Manuscript submitted to *The Econometrics Journal*, pp. 1–39.

Testing for parameter change epochs in GARCH time series

Stefan Richter[†], Weining Wang[‡] and Wei Biao Wu[*]

[†] *Institute of Applied Mathematics, Im Neuenheimer Feld 205, Heidelberg, Germany.*
E-mail: stefan.richter@iwr.uni-heidelberg.de
[‡] *University of York[‡].*
E-mail: weining.wang@york.ac.uk
[*] *University of Chicago.*
E-mail: wbwu@galton.uchicago.edu

Summary We develop a uniform test for detecting and dating the integrated or mildly explosive behaviour of a strictly stationary generalized autoregressive conditional heteroskedasticity (GARCH) process. Namely, we test the null hypothesis of a globally stable GARCH process with constant parameters against the alternative that there is an "abnormal" period with changed parameter values. During this period, the parameter-value change may lead to an integrated or mildly explosive behaviour 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 breaks, and then provide an algorithm to identify the periods of changes. Our theoretical results hold under mild moment assumptions on the innovations of the GARCH process. Technically, the existing properties for the quasi-maximum likelihood estimation (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 long time series. We apply the test to the conventional early-warning indicators of both the financial market and a representative of the emerging Fintech market, i.e. the Bitcoin returns.

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

1. INTRODUCTION

Volatility is an important indicator for economic and financial stability. There is growing evidence of the unstable behaviour of the historical volatility of numerous micro- and macro- level data, such as individual asset returns, VIX (the Chicago Board Options Exchange volatility index), inflation and unemployment. Bloom (2007) documents the unstable behaviour of the 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. Ignoring parameter change may thus lead to biased analysis in policymaking and forecasting. This motivates us to consider a general method of testing parameter constancy for models of volatility.

For modeling the volatility processes, the highly celebrated autoregressive conditional heteroskedasticity (ARCH) model proposed by Engle (1982) 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.

$$X_i^2 = \zeta_i^2 \sigma_i^2,$$
$$\sigma_i^2 = \alpha_0 + \sum_{j=1}^{r} \alpha_j X_{i-j}^2 + \sum_{k=1}^{s} \beta_k \sigma_{i-k}^2, \tag{1.1}$$

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 Paolella (2018) for more details of the model. Among various possible changes of parameters of the underlying process, moving from the covariance stationarity to the infinite variance has come to the center of our focus for its potential use of detecting periods of economic uncertainty. In addition to the case of integrated GARCH, we refer to the volatility process behaving more explosive after the change as a "mildly explosive" one, which can be considered as an analogue of a mildly explosive unit-root return process. The name "mildly explosive" follows Lee and Hansen (1994), who refer to a GARCH(1,1) model with $\alpha_1 + \beta_1 > 1$ as a "mildly explosive" one.

Is there empirical evidence of the existence of mildly explosive region of a GARCH model with fitted parameters? One often sees sudden, integrated or mildly explosive behaviour in the second moment of the process which bounces back after a while. For example, in Figure 1, we have plotted the squared returns from a realization of a piecewise mildly explosive GARCH(1,1) process and the squared log returns of Bitcoin at the bottom. The whole time span is set to be (July 19, 2010- April 13, 2013) with a mildly explosive period with changing parameter values (March 26, 2011 - Aug. 23, 2011). The two trajectories of the time series look rather similar. We see that the piecewise mildly explosive GARCH process in Figure 1 captures the mildly explosive behaviour of Bitcoin in the squared returns. Moreover, Figure 2 shows a rolling window fit of parameter values of a GARCH(1,1) model using Bitcoin data. We can see clear signs of time varying parameters. In particular, there are regions of the estimated parameters falling out of the covariance stationary regime ($\hat{\alpha}_1 + \hat{\beta}_1 \geq 1$). Such kind of data phenomena suggest that the underlying processes have time varying parameters, calling for a rigorous quantitative treatment for detection of change periods and making corresponding inference.

