

IRTG 1792 Discussion Paper 2020-018



A supreme test for periodic explosive GARCH

Stefan Richter ^{*}
Weining Wang ^{*2}
Wei Biao Wu ^{*3}



- * Universität Heidelberg, Germany
- *2 Humboldt-Universität zu Berlin, Germany
- *3 University of Chicago, USA

This research was supported by the Deutsche
Forschungsgesellschaft through the
International Research Training Group 1792
"High Dimensional Nonstationary Time Series".

<http://irtg1792.hu-berlin.de>
ISSN 2568-5619

International Research Training Group 1792

A supreme test for periodic explosive GARCH

Stefan Richter^{1,*}, Weining Wang^{2,**} and Wei Biao Wu^{3,†}

¹*Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany. e-mail: *stefan.richter@iwr.uni-heidelberg.de*

²*Department of Economics, City, U of London, center of applied statistics, HU Berlin, Unter den Linden 6, 10099 Berlin, Germany. e-mail: **weining.wang@city.ac.uk*

³*Department of Statistics, University of Chicago, 5734 S. University Avenue, Chicago, IL 60637, USA. e-mail: †wbwu@galton.uchicago.edu*

Abstract: We develop a uniform test for detecting and dating explosive behavior of a strictly stationary GARCH(r, s) (generalized autoregressive conditional heteroskedasticity) process. Namely, we test the null hypothesis of a globally stable GARCH process with constant parameters against an alternative where there is an 'abnormal' period with changed parameter values. During this period, the change may lead to an explosive behavior of the volatility process. It is assumed that both the magnitude and the timing of the breaks are unknown. We develop a double supreme test for the existence of a break, and then provide an algorithm to identify the period of change. Our theoretical results hold under mild moment assumptions on the innovations of the GARCH process. Technically, the existing properties for the QMLE in the GARCH model need to be reinvestigated to hold uniformly over all possible periods of change. The key results involve a uniform weak Bahadur representation for the estimated parameters, which leads to weak convergence of the test statistic to the supreme of a Gaussian Process. In simulations we show that the test has good size and power for reasonably large time series lengths. We apply the test to Apple asset returns and Bitcoin returns.

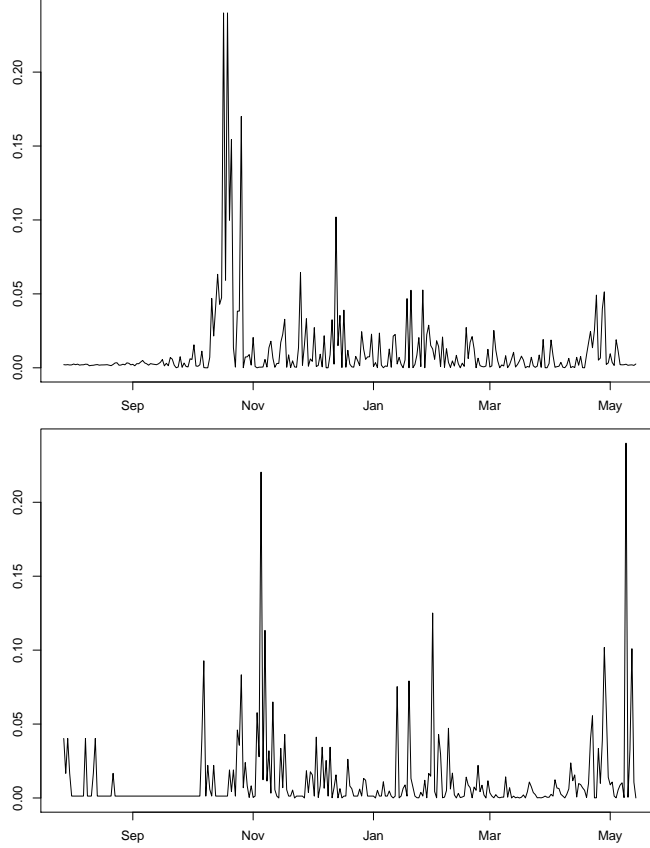
Keywords and phrases: GARCH, IGARCH, Change-point Analysis, Concentration Inequalities, Uniform Test.

1. Introduction

Volatility is an important indicator for economic and financial stability. There is a growing evidence for the unstable behavior of the historical volatility of numerous micro level as well as macro level data, such as individual stocks prices, asset returns, VIX, inflation and unemployment. [2] document the unstable behavior of higher moments of many economic variables, such as R&D (research and develop) rates related to the uncertainty about future productivity. It is understood that the nature of uncertainty is the unpredictability of any model to the future path of a time series. Therefore it may be connected with a change of the parameter values in the underlying data generating process. A direct empirical fact is that one often sees a sudden explosive behavior in the second moment of the process which bounces back after a while. For example, in Figure 1 we have plotted a realization of a piece-wise explosive GARCH(1,1) process and the log returns of Bitcoin. The whole time span is set to be (July 28, 2010- May 14, 2011) with an explosive period with changing parameter values (December 30, 2010 - October 15, 2011). The two trajectories of the time series look rather similar. Such kind of data phenomena suggest that the underlying processes have time varying parameters, calling for a rigorous econometric treatment for detection of change periods and corresponding inference. We see that the piecewise explosive GARCH process in Figure 1 captures the explosive behavior of Bitcoin in the squared returns. The aim

of our paper is to develop a generalised supreme test for GARCH models which is able to detect exuberance behaviour periods (periods with explosive parameter values) which are associated with the empirical phenomena of explosiveness in the second moment.

FIGURE 1. *Plot of simulated piecewise explosive squared returns (upper panel) versus Bitcoin squared log returns (lower panel). A stable period (20100728- 20110514) versus an explosive period (20100901-20110301).*



The highly celebrated autoregressive conditional heteroskedasticity (ARCH) model proposed by [6] is important for describing the pervasive phenomena of heteroskedasticity presented in many time series. One key generalization of ARCH is the GARCH model, i.e.

$$\begin{aligned} X_i^2 &= \zeta_i^2 \sigma_i^2, \\ \sigma_i^2 &= \alpha_0 + \sum_{j=1}^r \alpha_j X_{i-j}^2 + \sum_{k=1}^s \beta_k \sigma_{i-k}^2, \end{aligned} \quad (1)$$

where the conditional variance σ_i^2 depends on the past observations X_{i-j}^2 but also on the historical conditional variance σ_{i-k}^2 . ζ_i are assumed to be i.i.d. innovations, see for example [16] for more details of the model.

Numerous estimation methods for the parameters of GARCH models have been proposed, and the consistency and asymptotic normality has been carefully studied in the literature. A conventional estimation approach is the quasi maximum likelihood estimation (QMLE), e.g. [3]. Also [7] study

quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. [17] propose a least absolute deviation estimator. [12] establish consistency and asymptotic normality of the quasi-maximum likelihood estimator in the linear ARCH model. It is well known that under the assumption of strict stationarity of a GARCH model, there is still a region of parameter values allowing for realizations with unstable volatility behavior. The leading case is the "IGARCH" process. [15] looks at the behavior of an "IGARCH" process, and it is known that the unconditional mean of the IGARCH's conditional variance is not finite, which implies infinite second or higher moments (i.e. eruptive behavior). [14] provide an asymptotic theory for a strictly stationary GARCH(1,1) quasi-maximum likelihood estimator allowing for the case of IGARCH and explosive conditional variance and even nonstationarity. [12] consider asymptotic inference for nonstationary GARCH model.

Despite the rich empirical literature which suggests the existence of an unstable moment period of a GARCH process, there is only sparse literature on determining and testing the period of explosiveness. [9] provide a test for testing strict stationarity of a GARCH(1,1) process. Their results however do not lead to a handy tool for checking for the existence of an explosive period and determining its range. There is also a large and important literature on testing for explosiveness and dating the period of instability in the price or dividend processes of a financial asset using a supreme unit root test for bubbles. See for example [18] for a left-tailed augmented Dickey-Fuller test (ADF) for the explosive behavior in the 1990s Nasdaq. [10] considers such kind of bubble tests for crypto-currencies. [11] investigate a bubble test with a smooth time varying volatility function. The underlying models focus usually on unit root or explosive auto-regressive (AR) processes to test the change of the AR(1) coefficient. Often the variance of the errors stays the same or vary smoothly after the explosion which means that the volatility increase is mostly driven by the increase of the AR parameter. In our model we choose a different approach to model an explosion of volatility: we describe the evolution of the data-generating process by a GARCH process and therefore link the source of a change in the volatility to a change of the parameters in the volatility recursion.

It is worth noting that unlike a bubble test for an AR process, it is quite debatable to link a direct cause of the bursting behavior to the volatility process, see [13]. In contrary, volatility bursting can also be related to time varying risk aversion, sentiment, bubble or uncertainty. Nevertheless, we are trying to establish a theoretical rigorous framework of testing for the explosive interval using a GARCH model for the volatility process. It should be stressed that we focus on one aspect of the parameter and model uncertainty, namely changes in the parameters driving the volatility over time. We do not claim that our method can directly identify the cause of this behavior. In sum, we develop a change-point test for detecting possible unstable behavior of a strictly stationary GARCH(r, s) process. The null hypothesis is a GARCH process with globally constant parameters while the alternative is the existence of a period where the parameter values change to another (higher) value. This increase potentially leads to a period of explosive volatility.

Assuming that no information on the period and the change itself is available, we develop a test statistic based on supremes which searches over all possible sub-windows of the data. We prove asymptotic consistency and provide a limit distribution of our test statistic. It is important that the test is not of unit-root-type since hypothesis and alternative are still in the regime where the GARCH process is strictly stationary. The theoretical contributions are uniform consistency statements of the QML estimators over an arbitrary observation period, a uniform weak Bahadur representation and corresponding uniform distributional limit results. For the proofs we carve out

the essential analytical properties of the likelihood functions and use new concentration inequalities from [23] leading to mild moment assumptions.

We introduce some notations we use throughout the paper. For $q > 0$ and vector $v = (v_1, \dots, v_d)^\top \in \mathbb{R}^d$, let $|v|_1 := \sum_{i=1}^d |v_i|$. For matrices $A \in \mathbb{R}^{d \times d}$, we similarly use $|A|_1 := \sum_{i,j=1}^d |A_{i,j}|$. We denote by $|A|_2 = \max_{|v|=1} |Av|_2$ the spectral norm of A . We use $Z_n \xrightarrow{d} Z$ and $Z_n \xrightarrow{p} Z$ to denote convergence in distribution and convergence in probability for random variables Z_n, Z . For some sequence $(y_j)_{j \in \mathbb{N}}$ of real numbers and some sequence $(\chi_j)_{j \in \mathbb{N}}$ of nonnegative real numbers, define the weighted seminorm

$$|y|_{\chi,q} := \left(\sum_{j=1}^{\infty} \chi_j |y_j|^q \right)^{1/q}.$$

For some sequences (a_n) and (b_n) of positive numbers, write $a_n = O(b_n)$ or $a_n = o(b_n)$ if there exists a positive constant C such that $a_n/b_n \leq C$ or $a_n/b_n \rightarrow 0$ respectively. For two sequences of random variables (X_n) and (Y_n) , write $X_n = o_p(Y_n)$ (resp. $X_n = O_p(Y_n)$) if $X_n/Y_n \rightarrow 0$ in probability (X_n/Y_n is bounded in probability). For some random variable Z , define $\|Z\|_q := (\mathbb{E}|Z|^q)^{1/q}$. If $\|\cdot\|_q$ is applied to a matrix, this is meant by a component-wise operation. For the i.i.d. random variables $\zeta_i, i \in \mathbb{Z}$ used in the model definition (5), let $\mathcal{F}_i := (\zeta_i, \zeta_{i-1}, \dots)$. With some abuse of notation, we refer to \mathcal{F}_i also as the σ algebra generated by the entries of \mathcal{F}_i . For $X_i = h(\mathcal{F}_i)$, $i \in \mathbb{Z}$ with some measurable function h , we define the functional dependence measure (cf. [21]),

$$\delta_q(k) := \|X_i - X_i^*\|_q,$$

where $X_i^* = h(\mathcal{F}_i^*)$ and $\mathcal{F}_i^* := (\zeta_i, \dots, \zeta_1, \zeta_0^*, \zeta_{-1}, \zeta_{-2}, \dots)$ with ζ_0^* being an independent copy of ζ_0 . Our text is organized as follows. Section 2 provides the results for the important GARCH(1,1) model and the corresponding test procedure. Section 3 is concerning the estimation and theoretical results in a general GARCH(r, s) model. In particular, Section 3.1 introduces the framework of the QMLE and the consistency of the QMLE, and Section 3.2 presents the theoretical foundations of our uniform test and discusses the estimation of the covariance matrix of the QMLE appearing in the test statistic. In Section 4 we analyze the size and the power of our test in simulations, while Section 5 discusses the behavior of the test in examples from practice. The technical proofs are delegated to the Appendix.

2. A supreme explosiveness test for GARCH(1,1)

In this section, we introduce our model by starting with a simple testing framework for the GARCH(1,1) model. Then we will provide a rigorous theoretical treatment by starting with a more general GARCH(r, s) model in the following section. We consider first of all the baseline GARCH(1,1) model over the whole sample period with possibly time-varying parameters,

$$\begin{aligned} X_i &= \zeta_i \sigma_i, \\ \sigma_i^2 &= \alpha_0(i) + \alpha_1(i) X_{i-1}^2 + \beta_1(i) \sigma_{i-1}^2, \quad i \in \mathbb{Z}, \end{aligned} \tag{2}$$

where ζ_i is an i.i.d. sequence of random variables with $\mathbb{E}\zeta_1 = 0$, $\mathbb{E}\zeta_1^2 = 1$, and $\alpha_0(i), \alpha_1(i), \beta_1(i) > 0$ are the underlying parameters at each time point. We collect data of this model at time points $1, \dots, n$.

We summarize the parameters into $\theta(i) = (\alpha_0(i), \alpha_1(i), \beta_1(i))'$. In the case that the parameters are constant, i.e. $\theta(i) \equiv \theta = (\alpha'_0, \alpha'_1, \beta'_1)'$, the top Lyapunov exponent associated with this model according to [5] is

$$\gamma(\theta) = \mathbb{E} \log(\alpha_1 \zeta_1^2 + \beta_1).$$

It is worth noting that $\gamma(\theta) < 0$ allows (for instance) the IGARCH case, i.e. $\alpha_1 + \beta_1 = 1$. The aim of this paper is to construct tests for a period of changed parameters in the GARCH model, which allows for the explosiveness of the variances of the process. Namely, we would like to test whether there exists a period $\{n_1, \dots, n_2\}$ (with $1 < n_1 < n_2 < n$), where the parameter values in (2) change their values compared with $\{1, \dots, n\}$. The task breaks into two parts: First, checking for the existence of a change, for which a uniform test is needed. Second, one has to identify the period of the change and to estimate the corresponding parameters.

For our studies, let

$$\Theta = \{\theta = (\alpha_0, \alpha_1, \beta_1) \in \mathbb{R}^3 : \gamma(\theta) < 0, \alpha_0, \alpha_1, \beta_1 > 0\}$$

be the parameter space which contains all possible configurations of $\theta = (\alpha_0, \alpha_1, \beta_1)$.

Let $\theta^*(i) = (\alpha_0^*(i), \alpha_1^*(i), \beta_1^*(i))'$ denote the true parameter in the baseline model, which possibly has a period of change in $\{\lfloor n\tau_1^* \rfloor + 1, \dots, \lfloor n\tau_2^* \rfloor\}$ (where $\tau_1^*, \tau_2^* \in [0, 1]$, $\tau_1^* < \tau_2^*$) with a magnitude $\Delta^* = [\delta_1^*, \delta_2^*, \delta_3^*]'$. Namely,

$$\theta^*(i) = \begin{cases} \theta^*, & i \leq \lfloor n\tau_1^* \rfloor, \\ \theta^* + \Delta^*, & \lfloor n\tau_1^* \rfloor + 1 \leq i \leq \lfloor n\tau_2^* \rfloor, \\ \theta^*, & i > \lfloor n\tau_2^* \rfloor. \end{cases}$$

Here $\lfloor x \rfloor$ denotes the flooring operator, i.e. the largest integer smaller or equal to x . An interesting question is to test whether the process is stable, i.e. $\alpha_1^*(i) + \beta_1^*(i) < 1$ for all time points $i = 1, \dots, n$ versus the hypothesis that there exists a period explosiveness in which $\alpha_1^*(i) + \beta_1^*(i) > 1$ for some i . $\alpha_1^*(i) + \beta_1^*(i)$ is referred to as the persistency parameter in our setting. Graphically, this corresponds to the question if there exists regions where the process leaves the variance-stationary regime (i.e. the variance explodes).

We formulate our hypothesis therefore in the following way, with $\alpha_1^* + \beta_1^* = c$ ($c \leq 1$),

$$H_0 : \delta_2^* + \delta_3^* = 0 \quad \text{v.s.} \quad H_1 : \delta_2^* + \delta_3^* > 0.$$

To construct a test, we first derive estimators for the parameters. If the period described by τ_1^*, τ_2^* is known, we can use a standard quasi maximum likelihood estimation approach. It is not hard to see from (2), that in the case of constant parameters $\theta(i) \equiv \theta$,

$$\sigma_i^2 = \alpha_0 / (1 - \beta_1) + \alpha_1 \sum_{k=1}^{\infty} \beta_1^k X_{i-1-k}^2 \quad \text{a.s.}$$

The truncated version which can be calculated from a sample is

$$\sigma_i^{2c} = \alpha_0 / (1 - \beta_1) + \alpha_1 \sum_{k=1}^{i-2} \beta_1^k X_{i-1-k}^2.$$

The quasi likelihood approach is to use the negative log likelihood function

$$L_{n,\tau_1,\tau_2}^c(\theta) := \frac{1}{n} \sum_{i=\lfloor n\tau_1 \rfloor + 1}^{\lfloor n\tau_2 \rfloor} \ell(X_i^2, Y_i^c, \theta),$$

where $Y_i^c := (X_{i-1}^2, \dots, X_1^2, 0, 0, \dots)$ and

$$\ell(X_i^2, Y_i^c, \theta) := \frac{1}{2} \left(\frac{X_i^2}{\sigma_i^{2c}} + \log \sigma_i^{2c} \right).$$

The estimated parameter with observations during any given period $\{\lfloor n\tau_1 \rfloor + 1, \dots, \lfloor n\tau_2 \rfloor\}$ is defined to be $\hat{\theta}_{n,\tau_1,\tau_2} = \operatorname{argmin}_{\theta \in \Theta} L_{n,\tau_1,\tau_2}^c(\theta)$. It can be shown that under regularity conditions, $\hat{\theta}_{n,\tau_1,\tau_2}$ is asymptotically normal with covariance matrix

$$\Sigma = V(\theta^*)^{-1} I(\theta^*) V(\theta^*)^{-1}, \quad (3)$$

where

$$V(\theta) := \mathbb{E}[\nabla_{\theta}^2 \ell(X_i^2, Y_i, \theta)],$$

and

$$I(\theta) := \mathbb{E}[\nabla_{\theta} \ell(X_i^2, Y_i, \theta) \cdot \nabla_{\theta} \ell(X_i^2, Y_i, \theta)'].$$

Now we turn to our first step, i.e. a uniform test of the existence of the period of change. For given τ_1, τ_2 , the test statistic associated with our hypothesis H_0 of interest is

$$T_{\tau_1,\tau_2} = (\tau_2 - \tau_1)^{\chi} (H' \bar{\Sigma}_{n,\tau_1,\tau_2} H)^{-1/2} \{ \hat{\alpha}_{1,n,\tau_1,\tau_2} + \hat{\beta}_{1,n,\tau_1,\tau_2} - c \},$$

where χ is a scaling factor chosen for numerical stability, $\hat{\alpha}_{1,n,\tau_1,\tau_2}, \hat{\beta}_{1,n,\tau_1,\tau_2}$ are the second and third component of $\hat{\theta}_{n,\tau_1,\tau_2}$, and $\bar{\Sigma}_{n,\tau_1,\tau_2}$ is an estimator of Σ using observations outside $\{\lfloor n\tau_1 \rfloor, \dots, \lfloor n\tau_2 \rfloor\}$. For instance, we can set $\chi = 0.5$, $\bar{\Sigma}_{n,\tau_1,\tau_2}$ to be the standard covariance matrix estimator obtained by replacing V, I with their empirical counterparts (see (10) below) with observations outside $\{\lfloor n\tau_1 \rfloor, \dots, \lfloor n\tau_2 \rfloor\}$ and $H = (0, 1, 1)^\top$.

