_2, c=(Th_2)^{1/2}\log^{-1/2}(T)$, we have

$$\begin{aligned}
\sum_{s=1}^N P\left(|\bar{Z}_{s,T}|>x, V_{s,T}^2\leqslant y\right) &\leqslant \sum_{s=1}^N \exp\left(-\frac{x^2}{2(y+cx)}\right)\\
&= \sum_{s=1}^N \exp\left(-\frac{\mathbb{C}^2 Th_2\log(T)}{2(\mathbb{C}Th_2+\mathbb{C}Th_2)}\right)\\
&= \sum_{s=1}^N \exp\left(-\frac{\mathbb{C}\log(T)}{4}\right)\\
&= N\times T^{-\mathbb{C}/4}\to 0,
\end{aligned}$$

if we choose $\mathbb{C}$ large enough. On the other hand, in view of $\max_s V_{s,T}^2 = O_p(Th_2)$ and the definition of y, $P(\max_s V_{s,T}^2>y)$ can be made arbitrarily small by choose $\mathbb{C}$ large enough. In total we have

$$P\left(\max_s |\bar{Z}_{s,T}| > \mathbb{C}T^{1/2}h_2^{1/2}\log^{1/2}(T)\right) \to 0$$

thus (S.12) is proved.

For (S.13), notice that

$$\begin{aligned}
|\tilde{Z}_{s,T}| &= \left|\sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right)(\tilde{z}_i - E(\tilde{z}_i|\mathcal{F}_{i-1}))\right| \\
&\leqslant \sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right)|\tilde{z}_i| + \sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right)|E(\tilde{z}_i|\mathcal{F}_{i-1})| \\
&= Z_1 + Z_2.
\end{aligned}$$

First consider Z_1. Notice that by the definition of $|\tilde{z}_i|$,

$$\begin{aligned}
P\left(\max_s \sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right)|\tilde{z}_i| > 0\right) &\leqslant P\left(\max_{1\leqslant i\leqslant T}|\tilde{z}_i| > 0\right) \\
&= P\left(\max_{1\leqslant i\leqslant T}|z_i\mathbb{I}(|z_i|>c)| > 0\right) \\
&\leqslant \sum_{i=1}^{T} P(|z_i\mathbb{I}(|z_i|>c)| > 0) \\
&\leqslant \sum_{i=1}^{T} P(|z_i| > c) \\
&\leqslant \mathbb{C}\frac{TE|z_i|^{p/2}}{c^{p/2}} \sim \mathbb{C}\frac{TE|u_i|^p}{c^{p/2}}
\end{aligned}$$

where in the third inequality, we have used the fact that $\{|z_i| > c\}$ implies $\{|z_i\mathbb{I}(|z_i|>c)| > 0\}$. Using the definition of c, the above becomes

$$P\left(\max_s \sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right)|\tilde{z}_i| > 0\right) \leqslant \mathbb{C}\frac{T\log^{p/4}(T)}{(Th_2)^{p/4}} = \mathbb{C}\left(\frac{\log(T)}{T^{1-\frac{4}{p}}h_2}\right)^{p/4}$$

which is $o_p(1)$ under C3. We thus have $\max_s |Z_1| = o_p(1) = o_p\left(\sqrt{Th_2\log(T)}\right)$. Now, for Z_2,

$$\begin{aligned}
\sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right)|E(\tilde{z}_i|\mathcal{F}_{i-1})| &\leqslant \sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right)E(|z_i\mathbb{I}(|z_i|>c)||\mathcal{F}_{i-1}) \\
&\leqslant \sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right)\frac{E(|z_i|^{p/2}|\mathcal{F}_{i-1})}{c^{p/2-1}}.
\end{aligned}$$

Thus,

$$
\begin{aligned}
\max_s |Z_2| &\leqslant \max_s \left| \sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right) \frac{E(|u_i|^p|\mathcal{F}_{i-1})}{c^{p/2-1}} \right| \\
&\leqslant \mathbb{C}\frac{1}{c^{p/2-1}} \max_s \left| \sum_{i=1}^{T} K\left(\frac{i/T-\tau_s}{h_2}\right) \right| \\
&= O_p\left(\frac{1}{c^{p/2-1}} Th_2\right) = O_p\left(\frac{\log^{(p-2)/4}(T)}{(Th_2)^{(p-6)/4}}\right).
\end{aligned}
$$

By Assumption C3 and straightforward algebra, we have $\max_s |Z_2| = o_p(1)$. Combining results for the uniform rates of Z_1 and Z_2 we have proved (S.13).