The aim of our paper is to develop a generalised uniform test for GARCH models which is able to detect exuberant behaviour periods (periods with integrated or mildly explosive parameter values) which are associated with the empirical phenomena of mild explosiveness in the second moment. The test is constructed by looking at the supreme of Wald-type test statistics over all possible intervals with changing parameters. Numerous estimation methods for the parameters of GARCH models have been proposed, and their consistency and asymptotic normality have been carefully studied in the literature. A conventional estimation approach is the QMLE, e.g. Bollerslev and Wooldridge (1992). Also Fan et al. (2014) study QMLE of GARCH models with heavy-tailed likelihoods. Peng and Yao (2003) propose a least absolute deviation estimator. Jensen and Rahbek (2004) 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 the strict stationarity of a GARCH model, there is still a region of parameter values allowing for realizations with unstable volatility behaviour. The leading case is the "IGARCH" pro-

Figure 1. A plot of simulated piecewise mildly explosive squared returns in the upper panel versus the actual realized Bitcoin squared log returns after fitting an ARMA (autoregressive moving average) model (lower panel). The simulation plot in the upper panel is according to the fitted stable period (20100719- 20130413, $\hat{\alpha}_1 = 0.19$, $\hat{\beta}_1 = 0.79$) versus a mildly explosive period (20110326-20110823, $\hat{\alpha}_1 = 0.69$, $\hat{\beta}_1 = 0.45$).

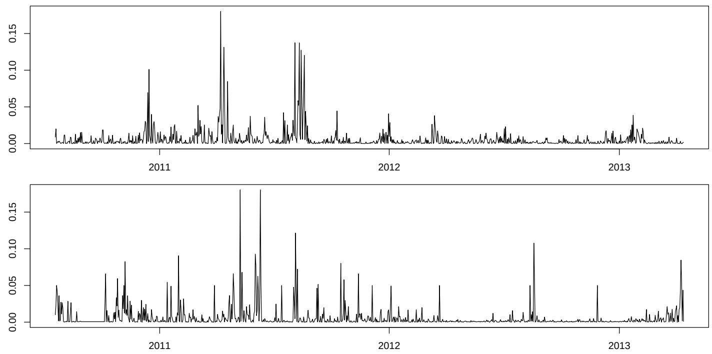
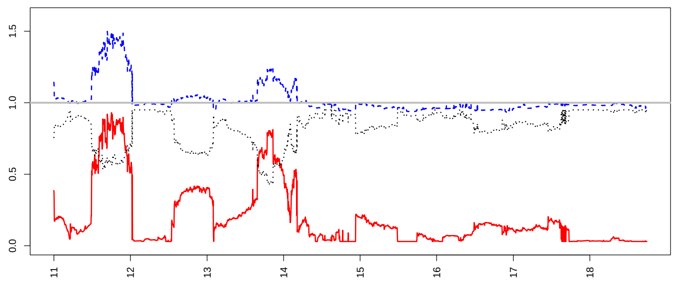


Figure 2. A plot of estimated GARCH(1,1) parameters using the Bitcoin data over a rolling window of size 200. $\hat{\alpha}_1 + \hat{\beta}_1$ estimate persistence parameter (blue dash line), $\hat{\alpha}_1$ (red solid line), $\hat{\beta}_1$ (black dotted line), threshold of mild explosiveness ($\alpha_1 + \beta_1 = 1$).