The feasible search set is defined to be $R_{\kappa,\kappa'} := \{(\tau_1, \tau_2) \in [0, 1]^2 : \tau_1 < \tau_2, 1 - \kappa' \geq \tau_2 - \tau_1 \geq \kappa\}$ (with some $\kappa, \kappa' > 0$, for instance $\kappa = \kappa' = 0.1$), ensuring proper estimation of Σ due to $1 - \kappa' \geq \tau_2 - \tau_1$ and a change detection based on enough samples due to $\tau_2 - \tau_1 \geq \kappa$. The supreme of $\sqrt{n} T_{\tau_1,\tau_2}$ converges asymptotically to the supreme of a Gaussian process, namely $\left\{ \frac{B(\tau_2) - B(\tau_1)}{(\tau_2 - \tau_1)^{1-\chi}} \right\}$, where $B(\cdot)$ is a 1-dimensional Brownian motion. We show this formally in Theorem 3.9 in Section 3.2.

Empirically, we cannot exhaust all the values $(\tau_1, \tau_2) \in R_{\kappa,\kappa'}$. We therefore need to restrict the calculation of the supreme to a set of grid points as an approximation to our supreme test statistics.

We summarize our uniform test as follows.

Step 0 Choose some $L > 0$ (number of grids associated with detection accuracy). The corresponding grid points are $\mathcal{G} = \left\{ \frac{j}{L} : j = 0, \dots, L \right\}$ on the time line.

Step 1 Let $H = (0, 1, 1)'$.

Step 2 Choose values for $\kappa, \kappa' \in (0, 1)$, and $\chi \in (0, 1)$.

Step 3 For each given interval $(\tau_1, \tau_2) \in R_{\kappa,\kappa'} \cap \mathcal{G}^2$, determine the associated QLME $\hat{\theta}_{n,\tau_1,\tau_2}$ defined in (8) and calculate $\bar{\Sigma}_{n,\tau_1,\tau_2}$ as in (10). Then we can calculate the test statistics as

$$\hat{B}_n(\tau_1, \tau_2) := \sqrt{n} (\tau_2 - \tau_1)^{\chi} \{ H' \bar{\Sigma}_{n,\tau_1,\tau_2} H \}^{-1/2} \{ H' \hat{\theta}_{n,\tau_1,\tau_2} - H' \theta^* \}.$$

(cf. (11)).

Step 4 For the critical value of this test, we can approximate the quantile of the test statistics via simulation under the null hypothesis H_0 : For large N (e.g. $N = 10000$), generate i.i.d. $\varepsilon_{i,k} \sim N(0, 1)$, $i = 1, \dots, n$ and calculate

$$\hat{\mu}_{n,k} := \sup_{(\tau_1, \tau_2) \in R_{\kappa, \kappa'} \cap \mathcal{G}^2} \frac{1}{\sqrt{n}} \frac{\sum_{i=[n\tau_1]+1}^{[n\tau_2]} \varepsilon_{i,k}}{(\tau_2 - \tau_1)^{1-\chi}}.$$

We define $\hat{q}_{W,\delta} := \hat{\mu}_{n, \lceil [N \cdot \delta] \rceil}$, where $\hat{\mu}_{n, [1]}, \dots, \hat{\mu}_{n, [N]}$ are the order statistics of $\hat{\mu}_{n,1}, \dots, \hat{\mu}_{n,N}$. Figure 2 shows how one calculates the supreme test statistics over different windows associated with the grid points.

Step 5 We can now make a test decision based on critical values from the previous steps. If

$$\sup_{(\tau_1, \tau_2) \in R_{\kappa, \kappa'} \cap \mathcal{G}^2} \hat{B}_n(\tau_1, \tau_2) > \hat{q}_{W,\delta}, \quad (4)$$

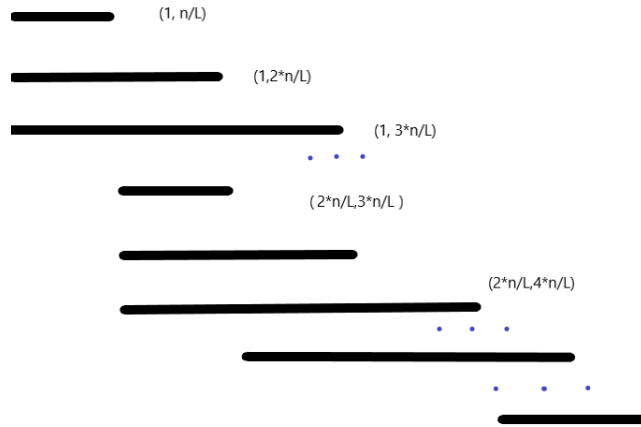
there is a significant shock in the parameter values. In this case, one can estimate the true shock period as $[\hat{\tau}_1^*, \hat{\tau}_2^*]$

$$(\hat{\tau}_{1,n}, \hat{\tau}_{2,n}) \in \operatorname{argmax}_{(\tau_1, \tau_2) \in R_{\kappa, \kappa'}} (\tau_2 - \tau_1)^\chi \bar{\Sigma}_{n, \tau_1, \tau_2, H}^{-1/2} \{H' \hat{\theta}_{n, \tau_1, \tau_2} - H' \theta^*\}.$$

If instead (4) does not hold, we conclude that there is no evidence for a period of parameter change.

Step 6 In case of the significance of our uniform test in Step 5, we re-estimate the parameter $\theta_{n, \hat{\tau}_1, \hat{\tau}_2}$, and produce the confidence interval based on Theorem 3.9.

FIGURE 2. Plot of the windows where the supreme is calculated.



We name our test as a GSRWW test (stands for GARCH supreme and the author abbreviations). The GSRWW test's limit distribution is $\sup_{\tau_1, \tau_2 \in R_{\kappa, \kappa'}} \left\{ \frac{B(\tau_2) - B(\tau_1)}{(\tau_2 - \tau_1)^{1-\chi}} \right\}$ (with $B(\cdot)$ a standard Brownian motion).

3. A supreme test for $GARCH(r, s)$

In this section, we formulate the test and provide necessary theoretical results in a general $GARCH(r, s)$ model. For $r, s \in \mathbb{N}$, $\theta(i) = (\alpha_0(i), \alpha_1(i), \dots, \alpha_r(i), \beta_1(i), \dots, \beta_s(i))'$, we consider the $GARCH(r, s)$ model

$$\begin{aligned} X_i^2 &= \zeta_i^2 \sigma_i^2, \\ \sigma_i^2 &= \alpha_0(i) + \sum_{j=1}^r \alpha_j(i) X_{i-j}^2 + \sum_{k=1}^s \beta_k(i) \sigma_{i-k}^2. \end{aligned} \quad (5)$$

Here, ζ_i are i.i.d. innovations with $\mathbb{E}\zeta_1 = 0$ and $\mathbb{E}\zeta_1^2 = 1$. We first analyze the model in the case that the parameters are constant, i.e. $\theta(i) \equiv \theta = (\alpha_0, \alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s)'$. We first present our set of assumptions ensuring the existence of a unique stationary solution to our model in (5). Define $f(\theta) = (\alpha_1, \dots, \alpha_r, \beta_1, \dots, \beta_s)'$ and let $e_j = (0, \dots, 0, 1, 0, \dots, 0)' \in \mathbb{R}^{r+s}$ be the unit column vector with j th element being 1, $1 \leq j \leq r + s$. Define the $(r + s) \times (r + s)$ -matrix

$$A_i(\theta) = (f(\theta)\zeta_i^2, e_1, \dots, e_{r-1}, f(\theta), e_{r+1}, \dots, e_{r+s-1})'.$$

Recall that $|A|_2$ is the spectral norm of a quadratic matrix A . Define the top Lyapunov exponent of $A_i(\theta)$,

$$\gamma(\theta) := \lim_{i \rightarrow \infty} \frac{1}{i} \log |A_i(\theta) A_{i-1}(\theta) \dots A_1(\theta)|_2.$$

which exists if $\mathbb{E}|\zeta_0^2|^a < \infty$ for some $a > 0$ (cf. [8]).

Assumption 3.1. *Suppose that*

(A1) ζ_0^2 has a non-degenerate distribution with $\mathbb{E}\zeta_0^2 = 1$.

(A2) Let $\alpha_{min} > 0$, and

$$\tilde{\Theta} = \{\theta \in \mathbb{R}_{\geq 0}^{r+s+1} : \alpha_0 \geq \alpha_{min}, \gamma(\theta) < 0 \text{ a.s.}, \sum_{j=1}^s \beta_j < 1\}. \quad (6)$$

Let $\Theta \subset \tilde{\Theta}$ be compact. Assume that $\theta^* \in \text{int}(\Theta)$.

(A3) Let $\mathcal{A}_\theta(z) := \sum_{i=1}^r \alpha_i z^i$, $\mathcal{B}_\theta(z) := 1 - \sum_{j=1}^s \beta_j z^j$. If $s > 0$, $\mathcal{A}_{\theta^*}(z)$ and $\mathcal{B}_{\theta^*}(z)$ have no common root, $\mathcal{A}_{\theta^*}(1) \neq 0$ and $\alpha_r^* + \beta_s^* \neq 0$.

Condition (A2) $\gamma(\theta) < 0$ ganrantees the strict stationarity of the GARCH process. Note that this includes parameter values corresponding to IGARCH or explosive GARCH $\sum_i \alpha_i + \sum_j \beta_j \geq 1$. Most of the following proposition is in the spirit of the introductory comments in [8]. However, to obtain appropriate statements for the functional dependence measure, we have to modify some arguments in the proof.

Proposition 3.2 (Existence of the GARCH model). *Let Assumption 3.1 hold. Then:*

- (i) (5) has a unique stationary solution $X_i^2 = H(\mathcal{F}_i)$, $i \in \mathbb{Z}$.
- (ii) There exists $q > 0$ with $\|X_0^2\|_q \leq D$ and $\delta_q^{X^2}(k) = O(c^k)$ for some $0 < c < 1$.

(iii) $\lambda_{\max}(B(\theta)) < 1$, where

$$B(\theta) = \begin{pmatrix} \beta_1 & \beta_2 & \dots & \dots & \beta_s \\ 1 & 0 & \dots & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 & 0 \end{pmatrix}.$$

3.1. QMLE in GARCH(r, s) and its consistency

In this subsection, we describe the QMLE and provide a theorem on its uniform consistency. For estimation of the parameters $\theta \in \Theta$, we consider the following QML approach. We denote by $Y_i^c := (X_{i-1}^2, X_{i-2}^2, \dots, X_1^2, 0, 0, \dots)$ the observed data until time $i - 1$. For $0 \leq \tau_1 < \tau_2 \leq 1$,

$$L_{n,\tau_1,\tau_2}^c(\theta) := \frac{1}{n} \sum_{i=\lfloor n\tau_1 \rfloor + 1}^{\lfloor n\tau_2 \rfloor} \ell(X_i^2, Y_i^c, \theta),$$

where

$$\ell(x, y, \theta) := \frac{1}{2} \left(\frac{x}{\sigma(y, \theta)^2} + \log \sigma(y, \theta)^2 \right)$$

and $\sigma(y, \theta)^2$ follows the recursion

$$\sigma(y, \theta)^2 = \alpha_0 + \sum_{j=1}^r \alpha_j y_j + \sum_{k=1}^s \beta_k \sigma((y_{k+1}, y_{k+2}, \dots), \theta)^2. \quad (7)$$

The analytic definition of the recursion of $\sigma(y, \theta)^2$ is formulated in a forward way (using y_1, y_2, \dots instead of y_{-1}, y_{-2}, \dots) because we plug in $y = Y_i^c$ which is formulated in a backward way, leading to the usual quasi-likelihood approach for GARCH models. Note that $\sigma(Y_i^c, \theta)$ in (7) terminates after a finite number of steps due to zeros in Y_i^c . Moreover, instead of using the truncated version $Y_i^c = (X_{i-1}^2, X_{i-2}^2, \dots, X_1^2, 0, \dots, 0)$ which corresponds to assuming that all initial values $X_0^2 = X_{-1}^2 = \dots = 0$, one can also use different initial values like $X_0^2 = X_{-1}^2 = \dots = \alpha_0$ or $X_0^2 = X_{-1}^2 = \dots = X_1^2$ as investigated in [8]. For a discussion of different initial values, consider [4] (in the case of strict stationarity).

With the defined likelihood function, for $0 \leq \tau_1 < \tau_2 \leq 1$, an estimator $\hat{\theta}_{n,\tau_1,\tau_2}$ of θ in the observation interval $i = \lfloor n\tau_1 \rfloor + 1, \dots, \lfloor n\tau_2 \rfloor$ is obtained by

$$\hat{\theta}_{n,\tau_1,\tau_2} := \operatorname{argmin}_{\theta \in \Theta} L_{n,\tau_1,\tau_2}^c(\theta). \quad (8)$$

Theorem 3.3 (Consistency of $\hat{\theta}_n$). *Let Assumption 3.1 hold. Then for each $\kappa > 0$,*

$$\sup_{0 \leq \tau_1 < \tau_2 \leq 1, |\tau_1 - \tau_2| \geq \kappa} |\hat{\theta}_{n,\tau_1,\tau_2} - \theta^*|_1 \xrightarrow{p} 0.$$

3.2. Limiting distribution

Given the consistency of our QMLE in a GARCH(r, s) model, we provide a distribution theorem for $\hat{\theta}_{n, \tau_1, \tau_2}$ which allows us to obtain critical values for the uniform test defined in Section 2 and more general tests. In our GSRWW test, we expect that over some observation period $\lfloor n\tau_1^* \rfloor + 1, \dots, \lfloor n\tau_2^* \rfloor$, certain combinations of the parameters are large. For instance in the GARCH(1,1) model with $\theta = (\alpha_0, \alpha_1, \beta_1)$, one can observe (partly) explosive behavior even in the stationary case if $\alpha_1 + \beta_1$ is large. To model the explosive behavior of the volatility process, we propose the following alternative 'shocked' $GARCH^{sh}(r, s)$ model where a change of parameter values happens at time $\lfloor n\tau_1^* \rfloor$. It pushes the parameters in a specific direction $H \in \mathbb{R}^{(r+s+1) \times 1}$ which lasts until $\lfloor n\tau_2^* \rfloor$, where the parameter values go back to their initial states.

In the following, we assume that $\theta^*(i) = (\alpha_0^*(i), \alpha_1^*(i), \dots, \alpha_r^*(i), \beta_1^*(i), \dots, \beta_s^*(i))'$ denotes the true parameter in the baseline model, which possibly has a period of change in $\{\lfloor n\tau_1^* \rfloor + 1, \dots, \lfloor n\tau_2^* \rfloor\}$ (where $\tau_1^*, \tau_2^* \in [0, 1]$, $\tau_1^* < \tau_2^*$). We suppose that the variation of θ^* over time reads

$$\theta^*(i) = \begin{cases} \theta^*, & i \leq \lfloor n\tau_1^* \rfloor, \\ \theta^* + H\Delta^*, & \lfloor n\tau_1^* \rfloor + 1 \leq i \leq \lfloor n\tau_2^* \rfloor, \\ \theta^*, & i > \lfloor n\tau_2^* \rfloor, \end{cases}$$

with some magnitude $\Delta^* \geq 0$ such that $\theta^* + H\Delta^* \in \Theta$. If $\Delta^* = 0$, $\theta^*(i) \equiv \theta^*$ is constant over time and no change of parameter values happens; otherwise there is a change. We call the model (5) with the above parameter configuration the shocked GARCH(r, s) model, $GARCH^{sh}(r, s)$ for short.

The condition $\theta^* + H\Delta^* \in \Theta$ means that even in the alternative we assume that the observed process is strictly stationary. It should be noted that the space of allowed parameter configurations can be relaxed even further by sacrificing the estimation accuracy of the constant term $\alpha_0^*(i)$ (cf. [9]).

We now give some important examples.

Example 3.4 ($GARCH^{sh}(1, 1)$). Here, $\theta = (\alpha_0, \alpha_1, \beta_1)'$. Fix some $\bar{\alpha}_1$, for instance $\bar{\alpha}_1 = 1$, where we understand $\alpha_1^* = \bar{\alpha}_1$ as a 'stable' parametrization of the process without a change. Assume that X_i^2 follows a $GARCH^{sh}(1, 1)$ model with some $\Delta^* \geq 0$. The existence of a break period is related to testing $\alpha_1^*(i) = \bar{\alpha}_1$ for all $i = 1, \dots, n$ against $\alpha_1^*(i) > \bar{\alpha}_1$ for some i . Thus it corresponds to $H = (0, 1, 0)'$,

$$H_0 : \Delta^* = 0 \quad v.s. \quad H_1 : \Delta^* > 0.$$

Example 3.5 ($GARCH^{sh}(r, s)$). For some fixed constants $\bar{\alpha}$, for instance $\bar{\alpha} = 1$, where we understand $\sum_{j=1}^r \alpha_j^* = \bar{\alpha}$ as a 'stable' parametrization. Assume that X_i^2 follows a $GARCH^{sh}(r, s)$ model with some $\Delta^* \geq 0$. The existence of an explosive period is then related to the question that whether $\sum_{j=1}^r \alpha_j^*(i) = \bar{\alpha}$ for all $i = 1, \dots, n$ or $\sum_{j=1}^r \alpha_j^*(i) > \bar{\alpha}$ for some i . This is our test with $H = (0, 1, \dots, 1, 0, \dots, 0)'$,

$$H_0 : \Delta^* = 0 \quad v.s. \quad H_1 : \Delta^* > 0.$$

To formulate the hypotheses in a more general way, we propose the following

$$H_0 : H'\theta^* = \bar{c} \quad v.s. \quad H_1 : H'\theta^* > \bar{c},$$

where $H \in \mathbb{R}^{(r+s+1) \times 1}$ is a vector, $\bar{c} \geq 0$. Motivated by the consistency result Theorem 3.3, we propose a test based on the supreme distance between the estimated targeting parameter value

and the value under the null,

$$\hat{B}_{n,H}^{(\chi)} := \sqrt{n} \sup_{0 \leq \tau_1 < \tau_2 \leq 1, |\tau_1 - \tau_2| \geq \delta} (\tau_2 - \tau_1)^\chi \{H' \hat{\theta}_{n,\tau_1,\tau_2} - H' \theta^*\},$$

where $\kappa > 0$ is some fixed parameter specifying the minimum length of the break period to be detected, and $\chi \in [0, 1]$ is a scaling parameter which can be chosen arbitrarily, for instance $\chi = \frac{1}{2}$.

Remark 3.6 (Consistency of the test statistic under the alternative). *If H_1 is true, there exists $\tau_1^* < \tau_2^*$ such that for $\lfloor n\tau_1^* \rfloor + 1 \leq i \leq \lfloor n\tau_2^* \rfloor$, $H' \theta^*(i) = H' \theta^* + H' H \Delta^* > H' \theta^*$, and by applying Theorem 3.3,*

$$\hat{B}_{n,H}^{(\chi)} \xrightarrow{p} \infty,$$

which ensures that our test has asymptotic power 1.

We conjecture that the result can be extended even to nonstationary alternatives where $\theta^* + H \Delta^* \notin \Theta$ as long as $H'(1, 0, \dots, 0) = 0$ ([9] find out that one cannot expect $\hat{\alpha}_0$ to be consistently estimated in the nonstationary regime). Therefore in practice, we can fix α_0 to be a constant and construct hypothesis in terms of other parameters.

To obtain the critical values of our test, we need to derive quantiles for the test statistics $\hat{B}_{n,H}^{(\chi)}$, which can be inferred by its limit distribution. The asymptotic distribution of $\hat{\theta}_{n,\tau_1,\tau_2}$ is strongly connected to the Fisher information matrices

$$V(\theta) := \mathbb{E}[\nabla_\theta^2 \ell(X_i^2, Y_i, \theta)],$$

and

$$I(\theta) := \mathbb{E}[\nabla_\theta \ell(X_i^2, Y_i, \theta) \cdot \nabla_\theta \ell(X_i^2, Y_i, \theta)'].$$

We then present some properties regarding $V(\theta), I(\theta)$. From [8] (proof of Theorem 2.2, part (ii) therein), we directly obtain (ii),(iii) of the following Proposition:

Proposition 3.7. *(Properties of the variance covariance matrix of the parameter estimates) Let Assumption 3.1 hold. Assume that $\mu_4 := \mathbb{E}\zeta_0^4 < \infty$. Then:*

- (i) *There exists $\iota > 0$ such that for all $\theta \in \Theta$ with $|\theta - \theta^*| < \iota$, $V(\theta)$ and $I(\theta)$ are finite.*
- (ii) *$I(\theta^*)$ is nonsingular. It holds that $I(\theta^*) = \frac{\mu_4 - 1}{2} V(\theta^*)$.*
- (iii) *$\mathbb{E} \nabla_\theta \ell(X_i^2, Y_i, \theta^*) = 0$.*

Next we provide results to quantify the difference $\hat{\theta}_{n,\tau_1,\tau_2} - \theta^*$ by a simple linear form uniformly in τ_1, τ_2 . In the following, let $\kappa \in (0, 1)$. We restrict ourselves to the case where $|\tau_1 - \tau_2| \geq \kappa$, i.e.

$$(\tau_1, \tau_2) \in R_\kappa := \{(\tau_1, \tau_2) \in [0, 1]^2 : \tau_1 < \tau_2, |\tau_1 - \tau_2| \geq \kappa\},$$

where $n \cdot \kappa$ then can be understood as the minimum size of a shock period which can be detected. In practice, $\kappa \in (0, 1)$ can be chosen using a cross validation method so that the restriction on R_κ does not affect the estimation accuracy.

Theorem 3.8 (Weak Bahadur Representation). *Let Assumption 3.1 hold. Assume that for some $a > 0$, $\mathbb{E}|\zeta_0|^{4+a} < \infty$. Then for each $\kappa > 0$,*

$$\sup_{(\tau_1, \tau_2) \in R_\kappa} \left| \{\hat{\theta}_{n,\tau_1,\tau_2} - \theta^*\} + ((\tau_2 - \tau_1)V(\theta^*))^{-1} \cdot \nabla_\theta L_{n,\tau_1,\tau_2}(\theta^*) \right| = O_p(\log(n)^3 n^{-1}).$$

With the linearization of our parameter estimation, we further obtain the limit distribution of $\hat{\theta}_{n,\tau_1,\tau_2}$ uniformly in $(\tau_1, \tau_2) \in R_\kappa$ under H_0 by using Gaussian approximation results from [22]. It naturally implies the pointwise convergence results from [8] but is much stronger since it can be used as a starting point to apply theorems (such as the continuous mapping theorem) from the empirical process theory. Let $\ell^\infty(T)$ denote the space of bounded functions $f : T \rightarrow \mathbb{R}$, cf. [19], Section 18, Example 18.5. As a direct consequence of the uniform Bahardur representation, we can derive the distribution of a simple test statistics under the null.

Theorem 3.9 (Asymptotic distribution of $\hat{\theta}_{n,\tau_1,\tau_2}$). *Assume assumption 3.1. Suppose that there exists $a' > 0$ such that $\mathbb{E}|\zeta_0|^{4+a'} < \infty$. Fix $\kappa > 0$. Then on $\ell^\infty(R_\kappa)^{r+s+1}$,*

$$\sqrt{n}\{\hat{\theta}_{n,\tau_1,\tau_2} - \theta^*\} \xrightarrow{d} \Sigma^{1/2}\left\{\frac{B(\tau_2) - B(\tau_1)}{\tau_2 - \tau_1}\right\},$$

where $B(\cdot)$ is a standard $(r + s + 1)$ -dimensional Brownian motion and

$$\Sigma := V(\theta^*)^{-1}I(\theta^*)V(\theta^*)^{-1} = \frac{\mu_4 - 1}{2} \cdot V(\theta^*)^{-1},$$

where $\mu_4 := \mathbb{E}\zeta_0^4$.

With a more general formulation, we obtain the limit distribution of $\hat{B}_{n,H}^{(\chi)}$ with the continuous mapping theorem. We state a slightly more general result by letting $H \in \mathbb{R}^{(r+s+1) \times d}$ which allows to detect for more than one deviation from a 'stable' state.

Corollary 3.10 (Limit distribution of $\hat{B}_{n,H}^{(\chi)}$). *Suppose that Assumption 3.1 holds. Suppose that there exists $a' > 0$ such that $\mathbb{E}|\zeta_0|^{4+a'} < \infty$. Fix $\kappa > 0$. Let $H \in \mathbb{R}^{(r+s+1) \times d}$ be a matrix with full rank. Let $\Sigma_H := H'\Sigma H$. Then*

$$\begin{aligned} \hat{B}_{n,H}^{(\chi)} &= \sqrt{n} \sup_{(\tau_1, \tau_2) \in R_\kappa} (\tau_2 - \tau_1)^\chi \{H'\hat{\theta}_{n,\tau_1,\tau_2} - H'\theta^*\} \\ &\xrightarrow{d} \Sigma_H^{1/2} \sup_{(\tau_1, \tau_2) \in R_\kappa} \left\{ \frac{B(\tau_2) - B(\tau_1)}{(\tau_2 - \tau_1)^{1-\chi}} \right\}, \end{aligned}$$

where $B(\cdot)$ is a standard d -dimensional Brownian motion.

3.3. Estimation of Σ

In this subsection, we discuss how to estimate the variance covariance matrix of our QMLE. which is needed to use the proposed test without prior knowledge of the parameters. We have seen that

$$\begin{aligned} \Sigma &= V(\theta^*)^{-1}I(\theta^*)V(\theta^*)^{-1} \\ &= \frac{\mu_4 - 1}{2} \cdot V(\theta^*)^{-1}. \end{aligned} \tag{9}$$

Here we restrict ourselves to estimation of Σ via the representation (9) to avoid estimating μ_4 separately. To get a test with high power it seems reasonable to estimate Σ under the alternative. Recall that $R_{\kappa,\kappa'} := \{(\tau_1, \tau_2) \in [0, 1]^2 : \tau_1 < \tau_2, 1 - \kappa' \geq \tau_2 - \tau_1 \geq \kappa\}$. Let

$$\bar{L}_{n,\tau_1,\tau_2}^c(\theta) := \frac{1}{n} \sum_{i \in \{1, \dots, n\} \setminus \{[n\tau_1]+1, [n\tau_2]\}} \ell(X_i^2, Y_i^c, \theta)$$

and define the estimator of θ^* in the stationary regime,

$$\bar{\theta}_{n,\tau_1,\tau_2} := \operatorname{argmin}_{(\tau_1,\tau_2) \in R_{\kappa,\kappa'}} \bar{L}_{n,\tau_1,\tau_2}^c(\theta).$$

Now put

$$\begin{aligned} \bar{V}_{n,\tau_1,\tau_2}(\theta) &:= \frac{1}{1 - (\tau_2 - \tau_1)} \nabla_{\theta}^2 \bar{L}_{n,\tau_1,\tau_2}^c(\theta), \\ \bar{I}_{n,\tau_1,\tau_2}(\theta) &:= \frac{1}{n(1 - (\tau_2 - \tau_1))} \sum_{i \in \{1, \dots, n\} \setminus \{[n\tau_1]+1, [n\tau_2]\}} \nabla_{\theta} \ell(X_i^2, Y_i^c, \theta) \nabla_{\theta} \ell(X_i^2, Y_i^c, \theta)'. \end{aligned}$$

Then the following intermediate result holds:

Proposition 3.11 (Estimation of V, I). *Suppose that Assumption 3.1 holds. Suppose that there exists $a' > 0$ such that $\mathbb{E}|\zeta_0|^{4+a'} < \infty$. Fix $\kappa, \kappa' > 0$. Then:*

- (i) $\sup_{(\tau_1,\tau_2) \in R_{\kappa,\kappa'}} |\bar{V}_{n,\tau_1,\tau_2}(\bar{\theta}_{n,\tau_1,\tau_2}) - V(\theta^*)|_1 \xrightarrow{P} 0$,
- (ii) If additionally $\mathbb{E}|\zeta_0|^{8+a'} < \infty$, $\sup_{(\tau_1,\tau_2) \in R_{\kappa,\kappa'}} |\bar{I}_n(\bar{\theta}_{n,\tau_1,\tau_2}) - I(\theta^*)| \xrightarrow{P} 0$.

We propose the estimate

$$\bar{\Sigma}_{n,\tau_1,\tau_2} := \bar{V}_n(\bar{\theta}_{n,\tau_1,\tau_2})^{-1} \bar{I}_n(\bar{\theta}_{n,\tau_1,\tau_2}) \bar{V}_n(\bar{\theta}_{n,\tau_1,\tau_2})^{-1} \quad (10)$$

for Σ and $\bar{\Sigma}_{n,\tau_1,\tau_2,H} := H' \bar{\Sigma}_{n,\tau_1,\tau_2} H$. As a corollary of Theorem (3.10) and Proposition (3.11), we obtain

$$\begin{aligned} \hat{B}_n &:= \sqrt{n} \sup_{(\tau_1,\tau_2) \in R_{\kappa,\kappa'}} (\tau_2 - \tau_1)^{\chi} \bar{\Sigma}_{n,\tau_1,\tau_2,H}^{-1/2} \{H' \hat{\theta}_{n,\tau_1,\tau_2} - H' \theta^*\} \\ &\xrightarrow{d} \sup_{(\tau_1,\tau_2) \in R_{\kappa,\kappa'}} \left\{ \frac{B(\tau_2) - B(\tau_1)}{(\tau_2 - \tau_1)^{1-\chi}} \right\} =: W \end{aligned} \quad (11)$$

with $B(\cdot)$ a d -dimensional Brownian motion. The quantiles of the limit distribution of W can be obtained via simulation.

If significance is detected, τ_1^*, τ_2^* can be estimated by the choice

$$(\hat{\tau}_{1,n}, \hat{\tau}_{2,n}) \in \operatorname{argmax}_{(\tau_1,\tau_2) \in R_{\kappa,\kappa'}} (\tau_2 - \tau_1)^{\chi} \bar{\Sigma}_{n,\tau_1,\tau_2,H}^{-1/2} \{H' \hat{\theta}_{n,\tau_1,\tau_2} - H' \theta^*\},$$

which is motivated by (11). For the special case of one parameter change in a GARCH(1,1) model we will have the following simplified result.

Example 3.12 (Example (3.4) continued). *From Corollary (3.10), we obtain with under $H_0 : \alpha_1^* = 1$:*

$$\begin{aligned} &\sqrt{n} \sup_{(\tau_1,\tau_2) \in R_{\kappa,\kappa'}} (\tau_2 - \tau_1)^{\chi} (\bar{\Sigma}_{n,\tau_1,\tau_2})_{2,2}^{-1/2} \{\hat{\alpha}_{1,n,\tau_1,\tau_2} - 1\} \\ &\xrightarrow{d} \sup_{(\tau_1,\tau_2) \in R_{\kappa,\kappa'}} \left\{ \frac{B(\tau_2) - B(\tau_1)}{(\tau_2 - \tau_1)^{1-\chi}} \right\} \end{aligned}$$

with some 1-dimensional standard Brownian motion $B(\cdot)$.

4. Simulation

In this section, we conduct a simulation study for evaluating the performance of our methodology. The algorithm is summarized in Section 2. We consider the shocked GARCH(1,1)-model with $\theta^*(i) = (\alpha_0^*(i), \alpha_1^*(i), \beta_1^*(i))'$,

$$\theta^*(i) = \begin{cases} \theta^*, & i \leq \lfloor n\tau_1 \rfloor, \\ \tilde{\theta}^*, & \lfloor n\tau_1 \rfloor + 1 \leq i \leq \lfloor n\tau_2 \rfloor, \\ \theta^*, & i > \lfloor n\tau_2 \rfloor, \end{cases}$$

with $H_0 : H'\tilde{\theta}^* = H\theta^*$, and $H_1 : H'\tilde{\theta}^* > H\theta^*$, where

- i) $H = (0, 1, 0)'$ and $\alpha_0^* = 0.3$, $\alpha_1^* = 1.0$, $\beta_1^* = 0.25$ or
- ii) $H = (0, 1, 1)'$ and $\alpha_0^* = 0.3$, $\alpha_1^* = 0.4$, $\beta_1^* = 0.6$,

and

$$X_i^2 = \zeta_i^2 \sigma_i^2, \quad \sigma_i^2 = \alpha_0^*(i) + \alpha_1^*(i)X_{i-1}^2 + \beta_1^*(i)\sigma_{i-1}^2,$$

We now check the behavior of the test under the null hypothesis H_0 of no change. We use the test proposed in Section 2 with $\chi = \frac{1}{2}$, $\kappa = \kappa' = 0.1$ and a grid approximation of $L = 30$.

For $N = 1000$ replications and $n \in \{500, 1000, 2000\}$, $\delta \in \{0.90, 0.95\}$, we obtain the quantiles (cf. Step 4 in the algorithm in Section 2) $\hat{q}_{W,0.90} \approx 3.031$, $\hat{q}_{W,0.95} \approx 3.285$ and the results given in Table 1. We find that the performance of the test is quite unaffected by the choice of χ and therefore do not present analysis for different values of χ here. We can see from the table that as the sample size increases, the coverage probabilities would approach to the nominal level for both $\delta = 0.90, 0.95$. This illustrates a good performance of our test statistics.

TABLE 1
Averaged acceptance rate under null hypothesis H_0 . ($L = 30$).

n / δ	i)		ii)	
	0.90	0.95	0.90	0.95
500	0.866	0.907	0.864	0.903
1000	0.877	0.906	0.884	0.910
2000	0.896	0.918	0.859	0.913

To evaluate the test performance under the alternative, we consider $\delta = 0.95, 0.90$ and the cases

$$H'\tilde{\theta}^* - H\theta^* \in \{0.05, 0.1, 0.2\}$$

with a shock period of $\tau_2^* - \tau_1^* \in \{0.1, 0.2\}$, where we have chosen $\tau_1^* = 0.5$. The choice of τ_1^* does not have a big influence on the performance of the test, therefore we do not present simulation results for different τ_1^* here. The test results in different scenarios can be found in Table 2 and 3. It can be seen that our test shows good power under the alternative hypothesis, which is robust against different choices of break sizes and time length of breaks. We also find that as the sample size increases the power increases drastically.

TABLE 2

Rejection rate of the test (test power) under the alternative H_1 with different Δ^* and $\tau_2^* - \tau_1^*$. ($L = 30$), $\delta = 95\%$

		$\tau_2^* - \tau_1^* = 0.1$			$\tau_2^* - \tau_1^* = 0.2$		
n/Δ^*		$\Delta^* = 0.05$	$\Delta^* = 0.1$	$\Delta^* = 0.2$	$\Delta^* = 0.05$	$\Delta^* = 0.1$	$\Delta^* = 0.2$
i)	500	0.300	0.305	0.828	0.896	0.923	0.942
	1000	0.571	0.591	0.884	0.897	0.908	0.920
	2000	0.801	0.813	0.902	1.000	1.000	1.000
ii)	500	0.344	0.488	0.831	0.901	0.940	0.950
	1000	0.796	0.801	0.865	0.965	0.984	0.997
	2000	0.808	0.824	0.908	1.000	1.000	1.000

TABLE 3

Rejection rate of the test (test power) under the alternative H_1 with different Δ^* and $\tau_2^* - \tau_1^*$. ($L = 30$), $\delta = 90\%$

		$\tau_2^* - \tau_1^* = 0.1$			$\tau_2^* - \tau_1^* = 0.2$		
n/Δ^*		$\Delta^* = 0.05$	$\Delta^* = 0.1$	$\Delta^* = 0.2$	$\Delta^* = 0.05$	$\Delta^* = 0.1$	$\Delta^* = 0.2$
i)	500	0.376	0.498	0.855	0.917	0.936	0.945
	1000	0.689	0.711	0.900	0.903	0.998	0.998
	2000	0.868	0.884	0.922	1.000	1.000	1.000
ii)	500	0.398	0.711	0.866	0.931	0.949	0.960
	1000	0.803	0.832	0.899	0.995	0.999	0.999
	2000	0.945	0.968	0.996	1.000	1.000	1.000

5. Application

We now apply our test to real data. We collect daily historical Apple stock prices (open, high, low, close and adjusted prices) from Jan 4th, 2000 to Nov 12th, 2018 extracted from the Yahoo Finance website, <http://finance.yahoo.com/quote/AAPL/history?p=AAPL>. The data has logged the prices of the Apple stock everyday and comprises of the open, close, low, high and the adjusted close prices of the stock for the span of 18 years. The goal of our analysis is to discover the existence of a period of unstable behavior for the underlying volatility process, and to see whether our test can help with forecasting. Figure 3 shows the plotted adjusted price, log returns and absolute log returns of the Apple stock price. From the plot we observe that the returns fluctuate around the zero. During the years 2000, 2008-2009, 2014 and 2015-16, there are volatility peaks. We divide the data into a sequence of consecutive windows of 1000 observations. The log returns are stationary in all the windows (suggested by the ADF test) and serial correlation is taken out by fitting an ARMA process in advance and the following analysis is done on the residuals.

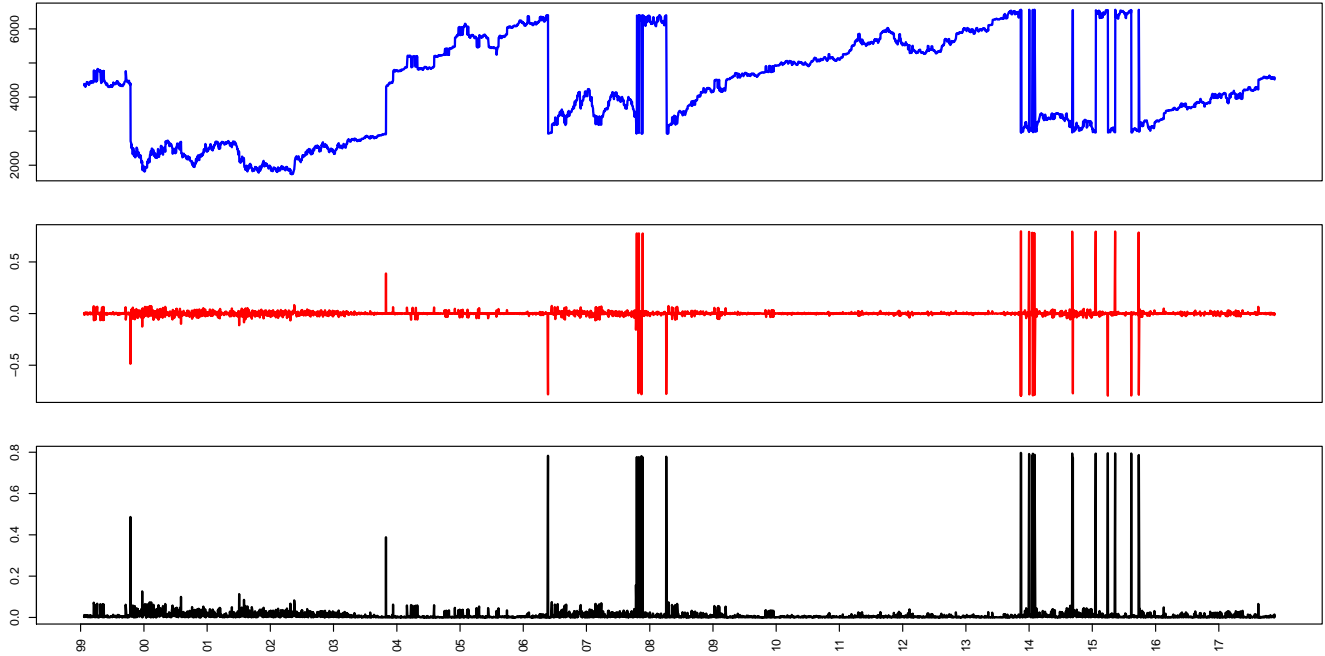
From the histogram and Q-Q plot of the time series in Figure 4 we observe a strong evidence of leptokurtic behavior. In Table 4 we present the detected periods of the explosive behavior. The

TABLE 4

Detected significant break periods for the apple returns, the corresponding persistency parameter ($\hat{\alpha}_1 + \hat{\beta}_1$) and the test statistics. (***) means significant at both 0.95, 0.90, and (**) means significant at only 0.90.

	$\hat{\tau}_1$	$\hat{\tau}_2$	in	out	test statistics
1	2000/01/04	2000/06/27	1.01	0.98	3.23(**)
2	2003/12/30	2004/10/15	1.27	0.98	3.76(***)
3	2008/06/12	2010/01/13	1.31	0.98	100(***)
4	2016/08/17	2018/09/11	1.33	0.88	224(***)

FIGURE 3. Plot of stock price (upper panel) and log returns of stock price of Apple (middle panel), absolute returns (lower panel).



GSRWW test identifies major financial crises such as the technological bubble in 2000 and the US subprime mortgage crisis between Dec. 2007 to June 2009. Furthermore, the test detects some short-lived instability such as the 2003 stock prices downturn.

As a second analysis, we gather the Bitcoin price series ranging from July 23, 2015 to August 21, 2018 at a daily frequency and is presented throughout diagnostic tests. The data source is <https://coinmarketcap.com/currencies/bitcoin/historical-data/>. We show the returns and the absolute returns for the Bitcoin in Figure 5. We can see that there are several high volatility periods. The volatility level is higher before 2013 followed by a stable period. Recently, the market volatility increased. The QQ plots and histograms are given in Figure 6, indicating the heavy-tailedness of the underlying distribution. We present the test results with a window of 1000 observations in Table 5. Again the log returns are stationary in all the windows (by results of an ADF test) and serial correlation is taken out by fitting an ARMA process in advance. We apply our tests to the residuals obtained. The GSRWW test indicates the presence of multiple market ‘euphoria’ episodes in the series. The GSSWW identifies the most significant high volatility period including the period covering the June 2016 crash, the crashes during summer 2017, the fear of market regulation in October 2017, and the massive crash that commenced in December 2017. Bitcoin are not controlled by any government, but speculators can use the test results as indicators of the market sentiment.

6. Conclusion

In this paper, we propose a supreme test for explosive GARCH process. Theoretical results on uniform parameter consistency and asymptotic distribution of the test statistics are provided. Our test is easy to implement and can help to effectively identify explosive periods with good sizes

FIGURE 4. *QQ plot and the histogram for the daily return of the stock prices.*

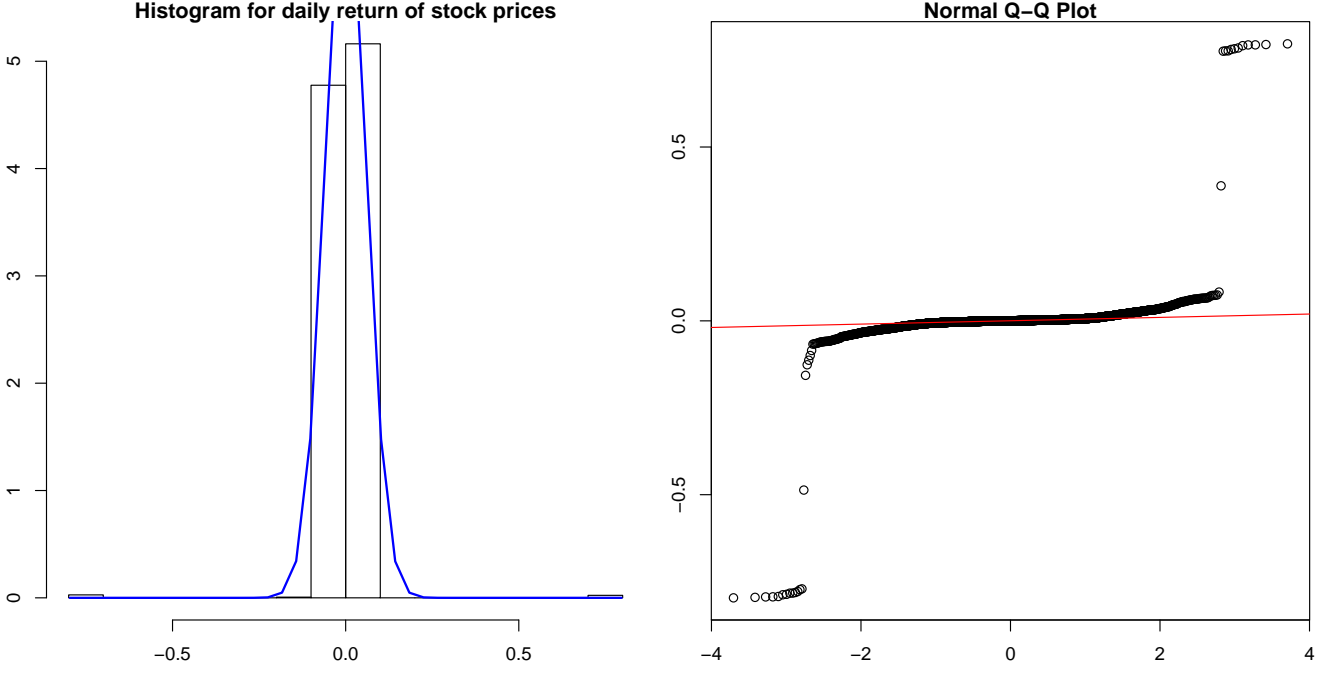


TABLE 5

*Detected significant break periods for the Bitcoin log returns, the corresponding persistency parameter ($\hat{\alpha}_1 + \hat{\beta}_1$) and the test statistics. (***) means significant at both 0.95, 0.90.*

	$\hat{\tau}_1$	$\hat{\tau}_2$	in	out	test statistics
1	2010/12/30	2011/10/15	1.14	1.00	11.68(***)
2	2013/04/14	2013/10/31	1.18	0.99	8.13(***)
3	2016/01/09	2017/12/09	1.05	0.99	19.04(***)

and power via a simulation study. Moreover, we track the volatile period of the Apple stock and Bitcoin returns (log) in our application. In further work one may extend the algorithm to online procedures allowing for real time detection of breaks.

7. Technical proofs and lemmata

7.1. Existence of GARCH models

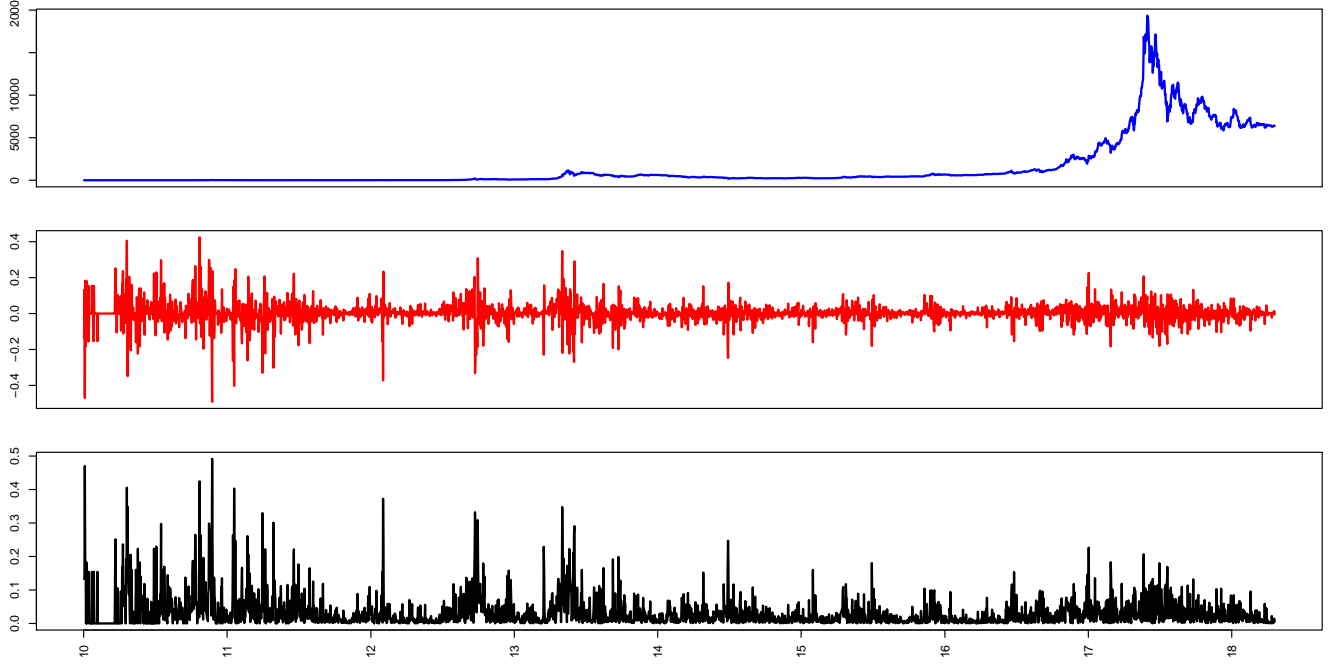
Proof of Proposition 3.2: Following the proof of Lemma 2.3 in [1], there exists $m \in \mathbb{N}$ such that $\mathbb{E} \log |A_m(\theta)A_{m-1}(\theta) \dots A_1(\theta)|_2 < 0$. The function $[0, \delta) \rightarrow [0, \infty)$, $s(t, \theta) = \mathbb{E}|A_m(\theta)A_{m-1}(\theta) \dots A_1(\theta)|_2^t$ fulfills $s'(0, \theta) = \mathbb{E} \log |A_m(\theta)A_{m-1}(\theta) \dots A_1(\theta)|_2 < 0$, thus $t \mapsto s(t, \theta)$ decreases in a neighborhood of 0. Since $s(0) = 1$, this implies that there exists $0 < q < \delta$ such that

$$\mathbb{E}|A_m(\theta)A_{m-1}(\theta) \dots A_1(\theta)|_2^q = s(q, \theta) < 1. \quad (12)$$

Define

$$\begin{aligned} P_i(\theta) &:= (X_i^2, \dots, X_{i-r+1}^2, \sigma_i^2, \dots, \sigma_{i-s+1}^2)', \\ a_i(\theta) &:= (\alpha_0 \zeta_i^2, 0, \dots, 0, \alpha_0, 0, \dots, 0)'. \end{aligned}$$

FIGURE 5. Plot of Bitcoin price (upper panel) and log returns of Bitcoin (middle panel), absolute returns (lower panel).



Following Section 3.1 in [20], the model (5) admits the representation

$$P_i(\theta) = A_i(\theta)P_{i-1}(\theta) + a_i(\theta). \quad (13)$$

Therefore, $P_i(\theta) = G_{\zeta_i, \theta}(P_{i-1})$ with $G_{\zeta_i, \theta}(y) = A_i(\theta) \cdot y + a_i(\theta)$. Let $W_n(y, \theta) := G_{\zeta_n, \theta} \circ G_{\zeta_{n-1}, \theta} \circ \dots \circ G_{\zeta_1, \theta}(y)$. Then we have

$$W_n(y, \theta) - W_n(y', \theta) = A_n(\theta)A_{n-1}(\theta) \cdot \dots \cdot A_1(\theta) \cdot (y - y').$$

Using the submultiplicativity of $\|\cdot\|_2$, we therefore have with (12) and some suitable constant $C > 0$:

$$\begin{aligned} \|W_n(y, \theta) - W_n(y', \theta)\|_2 &\leq \|A_n(\theta)A_{n-1}(\theta) \cdot \dots \cdot A_1(\theta)\|_2 \|y - y'\|_2 \\ &\leq C(s(q, \theta)^{q/m})^n. \end{aligned}$$

By Theorem 2 in [21], we obtain existence and a.s. uniqueness of $X_i^2 = H(t, \mathcal{F}_i)$, $\|X_0^2\|_q < \infty$ and $\delta_q^{X^2}(k) = O(c^k)$ with some $0 < c < 1$, i.e. (i) and (ii). (iii) is due to Proposition 1 in [8]. \square

7.2. Proofs for asymptotic theory

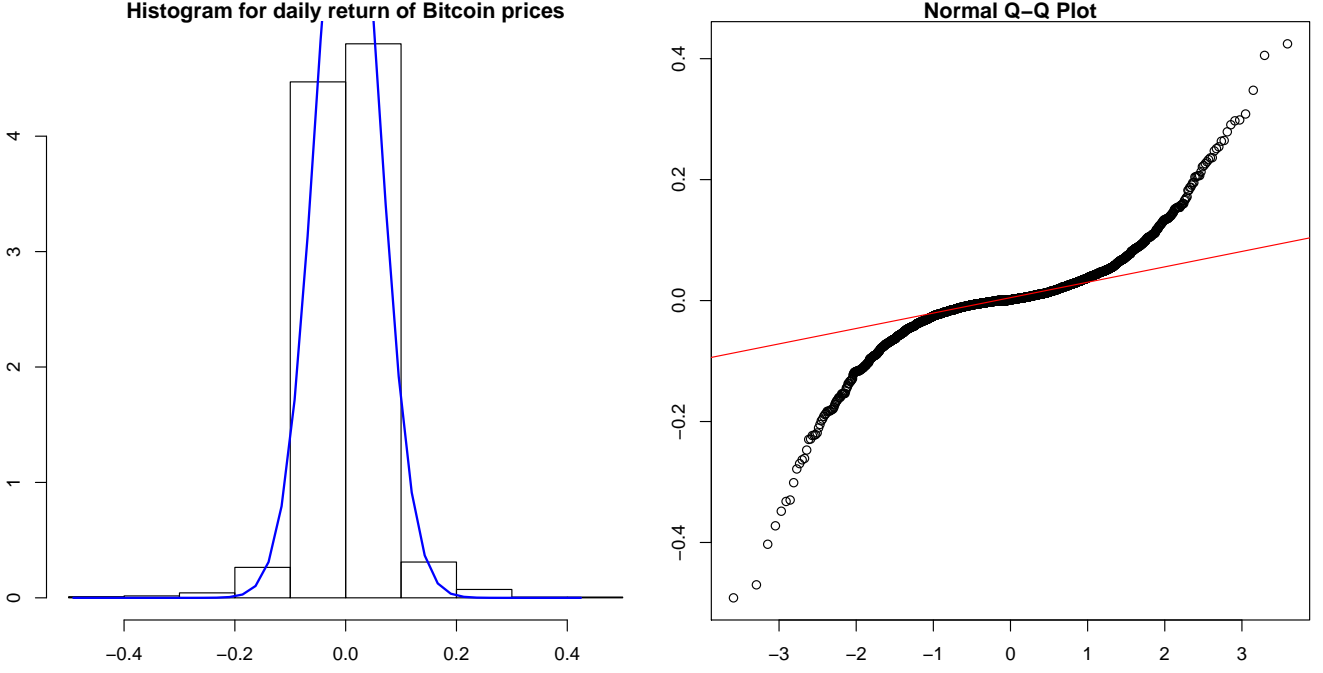
Observe that

$$L_{n, \tau_1, \tau_2}^c(\theta) = L_{n, \tau_2}^c(\theta) - L_{n, \tau_1}^c(\theta), \quad L_{n, r}^c(\theta) := \frac{1}{n} \sum_{i=1}^{\lfloor nr \rfloor} \ell(X_i^2, Y_i^c, \theta).$$

Define $L(\theta) := \mathbb{E}\ell(X_i^2, Y_i, \theta)$. This is well-defined due to $\mathbb{E} \max\{-\ell(X_i^2, Y_i, \theta), 0\} < \infty$ (cf. [8], Proof of Theorem 2.1).

To prove Theorem 3.3, we introduce some notation. For some sequence of real-valued random variables W_n , we write $W_n \xrightarrow{P} \infty$ if for each $M \in \mathbb{N}$, $\mathbb{P}(W_n < M) \rightarrow 0$ ($n \rightarrow \infty$).

FIGURE 6. *QQ plot and the histogram for the Bitcoin returns*



Lemma 7.1. *Let Assumption 3.1 hold. Let $\kappa > 0$, and $R := \{(\tau_1, \tau_2) \in [0, 1]^2 : \tau_1 < \tau_2, |\tau_1 - \tau_2| \geq \kappa\}$. For fixed $k \in \mathbb{N}$, let $V_k(\theta) := \{\theta' \in \Theta : |\theta' - \theta|_1 < 1/k\}$. Define $W_i^{(k)}(\theta) := \inf_{\theta' \in V_k(\theta)} \ell(X_i^2, Y_i, \theta')$. Then:*

(i) $\mathbb{E}W_1^{(k)}(\theta) \in \mathbb{R} \cup \{\infty\}$ and

$$\liminf_{n \rightarrow \infty} \inf_{(\tau_1, \tau_2) \in R} \frac{1}{n(\tau_2 - \tau_1)} \sum_{i=\lfloor n\tau_1 \rfloor + 1}^{\lfloor n\tau_2 \rfloor} W_i^{(k)}(\theta) \geq \mathbb{E}W_1^{(k)}(\theta) \quad a.s.$$

(ii) $L(\theta^*)$ is finite and

$$\sup_{(\tau_1, \tau_2) \in R} \frac{1}{n(\tau_2 - \tau_1)} \sum_{i=\lfloor n\tau_1 \rfloor + 1}^{\lfloor n\tau_2 \rfloor} \ell(X_i^2, Y_i, \theta^*) \rightarrow L(\theta^*) \quad a.s.$$

Proof. (i) Fix some $M \in \mathbb{N}$. It holds that $\mathbb{E}W_1^{(k)} \in \mathbb{R} \cup \{\infty\}$ and $\mathbb{E}[W_1^{(k)} \wedge M] \in \mathbb{R}$ since $\mathbb{E} \max\{-\ell(X_1^2, Y_1, \theta), 0\} < \infty$. Define $S_m := \sum_{i=1}^m W_i^{(k)}(\theta) \wedge M$. By the ergodic theorem, we have

$$\lim_{m \rightarrow \infty} \frac{1}{m} S_m = \mathbb{E}S_1 \quad a.s.$$

It holds that $\lfloor n\tau_2 \rfloor \rightarrow \infty$ uniformly in $\tau_2 \in [\kappa, 1]$, and thus

$$\sup_{\tau_2 \in [\kappa, 1]} \left| \frac{S_{\lfloor n\tau_2 \rfloor}}{\lfloor n\tau_2 \rfloor} - \mathbb{E}S_1 \right| \rightarrow 0 \quad a.s. \quad (14)$$

Furthermore we have that $\frac{S_m}{m}$ is a.s. bounded. We conclude that

$$\sup_{\tau_1 \in [0,1]} \left| \tau_1 \cdot \left(\frac{S_{\lfloor n\tau_1 \rfloor}}{\lfloor n\tau_1 \rfloor} - \mathbb{E}S_1 \right) \right| \rightarrow 0 \quad a.s. \quad (15)$$

[Proof: Fix some $\omega \in \Omega$ and let $C > 0$ be such that $|\frac{S_m(\omega)}{m}| \leq C$ for all $m \in \mathbb{N}$. For $\varepsilon > 0$,

$$\sup_{\tau_1 \in [0,1]} \left| \tau_1 \cdot \left(\frac{S_{\lfloor n\tau_1 \rfloor}}{\lfloor n\tau_1 \rfloor} - \mathbb{E}S_1 \right) \right| \leq \sup_{\tau_1 \in [\frac{\varepsilon}{C}, 1]} \left| \tau_1 \cdot \left(\frac{S_{\lfloor n\tau_1 \rfloor}}{\lfloor n\tau_1 \rfloor} - \mathbb{E}S_1 \right) \right| + \sup_{\tau_1 \in [0, \frac{\varepsilon}{C}]} \left| \tau_1 \cdot \left(\frac{S_{\lfloor n\tau_1 \rfloor}}{\lfloor n\tau_1 \rfloor} - \mathbb{E}S_1 \right) \right|$$

The second term is bounded by $\frac{\varepsilon}{C} \cdot C = \varepsilon$, while for the first term we can choose n large enough such that it is $\leq \varepsilon$.]

With the decomposition

$$\frac{1}{n(\tau_2 - \tau_1)} \sum_{i=\lfloor n\tau_1 \rfloor + 1}^{\lfloor n\tau_2 \rfloor} W_i^{(k)}(\theta) \wedge M = \frac{1}{\tau_2 - \tau_1} \left[\frac{\lfloor n\tau_2 \rfloor}{n} \cdot \frac{S_{\lfloor n\tau_2 \rfloor}}{\lfloor n\tau_2 \rfloor} - \frac{\lfloor n\tau_1 \rfloor}{n\tau_1} \cdot \tau_1 \cdot \frac{S_{\lfloor n\tau_1 \rfloor}}{\lfloor n\tau_1 \rfloor} \right]$$

and (14), (15), $\sup_{\tau_2 \in [\kappa, 1]} \left| \frac{\lfloor n\tau_2 \rfloor}{n} - \tau_2 \right| \leq n^{-1}$ we obtain

$$\inf_{(\tau_1, \tau_2) \in R} \frac{1}{n(\tau_2 - \tau_1)} \sum_{i=\lfloor n\tau_1 \rfloor + 1}^{\lfloor n\tau_2 \rfloor} W_i^{(k)}(\theta) \wedge M \rightarrow \mathbb{E}[W_1^{(k)}(\theta) \wedge M] \quad a.s.$$

Since $W_i^{(k)}(\theta) \geq W_i^{(k)} \wedge M$ and applying $M \rightarrow \infty$ on the r.h.s., we obtain

$$\liminf_{n \rightarrow \infty} \inf_{(\tau_1, \tau_2) \in R} \frac{1}{n(\tau_2 - \tau_1)} \sum_{i=\lfloor n\tau_1 \rfloor + 1}^{\lfloor n\tau_2 \rfloor} W_i^{(k)}(\theta) \geq \mathbb{E}W_1^{(k)}(\theta) \quad a.s.$$

(ii) The argument follows the same lines as (i). We obtain convergence since no truncation with M is needed. \square

Proof of Theorem 3.3. We make use of some results obtained in the proof of Theorem 2.1. in [8]. It was shown therein that $L(\theta) := \mathbb{E}\ell(X_i^2, Y_i, \theta)$ fulfills

$$\mathbb{E}|\ell(X_i^2, Y_i, \theta^*)| < \infty, \quad \forall \theta \neq \theta^* : L(\theta) > L(\theta^*). \quad (16)$$

Let $k \in \mathbb{N}$. Use the notation from Lemma 7.1. Let $\theta \neq \theta^*$. By Beppo-Levi's theorem, we have

$$\mathbb{E}W_1^{(k)}(\theta) \uparrow L(\theta) > L(\theta^*).$$

Thus, for each $\theta \neq \theta^*$ we can find $k(\theta) \in \mathbb{N}$ such that $\mathbb{E}W_1^{(k(\theta))}(\theta) > L(\theta^*)$.

Let $\varepsilon > 0$ and $\Theta_\varepsilon := \{\theta \in \Theta : |\theta - \theta^*| \geq \varepsilon\}$. Then Θ_ε is compact, and there exist finitely many $\theta_1, \dots, \theta_l$ with $\Theta_\varepsilon \subset \bigcup_{i=1}^l V_{k(\theta_i)}(\theta_i)$. Let

$$\delta := \min\left\{ \inf_{i=1, \dots, l} \mathbb{E}W_1^{(k(\theta_i))}(\theta_i) - L(\theta^*), 1 \right\} > 0.$$

Suppose that $\sup_{(\tau_1, \tau_2) \in R} |\hat{\theta}_{n, \tau_1, \tau_2} - \theta^*| \geq \varepsilon$. By the minimal property of $\hat{\theta}_{n, \tau_1, \tau_2}$ and dividing by $\tau_2 - \tau_1$, we conclude that

$$\begin{aligned}
0 &\geq \inf_{(\tau_1, \tau_2) \in R} \frac{1}{\tau_2 - \tau_1} \{L_{n, \tau_1, \tau_2}^c(\hat{\theta}_{n, \tau_1, \tau_2}) - L_{n, \tau_1, \tau_2}^c(\theta^*)\} \\
&= \inf_{i=1, \dots, l} \inf_{(\tau_1, \tau_2) \in R} \inf_{\theta' \in V_{k(\theta_i)}(\theta_i)} \frac{1}{\tau_2 - \tau_1} \{L_{n, \tau_1, \tau_2}^c(\theta') - L_{n, \tau_1, \tau_2}^c(\theta^*)\} \\
&\geq \inf_{i=1, \dots, l} \inf_{(\tau_1, \tau_2) \in R} \inf_{\theta' \in V_{k(\theta_i)}(\theta_i)} \frac{L_{n, \tau_1, \tau_2}^c(\theta')}{\tau_2 - \tau_1} - \sup_{(\tau_1, \tau_2) \in R} \frac{L_{n, \tau_1, \tau_2}^c(\theta^*)}{\tau_2 - \tau_1}
\end{aligned} \tag{17}$$

We furthermore have:

$$\begin{aligned}
&\inf_{(\tau_1, \tau_2) \in R} \inf_{\theta' \in V_{k(\theta_i)}(\theta_i)} \frac{L_{n, \tau_1, \tau_2}^c(\theta')}{\tau_2 - \tau_1} \\
&\geq \inf_{(\tau_1, \tau_2) \in R} \inf_{\theta' \in V_{k(\theta_i)}(\theta_i)} \frac{L_{n, \tau_1, \tau_2}^c(\theta')}{\tau_2 - \tau_1} - \kappa \cdot \sup_{(\tau_1, \tau_2) \in R} \sup_{\theta' \in \Theta} |L_{n, \tau_1, \tau_2}^c(\theta') - L_{n, \tau_1, \tau_2}(\theta')| \\
&\geq \inf_{(\tau_1, \tau_2) \in R} \frac{1}{n(\tau_2 - \tau_1)} \sum_{i=\lfloor n\tau_1 \rfloor}^{\lfloor n\tau_2 \rfloor} W_i^{(k(\theta_i))}(\theta_i) \\
&\quad - \kappa \cdot \sup_{(\tau_1, \tau_2) \in R} \sup_{\theta' \in \Theta} |L_{n, \tau_1, \tau_2}^c(\theta') - L_{n, \tau_1, \tau_2}(\theta')|.
\end{aligned} \tag{18}$$

By Lemma 7.2(i),

$$\mathbb{P}\left(\sup_{(\tau_1, \tau_2) \in R} \sup_{\theta' \in \Theta} |L_{n, \tau_1, \tau_2}^c(\theta') - L_{n, \tau_1, \tau_2}(\theta')| > \frac{\delta}{8}\right) = o(1). \tag{19}$$

By Lemma 7.1(i),

$$\mathbb{P}\left(\inf_{i=1, \dots, l} \inf_{(\tau_1, \tau_2) \in R} \frac{1}{n(\tau_2 - \tau_1)} \sum_{i=\lfloor n\tau_1 \rfloor}^{\lfloor n\tau_2 \rfloor} W_i^{(k(\theta_i))}(\theta_i) \leq L(\theta^*) + \frac{\delta}{2}\right) = o(1). \tag{20}$$

By Lemma 7.1(ii),

$$\mathbb{P}\left(\sup_{(\tau_1, \tau_2) \in R} \frac{L_{n, \tau_1, \tau_2}^c(\theta^*)}{\tau_2 - \tau_1} \geq L(\theta^*) + \frac{\delta}{8}\right) = o(1). \tag{21}$$

Inserting (19), (20) into (18) and afterwards using (21), we have

$$\begin{aligned}
&\mathbb{P}\left(\sup_{(\tau_1, \tau_2) \in R} |\hat{\theta}_{n, \tau_1, \tau_2} - \theta^*| \geq \varepsilon\right) \\
&\leq \mathbb{P}\left(0 \geq \inf_{i=1, \dots, l} \inf_{(\tau_1, \tau_2) \in R} \inf_{\theta' \in V_{k(\theta_i)}(\theta_i)} \frac{L_{n, \tau_1, \tau_2}^c(\theta')}{\tau_2 - \tau_1} - \sup_{(\tau_1, \tau_2) \in R} \frac{L_{n, \tau_1, \tau_2}^c(\theta^*)}{\tau_2 - \tau_1}\right) \\
&\leq \mathbb{P}\left(0 \geq (L(\theta^*) + \frac{\delta}{2}) - \frac{\delta}{8} - (L(\theta^*) + \frac{\delta}{8}) = \frac{\delta}{4}\right) + o(1) = o(1),
\end{aligned}$$

showing the assertion. □

Proof of Theorem 3.8. Let $\kappa > 0$ und define $R := \{(\tau_1, \tau_2) \in [0, 1]^2 : \tau_1 < \tau_2, |\tau_1 - \tau_2| \geq \kappa\}$. By Theorem 3.3, we have

$$\sup_{(\tau_1, \tau_2) \in R} |\hat{\theta}_{n, \tau_1, \tau_2} - \theta^*|_1 \rightarrow 0 \quad a.s. \quad (22)$$

Therefore, $\hat{\theta}_{n, \tau_1, \tau_2} \in \text{int}(\Theta)$ uniformly in $(\tau_1, \tau_2) \in R$ for n large enough. Thus there exists $\bar{\Theta} \subset \text{int}(\Theta)$ with $\hat{\theta}_{n, \tau_1, \tau_2} \in \bar{\Theta}$ (for n large enough, $(\tau_1, \tau_2) \in R$).

By a Taylor expansion, we have

$$\begin{aligned} \hat{\theta}_{n, \tau_1, \tau_2} - \theta^* &= -[\nabla_{\theta}^2 L_{n, \tau_1, \tau_2}^c(\bar{\theta})]^{-1} \nabla_{\theta} L_{n, \tau_1, \tau_2}^c(\theta^*) \\ &= -[(\tau_2 - \tau_1)V(\theta^*) + T_{n, \tau_1, \tau_2}(\bar{\theta}_{n, \tau_1, \tau_2})]^{-1} \nabla_{\theta} L_{n, \tau_1, \tau_2}^c(\theta^*), \end{aligned} \quad (23)$$

where $T_{n, \tau_1, \tau_2}(\bar{\theta}_n) = \nabla_{\theta}^2 L_{n, \tau_1, \tau_2}^c(\bar{\theta}_n) - (\tau_2 - \tau_1)V(\theta^*)$ and $\bar{\theta}_{n, \tau_1, \tau_2} \in \Theta$ with $|\bar{\theta}_{n, \tau_1, \tau_2} - \theta^*|_1 \leq |\hat{\theta}_{n, \tau_1, \tau_2} - \theta^*|_1$. By Lemma 7.2 and Lemma 7.4 and $|\frac{\lfloor n\tau_2 \rfloor - \lfloor n\tau_1 \rfloor}{n} - (\tau_1 - \tau_2)| \leq 2n^{-1}$, we have

$$\sup_{(\tau_1, \tau_2) \in R} |T_{n, \tau_1, \tau_2}(\bar{\theta}_{n, \tau_1, \tau_2})|_1 \leq \sup_{(\tau_1, \tau_2) \in R} |V(\bar{\theta}_{n, \tau_1, \tau_2}) - V(\theta^*)|_1 + O_p(\log(n)^{3/2} n^{-1/2}). \quad (24)$$

By Lemma 7.5(ii) applied to $p = q$ and $\bar{\Theta}$, we obtain $\iota > 0$, $C > 0$ and $\rho \in (0, 1)$ such that for $|\theta - \theta^*|_1 < \iota$,

$$\begin{aligned} |V(\theta) - V(\theta^*)|_1 &\leq C(1 + \| |Y_i|_{(\rho^j)_{j,q}}^q \|_1)(1 + \mathbb{E}\zeta_0^2) \cdot |\theta - \theta^*|_1 \\ &\leq C(1 + \frac{D^q}{1 - \rho})(1 + \mathbb{E}\zeta_0^2) \cdot |\theta - \theta^*|_1 =: \tilde{C} \cdot |\theta - \theta^*|_1. \end{aligned} \quad (25)$$

Since $\mathbb{E}\nabla_{\theta} \ell(Y_i, X_i^2, \theta^*) = 0$ (cf. Proposition 3.7(iii)), Lemma 7.2 and Lemma 7.4, we have

$$\sup_{(\tau_1, \tau_2) \in R} |\nabla_{\theta} L_{n, \tau_1, \tau_2}^c(\theta^*)|_1 = O_p(\log(n)^{3/2} n^{-1/2}). \quad (26)$$

Inserting (22) into (25), we obtain $\sup_{(\tau_1, \tau_2) \in R} |T_{n, \tau_1, \tau_2}(\bar{\theta}_{n, \tau_1, \tau_2})|_2 = o_p(1)$. From (23) and (26) we obtain

$$\sup_{(\tau_1, \tau_2) \in R} |\hat{\theta}_{n, \tau_1, \tau_2} - \theta^*|_1 = O_p(\log(n)^{3/2} n^{-1/2}). \quad (27)$$

By (23), we have

$$\begin{aligned} &|\hat{\theta}_{n, \tau_1, \tau_2} - \theta^* + ((\tau_2 - \tau_1)V(\theta^*))^{-1} \cdot \nabla_{\theta} L_{n, \tau_1, \tau_2}^c(\theta^*)|_2 \\ &\leq |((\tau_2 - \tau_1)V(\theta^*) + T_{n, \tau_1, \tau_2}(\bar{\theta}_{n, \tau_1, \tau_2}))^{-1}|_1 \cdot |T_{n, \tau_1, \tau_2}(\bar{\theta}_{n, \tau_1, \tau_2})|_1 \\ &\quad \times |((\tau_2 - \tau_1)V(\theta^*))^{-1} \nabla_{\theta} L_{n, \tau_1, \tau_2}^c(\theta^*)|_1. \end{aligned} \quad (28)$$

Using (24), (25) and (27), we have $\sup_{(\tau_1, \tau_2) \in R} |T_{n, \tau_1, \tau_2}(\bar{\theta}_{n, \tau_1, \tau_2})|_1 = O_p(\log(n)^{3/2} n^{-1/2})$. Inserting this and (26) into (28), we obtain

$$\sup_{(\tau_1, \tau_2) \in R} |\hat{\theta}_{n, \tau_1, \tau_2} - \theta^* + ((\tau_2 - \tau_1)V(\theta^*))^{-1} \cdot \nabla_{\theta} L_{n, \tau_1, \tau_2}^c(\theta^*)|_1 = O_p(\log(n)^3 n^{-1}).$$

Since $\sup_{(\tau_1, \tau_2) \in R} |\nabla_{\theta} L_{n, \tau_1, \tau_2}^c(\theta^*) - \nabla_{\theta} L_{n, \tau_1, \tau_2}(\theta^*)|_1 = O_p(n^{-1})$ by Lemma 7.2, the proof is finished. \square

Proof of Theorem 3.9. Let $R := \{(\tau_1, \tau_2) \in [0, 1]^2 : \tau_1 < \tau_2, |\tau_1 - \tau_2| \geq \kappa\}$. By Lemma 7.3 (applied with $M = 2 + \frac{a'}{4}$, $a = \frac{a'}{4}$), we obtain $C > 0$, $\rho \in (0, 1)$ and $\iota > 0$ such that

$$\delta_M^{\nabla_{\theta} \ell(Z, \theta^*)}(k) \leq C \rho^k,$$

and thus (component-wise) $\Delta_M^{\nabla_{\theta} \ell(Z, \theta^*)}(k) \leq \frac{C}{1-\rho} \rho^k$. Let $W_i := -V(\theta^*)^{-1} \nabla_{\theta} \ell(X_i^2, Y_i, \theta^*)$ and $S(j) := \sum_{i=1}^j W_i$. By Proposition 3.7(iii), $\mathbb{E}W_i = 0$ and

$$\Sigma := \text{Cov}(W_i) = V(\theta^*)^{-1} I(\theta^*) V(\theta^*)^{-1} = \frac{\mu_4 - 1}{2} V(\theta^*)^{-1}.$$

By Corollary 1 in [22], there exists a richer probability space and i.i.d. $V_1, V_2, \dots \sim N(0, I_{(r+s+1) \times (r+s+1)})$, a process $(\hat{S}(i))_{i=1, \dots, n}$ and $S^0(i) = \sum_{j=1}^i V_j$ such that $(S(i))_{i=1, \dots, n} \stackrel{d}{=} (\hat{S}(i))_{i=1, \dots, n}$ and

$$\max_{i=1, \dots, n} |\hat{S}(i) - \Sigma^{1/2} S^0(i)| = O_p(n^{1/\min\{M, 4\}} \log(n)^{3/2}).$$

With Theorem 3.8 we obtain:

$$\sup_{(\tau_1, \tau_2) \in R} \left| \sqrt{n}(\tau_2 - \tau_1)(\hat{\theta}_{n, \tau_1, \tau_2} - \theta^*) - n^{-1/2}(S(\lfloor n\tau_2 \rfloor) - S(\lfloor n\tau_1 \rfloor)) \right| = O_p(\log(n)^3 n^{-1/2}). \quad (29)$$

By the Gaussian approximation result above, on $D(R)^{r+s+1}$

$$n^{-1/2}(S(\lfloor n\tau_2 \rfloor) - S(\lfloor n\tau_1 \rfloor)) \stackrel{d}{=} n^{-1/2}(\hat{S}(\lfloor n\tau_2 \rfloor) - \hat{S}(\lfloor n\tau_1 \rfloor)) \quad (30)$$

and

$$\begin{aligned} & \sup_{(r_1, r_2) \in R} \left| n^{-1/2}(\hat{S}(\lfloor n\tau_2 \rfloor) - \hat{S}(\lfloor n\tau_1 \rfloor)) - \Sigma^{1/2} \cdot n^{-1/2}(S^0(\lfloor n\tau_2 \rfloor) - S^0(\lfloor n\tau_1 \rfloor)) \right| \\ &= O_p(n^{\frac{1}{\min\{M, 4\}} - \frac{1}{2}} \log(n)^{3/2}). \end{aligned} \quad (31)$$

By Donsker's theorem, it holds in $D[0, 1]^{r+s+1}$ that $n^{-1/2} S^0(\lfloor nr \rfloor) \xrightarrow{d} B(r)$ with some standard $(r+s+1)$ -dimensional Brownian motion B . Applying the continuous mapping theorem, we obtain in $D(R)^{r+s+1}$:

$$n^{-1/2} \Sigma^{1/2} \{S^0(\lfloor n\tau_2 \rfloor) - S^0(\lfloor n\tau_1 \rfloor)\} \xrightarrow{d} \Sigma^{1/2} \{B(\tau_2) - B(\tau_1)\}. \quad (32)$$

Combining (29), (30), (31) and (32), we obtain the result. \square

Proof of Proposition 3.11. Using similar arguments as in the proof of Theorem 3.3 (now with $1 - (\tau_2 - \tau_1) \geq \kappa$ instead of $\tau_2 - \tau_1 \geq \kappa$), we obtain

$$\sup_{(\tau_1, \tau_2) \in \bar{R}_\kappa} |\bar{\theta}_{n, \tau_1, \tau_2} - \theta^*| \xrightarrow{p} 0. \quad (33)$$

(i) By Lemma 7.2 and Lemma 7.4 and $|\frac{1 - (\lfloor n\tau_2 \rfloor - \lfloor n\tau_1 \rfloor)}{1 - (\tau_2 - \tau_1)} - 1| \leq n^{-1}$, we have

$$\sup_{(\tau_1, \tau_2) \in \bar{R}_\kappa} |\bar{V}_{n, \tau_1, \tau_2}(\theta) - (1 - (\tau_2 - \tau_1))V(\theta)|_1 \xrightarrow{p} 0.$$

By Lipschitz continuity of $V(\cdot)$ (cf. (25)) and (33), we obtain the result.

(ii) Define $g(x, y, \theta) := \nabla_{\theta} \ell(x, y, \theta) \cdot \nabla_{\theta} \ell(x, y, \theta)'$ and $\tilde{g}_{\tilde{\theta}}(\zeta, y, \theta) := g(R_{\zeta}(y, \tilde{\theta}), y, \theta)$, where $R_{\zeta}(y, \theta) := \zeta^2 \sigma(y, \theta)^2$. Let $\bar{\Theta} \subset \text{int}(\Theta)$ be some compact set. Using Lemma 7.5(ii) and (52), it is easy to see that for any $p > 0$, one can find $\iota > 0$, $C > 0$ and $\rho \in (0, 1)$ such that (component-wise),

$$\begin{aligned} & \sup_{\theta, \tilde{\theta} \in \bar{\Theta}, |\theta - \tilde{\theta}|_1 < \iota} |\tilde{g}_{\tilde{\theta}}(\zeta, y, \theta) - \tilde{g}_{\tilde{\theta}}(\zeta, y', \theta)| \\ & \leq C(1 + |y|_{(\rho^j)_{j, 2p}}^{2p} + |y'|_{(\rho^j)_{j, 2p}}^{2p}) |y - y'|_{(\rho^j)_{j, p}}^p (1 + \zeta^2)^2. \end{aligned} \quad (34)$$

and

$$\sup_{\theta, \theta', \tilde{\theta} \in \bar{\Theta}, |\theta - \tilde{\theta}|_1 < \iota, |\theta' - \tilde{\theta}|_1 < \iota} \frac{|\tilde{g}_{\tilde{\theta}}(\zeta, y, \theta) - \tilde{g}_{\tilde{\theta}}(\zeta, y, \theta')|}{|\theta - \theta'|_1} \leq C(1 + |y|_{(\rho^j)_{j, p}}^p)(1 + \zeta^2)^2. \quad (35)$$

In the following we will enlarge C, ρ and reduce ι if necessary without further notice. Note that $\sup_{\theta \in \Theta} |\nabla_{\theta}(\sigma(0, \theta)^2)| < \infty$ and thus (component-wise) $\sup_{\theta \in \Theta} |\nabla_{\theta} \ell(x, 0, \theta)| \leq C(1 + |x|)$. With Lemma 7.5(iii) we conclude that (component-wise) $\sup_{\theta \in \Theta} |\nabla_{\theta} \ell(x, y, \theta)| \leq C(1 + |y|_{(\rho^j)_{j, 1}}^2)(1 + |x|) \cdot |y|_{(\rho^j)_{j, 1}}$. Using again Lemma 7.5(iii), we obtain

$$\sup_{\theta \in \Theta} |g(x, y, \theta) - g(x, y', \theta)| \leq C(1 + |y|_{(\rho^j)_{j, 1}}^5 + |y'|_{(\rho^j)_{j, 1}}^5)(1 + |x|)^2 \cdot |y - y'|_{(\rho^j)_{j, 1}}.$$

This shows that g has similar properties as $\nabla_{\theta}^2 \ell$, but with factors $(1 + \zeta^2)^2$ in (34), (35) instead of $(1 + \zeta^2)$. Therefore, we obtain the same result as in (i) under the stated moment condition. \square

7.3. Technical lemmata

Proof of Proposition 3.7. (i) By Proposition 3.2, there exists $q > 0$ with $\|X_0^2\|_q < \infty$. From the bounds (52) (applied with $p = q$) we conclude that $V(\theta), I(\theta)$ are finite as long as $|\theta - \theta^*|$ is small enough.

(ii),(iii) This was already shown in [8], see the proof step (ii) of Theorem 2.2 (the missing $\frac{1}{2}$ is due to the different formulation of the likelihood). \square

Lemma 7.2 (Negligibility of truncation). *Let Assumption 3.1 hold. Then for $g = \nabla_{\theta}^l \ell$, $l = 0, 1, 2$ it holds that*

(i)

$$\sup_{r \in [0, 1]} \sup_{\theta \in \Theta} \left| \sum_{i=1}^{\lfloor nr \rfloor} g(X_i^2, Y_i^c, \theta) - \sum_{i=1}^{\lfloor nr \rfloor} g(X_i^2, Y_i, \theta) \right| = O_p(1).$$

(ii)

$$\sup_{r \in [0, 1]} \sup_{\theta \in \Theta} \frac{1}{n} \left| \sum_{i=1}^{\lfloor nr \rfloor} g(X_i^2, Y_i^c, \theta) - \sum_{i=1}^{\lfloor nr \rfloor} g(X_i^2, Y_i, \theta) \right| = 0 \quad a.s.$$

Proof of Lemma 7.2. Note that for arbitrary $0 < \tilde{q} \leq \min\{q, 1\}$ and random variables $Z_i = (Z_{i1}, Z_{i2}, \dots)$ with $\|Z_i\|_q \leq D$ it holds that

$$\| |Z_i|_{(\rho^j)_{j, 1}} \|_{\tilde{q}} \leq \left(\sum_{j=1}^{\infty} \rho^{\tilde{q}j} \|Z_{ij}\|_{\tilde{q}}^{\tilde{q}} \right)^{1/\tilde{q}} \leq D \left(\frac{1}{1 - \rho^{\tilde{q}}} \right)^{1/\tilde{q}} =: \tilde{D}(\tilde{q}).$$

Let

$$W_i := \sup_{\theta \in \Theta} |g(X_i^2, Y_i^c, \theta) - g(X_i^2, Y_i, \theta)|.$$

By Lemma 7.5(iii), we have with Hölder's inequality for $0 < q' \leq q$ chosen such that $0 < q'(l+3) \leq 1$:

$$\begin{aligned} & \|W_i\|_{q'} \\ &= \left\| \sup_{\theta \in \Theta} |g(X_i^2, Y_i^c, \theta) - g(X_i^2, Y_i, \theta)| \right\|_{q'} \\ &\leq C(1 + \| |Y_i|_{(\rho^j)_{j,1}} \|_{q'(l+3)}^{l+1} + \| |Y_i^c|_{(\rho^j)_{j,1}} \|_{q'(l+3)}^{l+1}) \cdot (1 + \|X_i^2\|_{q'(l+3)}) \| |y - y'|_{(\rho^j)_{j,1}} \|_{q'(l+3)} \\ &\leq C(1 + 2\tilde{D}(q')^{l+1})(1 + D) \cdot \left(\sum_{j=i}^{\infty} \rho^{q'(l+3)j} \|X_j^2\|_{q'(l+3)}^{q'(l+3)} \right)^{1/(q'(l+3))} \\ &\leq C(1 + 2\tilde{D}(q')^{l+1})(1 + D)\tilde{D}(q'(l+3))\rho^i =: \tilde{C} \cdot \rho^i. \end{aligned} \tag{36}$$

Therefore, we have

$$\begin{aligned} \left\| \sup_{r \in [0,1]} \sup_{\theta \in \Theta} \left| \sum_{i=1}^{\lfloor nr \rfloor} g(X_i^2, Y_i^c, \theta) - \sum_{i=1}^{\lfloor nr \rfloor} g(X_i^2, Y_i, \theta) \right| \right\|_{q'} &\leq \left\| \sum_{i=1}^n W_i \right\|_{q'} \\ &\leq \left(\sum_{i=1}^n \|W_i\|_{q'}^{q'} \right)^{1/q'} \\ &\leq \tilde{C} \left(\sum_{i=1}^n (\rho^{q'})^i \right)^{1/q'} < \infty, \end{aligned}$$

giving the result.

(ii) It holds that

$$\sup_{r \in [0,1]} \sup_{\theta \in \Theta} \frac{1}{n} \left| \sum_{i=1}^{\lfloor nr \rfloor} g(X_i^2, Y_i^c, \theta) - \sum_{i=1}^{\lfloor nr \rfloor} g(X_i^2, Y_i, \theta) \right| \leq \frac{1}{n} \sum_{i=1}^n W_i.$$

In the following we show that $W_i \rightarrow 0$ ($i \rightarrow \infty$) a.s.. Then the assertion follows with a Cesaro sum argument. Let $\varepsilon > 0$ be arbitrary. Then with Markov's inequality and (36),

$$\sum_{i=1}^{\infty} \mathbb{P}(|W_i| > \varepsilon) \leq \sum_{i=1}^{\infty} \frac{\tilde{C}^{q'}}{\varepsilon^{q'}} \rho^{q'i} < \infty,$$

showing $W_i \rightarrow 0$ with Borel-Cantelli's lemma. □

Let us use the abbreviation $Z_i := (X_i^2, Y_i)$. We now state results about the dependence measure of the stationary processes $g(Z_i, \theta)$, where $g \in \{\nabla_{\theta} \ell, \nabla_{\theta}^2 \ell\}$.

Lemma 7.3 (Dependence measures of $\nabla_{\theta} \ell, \nabla_{\theta}^2 \ell$). *Let Assumption 3.1 hold. Let $M \geq 1$. Assume that $\mathbb{E}|\zeta_0|^{2(M+a)} < \infty$ for some $a > 0$. Let $g \in \{\nabla_{\theta} \ell, \nabla_{\theta}^2 \ell\}$. Then there exists some $C > 0, \rho \in (0, 1), \iota > 0$ such that*

$$\sup_{|\theta - \theta^*|_1 < \iota} \delta_M^{g(Z, \theta)}(k) \leq C\rho^k, \quad \delta_M^{\sup_{|\theta - \theta^*|_1 < \iota} g(Z, \theta)}(k) \leq C\rho^k.$$

Proof of Lemma 7.3. We only prove the second assertion, the first is nearly the same. Let $(X_i^2)^* = H(\mathcal{F}_i^*)$ and $Z_i^* := ((X_i^2)^*, Y_i^*)$. Let $\kappa = \frac{a}{3}$. Choose $p > 0$ small enough such that $\frac{(M+3\kappa)}{\kappa}p \leq q$. By Hoelder's inequality ($\frac{M}{M+3\kappa} + \frac{\kappa}{M+3\kappa} + \frac{2\kappa}{M+3\kappa} = 1$) and Lemma 7.5(ii) there exists $\iota > 0$, $C > 0$, $\rho \in (0, 1)$ such that

$$\begin{aligned}
& \delta_M^{\sup_{|\theta-\theta^*|_1 < \iota} |g(Z, \theta)|} (i) \\
&= \left\| \sup_{|\theta-\theta^*|_1 < \iota} |g(Z_i, \theta)| - \sup_{|\theta-\theta^*|_1 < \iota} |g(Z_i^*, \theta)| \right\|_M \\
&\leq \left\| \sup_{|\theta-\theta^*|_1 < \iota} |g(Z_i, \theta) - g(Z_i^*, \theta)| \right\|_M \\
&= \left\| \sup_{|\theta-\theta^*|_1 < \iota} |\tilde{g}_{\theta^*}(\zeta_i, Y_i, \theta) - \tilde{g}_{\theta^*}(\zeta_i, Y_i^*, \theta)| \right\|_M \\
&\leq C(1 + \| |Y_i|_{(\rho^j)_{j, 2p}}^{2p} \|_{\frac{M+3\kappa}{2\kappa}} + \| |Y_i^*|_{(\rho^j)_{j, 2p}}^{2p} \|_{\frac{M+3\kappa}{2\kappa}}) \| |Y_i - Y_i^*|_{(\rho^j)_{j, p}}^p \|_{\frac{M+3\kappa}{\kappa}} (1 + \|\zeta^2\|_{M+3\kappa}) \\
&\leq C(1 + 2\frac{D^{2p}}{1-\rho})(1 + \|\zeta^2\|_{M+3\kappa}) \cdot \sum_{j=1}^{\infty} \rho^j \|X_{i-j}^2 - (X_{i-j}^*)^2\|_{\frac{(M+3\kappa)p}{\kappa}}^p \\
&\leq \tilde{C} \cdot \sum_{j=1}^i \rho^j [\delta_q^{X^2}(i-j)]^p,
\end{aligned}$$

where $\tilde{C} := C(1 + 2\frac{D^{2p}}{1-\rho})(1 + \|\zeta^2\|_{M+3\kappa})$. By Proposition 3.2(ii), it holds that $\delta_q^{X^2}(k) = O(c^k)$, which finishes the proof. \square

In the following we make use of results from [23]. Therefore we have to define $\Delta_q^Z(m) := \sum_{k=m}^{\infty} \delta_q^Z(k)$ and $\|Z\|_{q, \alpha} := \sup_{m \geq 0} (m+1)^\alpha \Delta_q^Z(m)$.

Lemma 7.4. *Let Assumption 3.1 hold. Additionally, assume that for some $a' > 0$, $\mathbb{E}|\zeta_0|^{4+a'} < \infty$. Then there exists $\iota > 0$ such that for $g = \nabla_{\theta}^l \ell$, $l = 1, 2$, it holds that*

$$\sup_{|\theta-\theta^*|_1 < \iota} \sup_{r \in [0, 1]} \left| \frac{1}{n} \sum_{i=1}^{\lfloor nr \rfloor} \{g(X_i^2, Y_i, \theta) - \mathbb{E}g(X_i^2, Y_i, \theta)\} \right| = O_p\left(\left(\frac{\log(n)^3}{n}\right)^{1/2}\right).$$

Proof of Lemma 7.4. Let $\iota > 0$ (is chosen below). Let $S_j(\theta) := \sum_{i=1}^j \{g(X_i^2, Y_i, \theta) - \mathbb{E}g(X_i^2, Y_i, \theta)\}$, $j = 1, \dots, n$. For fixed $n \in \mathbb{N}$, choose $d \in \mathbb{N}$ such that $2^{d-1} \leq n \leq 2^d$. For $i = 0, 1, \dots, d-1$, define

$$\Phi_i(\theta) := \max_{1 \leq k \leq 2^{d-i}} |S_{2^i \cdot k}(\theta) - S_{2^i(k-1)}|.$$

By a dyadic expansion of $j \in \{1, \dots, n\}$ we obtain

$$\max_{j=0, \dots, n} |S_j(\theta)| \leq \sum_{i=0}^{d-1} \Phi_i(\theta).$$

Note that

$$\begin{aligned}
& \sup_{|\theta-\theta^*|_1 < \iota} \sup_{r \in [0, 1]} \left| \sum_{i=1}^{\lfloor nr \rfloor} \{g(X_i^2, Y_i, \theta) - \mathbb{E}g(X_i^2, Y_i, \theta)\} \right| \\
&\leq \sum_{i=0}^{d-1} \sup_{|\theta-\theta^*|_1 < \iota} \Phi_i(\theta).
\end{aligned}$$

Thus, for $Q > 0$, by stationarity,

$$\begin{aligned}
& \mathbb{P}\left(\sup_{|\theta-\theta^*|_1 < \iota} \sup_{r \in [0,1]} \left| \sum_{i=1}^{\lfloor nr \rfloor} \{g(X_i^2, Y_i, \theta) - \mathbb{E}g(X_i^2, Y_i, \theta)\} \right| > Q(n \log(n)^3)^{1/2}\right) \\
& \leq \sum_{i=0}^{d-1} \mathbb{P}\left(\sup_{|\theta-\theta^*|_1 < \iota} \Phi_i(\theta) > \frac{Q(n \log(n)^3)^{1/2}}{d}\right) \\
& \leq \sum_{i=0}^{d-1} 2^{d-i} \cdot \mathbb{P}\left(\sup_{|\theta-\theta^*|_1 < \iota} |S_{2^i}(\theta)| > \frac{Q(n \log(n)^3)^{1/2}}{d}\right). \tag{37}
\end{aligned}$$

Since $\theta^* \in \text{int}(\Theta)$, there exists $\iota_1 > 0$ such that $\bar{\Theta} := \{\theta \in \Theta : |\theta - \theta^*|_1 \leq \iota_1\} \subset \text{int}(\Theta)$.

Apply Lemma 7.5(ii) with $p = q$ and Lemma 7.3 applied to $M = 2 + \frac{a'}{4}$, $a = \frac{a'}{4}$, we obtain corresponding $C > 0$, $\rho \in (0, 1)$, $0 < \iota < \iota_1$ such that the statements of the Lemmata hold true.

We now use a simple chaining argument. Let Θ_n be a discretization of $\Theta \subset \mathbb{R}^{r+s+1}$ such that for each $\theta \in \Theta$ there exists some $\theta' \in \Theta_n$ with $|\theta - \theta'|_1 \leq n^{-1}$.

We conclude that for $1 \leq m \leq n$, it holds that

$$\begin{aligned}
& \mathbb{P}\left(\sup_{|\theta-\theta^*|_1 < \iota} |S_m(\theta)| > \frac{Q(n \log(n)^3)^{1/2}}{d}\right) \\
& \leq \mathbb{P}\left(\sup_{\theta \in \Theta_n, |\theta-\theta^*|_1 < \iota} |S_m(\theta)| > \frac{Q(n \log(n)^3)^{1/2}}{2d}\right) \\
& \quad + \mathbb{P}\left(\sup_{\theta, \theta' \in \Theta, |\theta-\theta'|_1 \leq n^{-1}, |\theta-\theta^*|_1 < \iota, |\theta'-\theta^*|_1 < \iota} |S_m(\theta) - S_m(\theta')| > \frac{Q(n \log(n)^3)^{1/2}}{2d}\right) \\
& =: I_n + II_n. \tag{38}
\end{aligned}$$

By Lemma 7.3 applied to $M = 2 + \frac{a'}{4}$, $a = \frac{a'}{4}$, we have $\Delta_M^{\sup_{|\theta-\theta^*|_1 < \iota} |g(Z, \theta)|}(k) = O(\rho^k)$ and $\sup_{|\theta-\theta^*|_1 < \iota} \Delta_M^{g(Z, \theta)}(k) = O(\rho^k)$. Let $\alpha = \frac{1}{2}$. Then

$$W_{M, \alpha} := \left\| \sup_{|\theta-\theta^*|_1 < \iota} |g(Z, \theta)| \right\|_{M, \alpha} = \sup_{m \geq 0} (m+1)^\alpha \Delta_M^{\sup_{|\theta-\theta^*|_1 < \iota} |g(Z, \theta)|}(m) < \infty,$$

and

$$W_{2, \alpha} := \sup_{|\theta-\theta^*|_1 < \iota} \|g(Z, \theta)\|_{2, \alpha} = \sup_{m \geq 0} (m+1)^\alpha \sup_{|\theta-\theta^*|_1 < \iota} \Delta_2^{g(Z, \theta)}(m) < \infty.$$

Note that $l = 1 \wedge \log \#\Theta_n \leq (r+s+1) \log(n)$ and $Qn^{1/2} \log(n)^{3/2} \geq \sqrt{ml} W_{2, \alpha} + m^{1/M} l^{3/2} W_{M, \alpha} \gtrsim m^{1/2} \log(m)^{1/2} + m^{1/M} \log(m)^{3/2}$ for Q large enough.

By applying Theorem 6.2 of [23] with $q = M$ to $(g(Z_i, \theta))_{\theta \in \bar{\Theta}_n, |\theta-\theta^*|_1 < \iota}$, we have with some constants $C_\alpha > 0$:

$$\begin{aligned}
I_n & = \mathbb{P}\left(\sup_{\theta \in \Theta_n, |\theta-\theta^*|_1 < \iota} |S_m(\theta)| > \frac{Q(n \log(n)^3)^{1/2}}{2d}\right) \\
& \leq \frac{C_\alpha m \cdot l^{M/2} W_{M, \alpha}^M}{(Q/2d)^M (n^{1/2} \log(n)^{3/2})^M} + C_\alpha \exp\left(-\frac{C_\alpha (Q/2d)^2 n \log(n)^3}{m W_{2, \alpha}^2}\right) \\
& \leq O(m \cdot n^{-\frac{M}{2}} + n^{-2}), \tag{39}
\end{aligned}$$

for Q large enough, since $d \leq \log_2(n) + 1$ and $m \leq n$.

Since $g(Z_i, \theta) = \tilde{g}_{\theta^*}(\zeta_i, Y_i, \theta)$ and $g(Z_i, \theta') = \tilde{g}_{\theta^*}(\zeta_i, Y_i, \theta')$, we have with Lemma 7.5(ii):

$$\begin{aligned} & \sup_{\theta, \theta' \in \tilde{\Theta}, |\theta - \theta'|_1 \leq n^{-1}, |\theta - \theta^*|_1 < \iota, |\theta' - \theta^*|_1 < \iota} |g(Z_i, \theta) - g(Z_i, \theta')| \\ & \leq C(1 + |Y_i|_{(\rho^j)_{j,p}}^p)(1 + \zeta_i^2)n^{-1}. \end{aligned}$$

Thus

$$\begin{aligned} & \left\| \sup_{\theta, \theta' \in \Theta, |\theta - \theta'|_1 \leq n^{-1}, |\theta - \theta^*|_1 < \iota, |\theta' - \theta^*|_1 < \iota} \left\{ \mathbb{E}_0 g(Z_i, \theta) - \mathbb{E}_0 g(Z_i, \theta') \right\} \right\|_1 \\ & \leq 2C(1 + \| |Y_0|_{(\rho^j)_{j,p}}^p \|_1)(1 + \mathbb{E}\zeta_0^2)\frac{m}{n} \\ & \leq 2C\left(1 + \frac{D^p}{1 - \rho}\right)(1 + \mathbb{E}\zeta_0^2)\frac{m}{n} = O\left(\frac{m}{n}\right). \end{aligned}$$

With Markov's inequality, we therefore obtain

$$II_n \leq \frac{2\tilde{C}m}{(Q/2d)n^{3/2} \log(n)^{3/2}}. \quad (40)$$

Inserting (39) and (40) into (38) and then into (37), we obtain with some constant $\tilde{C} > 0$:

$$\begin{aligned} & \mathbb{P}\left(\sup_{|\theta - \theta^*|_1 < \iota} \sup_{r \in [0,1]} \left| \sum_{i=1}^{\lfloor nr \rfloor} \{g(X_i^2, Y_i, \theta) - \mathbb{E}g(X_i^2, Y_i, \theta)\} \right| > Q(n \log(n)^3)^{1/2} \right) \\ & \leq \tilde{C} \sum_{i=0}^{d-1} 2^{d-i} \cdot \left(2^i \cdot n^{-M/2} + n^{-2} + 2^i \cdot n^{-3/2} \log(n)^{1/2} \right) \\ & \leq \tilde{C}dn \cdot \left(n^{-M/2} + n^{-2} + n^{-3/2} \log(n)^{1/2} \right) \rightarrow 0, \end{aligned}$$

showing the assertion. \square

7.4. Analytical properties of the likelihood

For the following results, we derive some analytical properties of the likelihood we use. This allows us to separate analytical and stochastic treatment. For $p > 0$, some sequence $(y_j)_{j \in \mathbb{N}}$ of real numbers and some sequence $(\chi_j)_{j \in \mathbb{N}}$ of nonnegative real numbers, define the weighted seminorm

$$|y|_{\chi,p} := \left(\sum_{j=1}^{\infty} \chi_j |y_j|^p \right)^{1/p}.$$

Later, we will plug in $x = X_i$ and $y = Y_i$ into $\ell(x, y, \theta)$ and its derivatives. To make use of all connections between x, y , define $R_\zeta(y, \theta) := \zeta^2 \sigma(y, \theta)^2$, and

$$\tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta) := \ell(R_\zeta(y, \tilde{\theta}), y, \theta).$$

In the following Lemma 7.5(ii), we collect some analytical properties of $\tilde{\ell}_{\tilde{\theta}}$ to calculate functional dependence measures of $\ell(X_i^2, Y_i, \theta)$. The bounds in (iii) will be used to show that the truncated likelihood $\ell(X_i^2, Y_i^c, \theta)$ is near to $\ell(X_i^2, Y_i, \theta)$; for this argument we cannot use the connection between X_i^2 and Y_i .

Lemma 7.5. $\theta \mapsto \sigma(y, \theta)$ and $\theta \mapsto \ell(x, y, \theta)$ are three times continuously differentiable. Let $\bar{\Theta} \subset \text{int}(\Theta)$ be a compact subset. Then for any $p > 0$, there exists $\iota > 0$ and $C > 0$, $\rho \in (0, 1)$ such that (component-wise),

(i) for $l = 0, 1, 2, 3$:

$$\sup_{\theta \in \bar{\Theta}} \frac{|\nabla_{\theta}^l (\sigma(y, \theta)^2)|}{\sigma(y, \theta)^2} \leq C(1 + |y|_{(\rho^j)_j, p}^p), \quad \sup_{\theta, \tilde{\theta} \in \bar{\Theta}, |\theta - \tilde{\theta}|_1 < \iota} \frac{\sigma(y, \tilde{\theta})^2}{\sigma(y, \theta)^2} \leq C(1 + |y|_{(\rho^j)_j, p}^p).$$

(ii) for $l = 0, 1, 2$,

$$\begin{aligned} & \sup_{\theta, \tilde{\theta} \in \bar{\Theta}, |\theta - \tilde{\theta}|_1 < \iota} |\nabla_{\theta}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta) - \nabla_{\theta}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y', \theta)| \\ & \leq C(1 + |y|_{(\rho^j)_j, 2p}^{2p} + |y'|_{(\rho^j)_j, 2p}^{2p}) |y - y'|_{(\rho^j)_j, p}^p (1 + \zeta^2). \end{aligned}$$

and

$$\sup_{\theta, \theta', \tilde{\theta} \in \bar{\Theta}, |\theta - \tilde{\theta}|_1 < \iota, |\theta' - \tilde{\theta}|_1 < \iota} \frac{|\nabla_{\theta}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta) - \nabla_{\theta'}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta')|}{|\theta - \theta'|_1} \leq C(1 + |y|_{(\rho^j)_j, p}^p)(1 + \zeta^2).$$

(iii) for $l = 1, 2$,

$$\sup_{\theta \in \bar{\Theta}} |\nabla_{\theta}^l \ell(x, y, \theta) - \nabla_{\theta}^l \ell(x, y', \theta)| \leq C(1 + |y|_{(\rho^j)_j, 1}^{l+1} + |y'|_{(\rho^j)_j, 1}^{l+1})(1 + |x|) \cdot |y - y'|_{(\rho^j)_j, 1}.$$

Proof of Lemma 7.5. (i) From Proposition (3.2)(iii) we obtain that the following explicit representation holds, where $F(y, \theta) := (\alpha_0 + \sum_{j=1}^r \alpha_j y_j, 0, \dots, 0)'$:

$$\sigma(y, \theta)^2 = \sum_{k=0}^{\infty} (B(\theta)^k F(y_{k \rightarrow}, \theta))_1, \quad (41)$$

where $y_{k \rightarrow} = (y_{k+1}, y_{k+2}, \dots)$. We conclude that

$$\begin{aligned} \sigma(y, \theta)^2 &= \alpha_0 \sum_{k=0}^{\infty} (B(\theta))_{11} + \sum_{j=1}^r \alpha_j \sum_{k=0}^{\infty} (B(\theta)^k)_{11} y_{k+j} \\ &\stackrel{k'=k+j}{=} \alpha_0 \sum_{k=0}^{\infty} (B(\theta))_{11} + \sum_{k'=1}^{\infty} \left(\sum_{j=1}^r \alpha_j (B(\theta)^{k'-j})_{11} \mathbb{1}_{k' \geq j} \right) y_{k'} \\ &=: c_0(\theta) + \sum_{k'=1}^{\infty} c_{k'}(\theta) y_{k'}. \end{aligned} \quad (42)$$

From Proposition (3.2)(iii) we obtain that $c_j(\theta) \geq 0$ satisfies

$$\sup_{\theta \in \bar{\Theta}} |c_k(\theta)| \leq C \cdot \rho^k \quad (43)$$

with some $\rho \in (0, 1)$, $C > 0$ and $c_0(\theta) \geq \sigma_{min}^2 > 0$ (due to $\alpha_0 \geq \alpha_{min} > 0$).

Furthermore we conclude that $\sigma(y, \theta)^2$ is three times continuously differentiable w.r.t. θ with

$$\nabla_{\theta}^k (\sigma(y, \theta)^2) = \nabla_{\theta}^k c_0(\theta) + \sum_{k=1}^{\infty} \nabla_{\theta}^k c_k(\theta) \cdot y_k, \quad k \in \{0, 1, 2, 3\}, \quad (44)$$

where $(\nabla_{\theta}^k c_k(\theta))_k$ is still geometrically decaying with $\sup_{\theta \in \Theta} |\nabla_{\theta}^l c_l(\theta)|_{\infty} \leq C \cdot \rho^k$, say (enlarge $C > 0, \rho \in (0, 1)$ if necessary).

In the following we make use of some arguments that were already used in [8]. Note that for $j = 0, \dots, r$, we have $\partial_{\alpha_j} F(y, \theta) \leq \frac{1}{\alpha_j} F(y, \theta)$ and thus

$$\partial_{\alpha_j} c_k(\theta) \leq \frac{1}{\alpha_j} c_k(\theta). \quad (45)$$

For $j = 1, \dots, s$, we have (' \leq ' is meant component-wise)

$$\partial_{\beta_j} (B(\theta)^k) = \sum_{i=1}^k B(\theta)^{i-1} (\partial_{\beta_j} B(\theta)) B(\theta)^{k-i} \leq \frac{1}{\beta_j} k B(\theta)^k.$$

since $\partial_{\beta_j} B(\theta) \leq \frac{1}{\beta_j} B(\theta)$. We therefore obtain

$$\partial_{\beta_j} c_k(\theta) \leq \frac{1}{\beta_j} k \cdot c_k(\theta). \quad (46)$$

From (45) and (46) we obtain the inequalities

$$\partial_{\theta_j} c_k(\theta) \leq \frac{k+1}{\theta_j} c_k(\theta).$$

Similar argumentations lead to the bounds for higher order derivatives (cf. also [8]):

$$\partial_{\theta_{j_1}} \partial_{\theta_{j_2}} c_k(\theta) \leq \frac{(k+1)^2}{\theta_{j_1} \theta_{j_2}} c_k(\theta), \quad \partial_{\theta_{j_1}} \partial_{\theta_{j_2}} \partial_{\theta_{j_3}} c_k(\theta) \leq \frac{(k+1)^3}{\theta_{j_1} \theta_{j_2} \theta_{j_3}} c_k(\theta).$$

If $\bar{\Theta} \subset \text{int}(\tilde{\Theta})$ is some compact subspace, we therefore obtain with $C_1 := \max\{\frac{1}{\theta_j} : j = 1, \dots, r + s + 1, \theta \in \bar{\Theta}\}$ for arbitrary small $p > 0$:

$$\begin{aligned} \frac{\partial_{\theta_j} (\sigma(y, \theta)^2)}{\sigma(y, \theta)^2} &\leq C_1 \frac{\sum_{k=0}^{\infty} (k+1) c_k(\theta)}{\sum_{k=0}^{\infty} c_k(\theta)} \\ &\leq \frac{C_1 c_0(\theta)}{\sigma_{min}^2} + C_1 \sum_{k=1}^{\infty} (k+1) \frac{c_k(\theta) y_k}{c_0(\theta) + c_k(\theta) y_k} \\ &\leq \frac{C_1 c_0(\theta)}{\sigma_{min}^2} + \sum_{k=1}^{\infty} (k+1) \left(\frac{c_k(\theta)}{c_0(\theta)} \right)^p y_k^p, \end{aligned}$$

where we have used $\frac{x}{1+x} \leq x^s$ in the last inequality. Since $c_k(\theta)^s \leq C^s (\rho^s)^k$, we can find $\tilde{C} > 0, \tilde{\rho} \in (0, 1)$ such that

$$\sup_{\theta \in \bar{\Theta}} \frac{|\partial_{\theta_j} (\sigma(y, \theta)^2)|}{\sigma(y, \theta)^2} \leq \tilde{C} (1 + |y|_{(\tilde{\rho}^j)_j, p}^p),$$

and similarly for the higher order derivatives (component-wise):

$$\sup_{\theta \in \bar{\Theta}} \frac{|\nabla_{\theta}^l (\sigma(y, \theta)^2)|}{\sigma(y, \theta)^2} \leq \tilde{C} (1 + |y|_{(\tilde{\rho}^j)_j, p}^p), \quad l = 1, 2, 3.$$

For $\theta, \tilde{\theta} \in \bar{\Theta}$ and arbitrary small $p > 0$, choose $\delta > 0$ such that $\bar{\rho} := (1 + \delta)\rho^p < 1$. Then choose $\iota > 0$ such that $|\theta - \tilde{\theta}|_1 < \iota$ implies (component-wise) $B(\tilde{\theta}) \leq (1 + \delta)B(\theta)$. For $|\theta - \tilde{\theta}| < \iota$, it then holds that $c_k(\tilde{\theta}) \leq (1 + \delta)^k c_k(\theta)$. We conclude that

$$\begin{aligned} \frac{\sigma(y, \tilde{\theta})^2}{\sigma(y, \theta)^2} &\leq \frac{c_0(\tilde{\theta})}{\sigma_{min}^2} + \sum_{k=1}^{\infty} \frac{c_k(\tilde{\theta})y_k}{c_0(\theta) + c_k(\theta)y_k} \\ &\leq \frac{c_0(\tilde{\theta})}{\sigma_{min}^2} + \sum_{k=1}^{\infty} \frac{c_k(\tilde{\theta})}{c_k(\theta)} \cdot \left(\frac{c_k(\theta)}{c_0(\theta)}\right)^p y_k^p \\ &\leq \frac{c_0(\tilde{\theta})}{\sigma_{min}^2} + \frac{C^p}{\sigma_{min}^{2p}} \sum_{k=1}^{\infty} ((1 + \delta)\rho^p)^k y_k^p. \end{aligned}$$

We conclude that there exists $\bar{C} > 0$, $\bar{\rho} \in (0, 1)$ such that

$$\sup_{\theta, \tilde{\theta} \in \bar{\Theta}, |\theta - \tilde{\theta}|_1 < \iota} \frac{\sigma(y, \tilde{\theta})^2}{\sigma(y, \theta)^2} \leq \bar{C}(1 + |y|_{(\bar{\rho}^j)_j, p}^p).$$

(ii) From the differentiability of $\theta \mapsto \sigma(y, \theta)$ we obtain that $\theta \mapsto \ell(x, y, \theta)$ is three times continuously differentiable and

$$\ell(x, y, \theta) = \frac{1}{2} \left(\frac{x}{\sigma(y, \theta)^2} + \log(\sigma(y, \theta)^2) \right), \quad (47)$$

$$\nabla_{\theta} \ell(x, y, \theta) = \frac{\nabla_{\theta}(\sigma(y, \theta)^2)}{2\sigma(y, \theta)^2} \left(1 - \frac{x}{\sigma(y, \theta)^2} \right), \quad (48)$$

$$\begin{aligned} \nabla_{\theta}^2 \ell(x, y, \theta) &= \left[-\frac{\nabla_{\theta}(\sigma(y, \theta)^2) \nabla_{\theta}(\sigma(y, \theta)^2)'}{2\sigma(y, \theta)^4} + \frac{\nabla_{\theta}^2(\sigma(y, \theta)^2)}{2\sigma(y, \theta)^2} \right] \left(1 - \frac{x}{\sigma(y, \theta)^2} \right) \\ &\quad + \frac{\nabla_{\theta}(\sigma(y, \theta)^2) \nabla_{\theta}(\sigma(y, \theta)^2)'}{2\sigma(y, \theta)^4} \cdot \frac{x}{\sigma(y, \theta)^2}. \end{aligned} \quad (49)$$

For the corresponding quantity $\tilde{\ell}_{\tilde{\theta}}$ we obtain

$$\nabla_{\theta} \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta) = \frac{\nabla_{\theta}(\sigma(y, \theta)^2)}{2\sigma(y, \theta)^2} \left(1 - \frac{\sigma(y, \tilde{\theta})^2}{\sigma(y, \theta)^2} \zeta^2 \right).$$

By (i), we obtain that for $p > 0$, there exist constants $\iota > 0$, $C_2 > 0$, $\rho_2 \in (0, 1)$ such that (component-wise):

$$\sup_{|\theta - \tilde{\theta}|_1 < \iota} |\nabla_{\theta} \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta)| \leq C_2 (1 + |y|_{(\rho_2^j)_j, p/2}^{p/2}) (1 + (1 + |y|_{(\rho_2^j)_j, p/2}^{p/2}) \zeta^2). \quad (50)$$

By using

$$\begin{aligned} |y|_{(\rho_2^j)_j, p/2}^{p/2} &\leq \sum_{j=1}^{\infty} \rho_2^{j/2} \cdot \rho_2^{j/2} y_j^p \leq \left(\sum_{j=1}^{\infty} \rho_2^j \right)^{1/2} \left(\sum_{j=1}^{\infty} \rho_2^j y_j^p \right)^{1/2} \\ &= (1 - \rho_2)^{-1/2} |y|_{(\rho_2^j)_j, p}^{p/2}, \end{aligned} \quad (51)$$

we can obtain the more compact form

$$\sup_{|\theta - \tilde{\theta}|_1 < \iota} |\nabla_{\theta} \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta)| \leq C_3(1 + |y|_{(\rho_2^j)_{j,p}}^p)(1 + \zeta^2).$$

with some new constant $C_3 > 0$. Due to the similar structure, we can use similar techniques to obtain (component-wise):

$$\sup_{|\theta - \tilde{\theta}|_1 < \iota} |\nabla_{\theta}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta)| \leq C_3(1 + |y|_{(\rho_2^j)_{j,p}}^p)(1 + \zeta^2) =: M_p(y, \zeta), \quad l = 1, 2, 3. \quad (52)$$

From (44) we deduce that (component-wise) for $l = 0, 1, 2$ with some constant $C_4 > 0$, uniformly in $\theta, \theta' \in \tilde{\Theta}$:

$$|\nabla_{\theta}^l(\sigma(y, \theta)^2) - \nabla_{\theta'}^l(\sigma(y', \theta')^2)| \leq C_4|y - y'|_{(\rho^j)_{j,1}}, \quad (53)$$

By using $|\frac{1}{\sigma(y, \theta)^2} - \frac{1}{\sigma(y', \theta')^2}| \leq \frac{1}{\sigma_{min}^4}|\sigma(y, \theta)^2 - \sigma(y', \theta')^2|$ and the very rough bounds $\sigma(y, \theta)^2 \geq \sigma_{min}^2$, (53) and (44), we obtain (component-wise) with some constant $C_5 > 0$:

$$\sup_{\theta \in \tilde{\Theta}} |\nabla_{\theta} \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta) - \nabla_{\theta'} \tilde{\ell}_{\tilde{\theta}}(\zeta, y', \theta')| \leq C_5(1 + |y|_{(\rho^j)_{j,1}} + |y'|_{(\rho^j)_{j,1}})^2|y - y'|_{(\rho^j)_{j,1}}(1 + \zeta^2)$$

Similar results can be obtained for higher derivatives (component-wise), $l = 1, 2$:

$$\begin{aligned} & \sup_{\theta \in \tilde{\Theta}} |\nabla_{\theta}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta) - \nabla_{\theta'}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y', \theta')| \\ & \leq C_5(1 + |y|_{(\rho^j)_{j,1}} + |y'|_{(\rho^j)_{j,1}})^{1+l}|y - y'|_{(\rho^j)_{j,1}}(1 + \zeta^2) =: N_l(y, y', \zeta). \end{aligned} \quad (54)$$

Using (52) and (54), we have for $l = 1, 2$ and arbitrary small $p' > 0$ (use $\min\{1, x\} \leq x^{p'}$):

$$\begin{aligned} & \sup_{\theta \in \tilde{\Theta}} |\nabla_{\theta}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta) - \nabla_{\theta'}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y', \theta')| \\ & \leq \min\{M_p(y, \zeta) + M_p(y', \zeta), N_l(y, y', \zeta)\} \\ & = \{M_p(y, \zeta) + M_p(y', \zeta)\} \min\{1, \frac{N_l(y, y', \zeta)}{M_p(y, \zeta) + M_p(y', \zeta)}\} \\ & \leq \{M_p(y, \zeta) + M_p(y', \zeta)\} \left(\frac{N_l(y, \zeta)}{M_p(y, \zeta) + M_p(y', \zeta)}\right)^{p'} \\ & = \{M_p(y, \zeta) + M_p(y', \zeta)\}^{1-p'} N_l(y, \zeta)^{p'}. \end{aligned}$$

Choosing $p' \in (0, \min\{1, p'(1+l)\})$, we obtain

$$\begin{aligned} \{M_p(y, \zeta) + M_p(y', \zeta)\}^{1-p'} & \leq C_3^{1-p'}(1 + |y|_{(\rho_2^j)_{j,p}}^p + |y'|_{(\rho_2^j)_{j,p}}^p)(1 + \zeta^2)^{1-p'}, \\ N_l(y, y', \zeta)^{p'} & \leq C_5^{p'}(1 + |y|_{(\rho^{pj})_{j,p}}^p + |y'|_{(\rho^{pj})_{j,p}}^p)|y - y'|_{(\rho^{pj})_{j,p}}(1 + \zeta^2)^{p'}. \end{aligned}$$

With (51), $\rho_3 := \max\{\rho_2, \rho^p\}$ and some constant $C_6 > 0$

$$\sup_{\theta, \tilde{\theta} \in \tilde{\Theta}, |\theta - \tilde{\theta}|_1 < \iota} |\nabla_{\theta}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta) - \nabla_{\theta'}^l \tilde{\ell}_{\tilde{\theta}}(\zeta, y', \theta')| \leq C_6(1 + |y|_{(\rho_3^j)_{j,2p}}^{2p} + |y'|_{(\rho_3^j)_{j,2p}}^{2p})|y - y'|_{(\rho_3^j)_{j,p}}^p(1 + \zeta^2).$$

By using (52) and the mean value theorem, we obtain for $l = 1, 2$:

$$\begin{aligned} & \sup_{\theta, \theta', \bar{\theta} \in \bar{\Theta}, |\theta - \bar{\theta}|_1 < \iota, |\theta' - \bar{\theta}|_1 < \iota} \frac{|\nabla_{\theta}^l \tilde{\ell}_{\bar{\theta}}(\zeta, y, \theta) - \nabla_{\theta'}^l \tilde{\ell}_{\bar{\theta}}(\zeta, y, \theta')|}{|\theta - \theta'|_1} \\ & \leq \sup_{|\bar{\theta} - \bar{\theta}|_1 < \iota} |\nabla_{\theta}^{l+1} \tilde{\ell}_{\bar{\theta}}(\zeta, y, \bar{\theta})|_{\infty} \leq M_p(y, \zeta), \end{aligned}$$

giving the result.

(iii) Using the representations (48), (49) and the inequalities (53), (44) and $\sigma(y, \theta)^2 \geq \sigma_{min}^2$, this is an immediate consequence. \square

References

- [1] István Berkes, Lajos Horváth, and Piotr Kokoszka. GARCH processes: structure and estimation. *Bernoulli*, 9(2):201–227, 2003.
- [2] Nick Bloom. Uncertainty and the dynamics of R&D. *American Economic Review*, 97(2):250–255, 2007.
- [3] Tim Bollerslev and Jeffrey M Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, 11(2):143–172, 1992.
- [4] Philippe Bougerol and Nico Picard. Stationarity of garch processes and of some nonnegative time series. *Journal of econometrics*, 52(1-2):115–127, 1992.
- [5] Philippe Bougerol and Nico Picard. Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, 20(4):1714–1730, 1992.
- [6] Robert F Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007, 1982.
- [7] Jianqing Fan, Lei Qi, and Dacheng Xiu. Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business & Economic Statistics*, 32(2):178–191, 2014.
- [8] Christian Francq and Jean-Michel Zakoïan. Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli*, 10(4):605–637, 2004.
- [9] Christian Francq and Jean-Michel Zakoïan. Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. *Econometrica*, 80(2):821–861, 2012.
- [10] Christian Hafner. Testing for bubbles in cryptocurrencies with time-varying volatility, 2018.
- [11] David I Harvey, Stephen J Leybourne, and Yang Zu. Testing explosive bubbles with time-varying volatility. *Econometric Reviews*, 2018.
- [12] Søren Tolver Jensen and Anders Rahbek. Asymptotic normality of the qmle estimator of arch in the nonstationary case. *Econometrica*, 72(2):641–646, 2004.
- [13] Kyle Jurado, Sydney C Ludvigson, and Serena Ng. Measuring uncertainty. *American Economic Review*, 105(3):1177–1216, 2015.
- [14] Sang-Won Lee and Bruce E Hansen. Asymptotic theory for the garch (1, 1) quasi-maximum likelihood estimator. *Econometric theory*, 10(1):29–52, 1994.
- [15] Daniel B Nelson. Stationarity and persistence in the GARCH (1, 1) model. *Econometric theory*, 6(3):318–334, 1990.

- [16] Marc S. Paoletta. *Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH*. Chichester, 2018.
- [17] Liang Peng and Qiwei Yao. Least absolute deviations estimation for arch and garch models. *Biometrika*, 90(4):967–975, 2003.
- [18] Peter CB Phillips, Yangru Wu, and Jun Yu. Explosive behavior in the 1990s nasdaq: When did exuberance escalate asset values? *International economic review*, 52(1):201–226, 2011.
- [19] A. W. van der Vaart. *Asymptotic statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 1998.
- [20] Wei Biao Wu and Wanli Min. On linear processes with dependent innovations. *Stochastic Processes and their Applications*, 115(6):939 – 958, 2005.
- [21] Wei Biao Wu and Xiaofeng Shao. Limit theorems for iterated random functions. *J. Appl. Probab.*, 41(2):425–436, 2004.
- [22] Wei Biao Wu and Zhou Zhou. Gaussian approximations for non-stationary multiple time series. *Statist. Sinica*, 21(3):1397–1413, 2011.
- [23] D. Zhang and W. B. Wu. Gaussian Approximation for High Dimensional Time Series. *Ann. Statist.*, 45(5):1895–1919, 2017.

IRTG 1792 Discussion Paper Series 2020



For a complete list of Discussion Papers published, please visit
<http://irtg1792.hu-berlin.de>.

- 001 "Estimation and Determinants of Chinese Banks' Total Factor Efficiency: A New Vision Based on Unbalanced Development of Chinese Banks and Their Overall Risk" by Shiyi Chen, Wolfgang K. Härdle, Li Wang, January 2020.
- 002 "Service Data Analytics and Business Intelligence" by Desheng Dang Wu, Wolfgang Karl Härdle, January 2020.
- 003 "Structured climate financing: valuation of CDOs on inhomogeneous asset pools" by Natalie Packham, February 2020.
- 004 "Factorisable Multitask Quantile Regression" by Shih-Kang Chao, Wolfgang K. Härdle, Ming Yuan, February 2020.
- 005 "Targeting Customers Under Response-Dependent Costs" by Johannes Haupt, Stefan Lessmann, March 2020.
- 006 "Forex exchange rate forecasting using deep recurrent neural networks" by Alexander Jakob Dautel, Wolfgang Karl Härdle, Stefan Lessmann, Hsin-Vonn Seow, March 2020.
- 007 "Deep Learning application for fraud detection in financial statements" by Patricia Craja, Alisa Kim, Stefan Lessmann, May 2020.
- 008 "Simultaneous Inference of the Partially Linear Model with a Multivariate Unknown Function" by Kun Ho Kim, Shih-Kang Chao, Wolfgang K. Härdle, May 2020.
- 009 "CRIX an Index for cryptocurrencies" by Simon Trimborn, Wolfgang Karl Härdle, May 2020.
- 010 "Kernel Estimation: the Equivalent Spline Smoothing Method" by Wolfgang K. Härdle, Michael Nussbaum, May 2020.
- 011 "The Effect of Control Measures on COVID-19 Transmission and Work Resumption: International Evidence" by Lina Meng, Yinggang Zhou, Ruige Zhang, Zhen Ye, Senmao Xia, Giovanni Cerulli, Carter Casady, Wolfgang K. Härdle, May 2020.
- 012 "On Cointegration and Cryptocurrency Dynamics" by Georg Keilbar, Yanfen Zhang, May 2020.
- 013 "A Machine Learning Based Regulatory Risk Index for Cryptocurrencies" by Xinwen Ni, Wolfgang Karl Härdle, Taojun Xie, August 2020.
- 014 "Cross-Fitting and Averaging for Machine Learning Estimation of Heterogeneous Treatment Effects" by Daniel Jacob, August 2020.
- 015 "Tail-risk protection: Machine Learning meets modern Econometrics" by Bruno Spilak, Wolfgang Karl Härdle, October 2020.
- 016 "A data-driven P-spline smoother and the P-Spline-GARCH models" by Yuanhua Feng, Wolfgang Karl Härdle, October 2020.
- 017 "Using generalized estimating equations to estimate nonlinear models with spatial data" by Cuicui Lu, Weining Wang, Jeffrey M. Wooldridge, October 2020.
- 018 "A supreme test for periodic explosive GARCH" by Stefan Richter, Weining Wang, Wei Biao Wu, October 2020.

IRTG 1792, Spandauer Strasse 1, D-10178 Berlin
<http://irtg1792.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the IRTG 1792.