cess. Nelson (1990) looks at the behaviour of an Integrated-GARCH (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 behaviour). Lee and Hansen (1994) provide an asymptotic theory for a strictly stationary GARCH(1,1) QML estimator allowing for the case of IGARCH and mildly explosive conditional variance and even nonstationarity. Jensen and Rahbek (2004) consider asymptotic inference for a 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 integrated/mild explosiveness in an uniform manner. Francq and Zakoïan (2012) provide important estimation results on nonstationary GARCH models, and they also provide a test for parameter constancy of a GARCH(1,1) process without assuming strict stationarity. Complementing to their study, our focus is on the integrated/mildly explosive parameter region and we extend the test to a uniform context. There is also a large and extensive literature on testing for mild 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, Phillips et al. (2011) for a left-tailed, augmented Dickey-Fuller test (ADF) for the mildly explosive behaviour in the 1990s Nasdaq. Hafner (2018) considers such bubble tests for crypto-currencies. Harvey et al. (2018) investigate a bubble test with a smooth time varying volatility function. The underlying models focus usually on unit-root or mildly explosive auto-regressive (AR) processes to test the change of the AR(1) coefficient. Often, the variance of the errors stays the same or varies 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 a mild 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. In addition, there is literature on break detection for multiple break points for nonlinear time series, cf. Bardet et al. (2012), Davis et al. (2008) and Fryzlewicz and Subba Rao (2014). In particular, Bardet et al. (2012) derive a breakpoint detection procedure for general recursively defined time series via a penalized maximum likelihood method and prove its consistency. Their formulation is rather general, leading to the restriction that the considered time series have to be covariance stationary. Davis et al. (2008) propose a model with piece-wise stationary time series with independent segments. Fryzlewicz and Subba Rao (2014) invent a novel method to find break points and test for covariance stationary ARCH processes using a CUSUM (cumulative sum) statistics. Our test is different but complement to the above study as we propose breakpoint detectors for GARCH models in the non covariance-stationary regime and provide a solid theoretical backup via a uniform testing procedure for the presence of breakpoints.

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 behaviour to the volatility process; see Jurado et al. (2015). On the contrary, volatility bursting can also be related to time-varying risk aversion, sentiment, bubbles or uncertainty. Nevertheless, we are trying to establish a rigorous theoretical framework of testing for the mildly explosive interval using a GARCH model for the volatility process. It should be stressed that we focus on one aspect of the parameter, namely, changes in the parameters driving the volatility over time. We do not claim that our method can directly identify the cause of this behaviour. In sum, we develop a change-point test for detecting possible unstable behaviour 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 in which the parameter values change to another (higher) values. This increase potentially leads to a period of mildly 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 the alternative are still in the regime where the GARCH process is strictly stationary. The theoretical contributions are extending the existing theoretical results on GARCH QML estimators to uniform consistency statements over an arbitrary observation period. Besides, 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 Zhang and Wu (2017), leading to mild moment assumptions. Empirically, we find that our test is useful for the early identification of the critical periods of financial crisis for two important early-warning indicator of the economic condition.

We introduce some notations we use throughout the paper. For $q > 0$ and vector $v = (v_1, \ldots, v_d)^\top \in \mathbb{R}^d$, 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 sequences (a_n) and (b_n) of positive numbers, we write $a_n = O(b_n)$ or $a_n = o(b_n)$ if there exists a positive constant C such that $a_n/b_n \le C$ or $a_n/b_n \to 0$, respectively. For two sequences of random variables (X_n) and (Y_n), we write $X_n = o_p(Y_n)$ (resp. $X_n = O_p(Y_n)$) if $X_n/Y_n \to 0$ in probability (X_n/Y_n is bounded in probability).

Our text is organized as follows. Section 2 provides the results of the important GARCH(1,1) model and the corresponding test procedure. Section 3 concerns the estimation and theoretical results in a general GARCH(r, s) model. In particular, Section 3.1 introduces the framework of the QMLE and its consistency, 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 behaviour of the test in examples from practice. The technical proofs are delegated to the Appendix.

2. A UNIFORM MILD 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,

$$\begin{aligned} X_i &= \zeta_i \sigma_i, \\ \sigma_i^2 &= \alpha_0(i) + \alpha_1(i) X_{i-1}^2 + \beta_1(i)\sigma_{i-1}^2, \quad i \in \mathbb{Z}, \end{aligned} \tag{2.2}$$

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 $1, \ldots, n$.

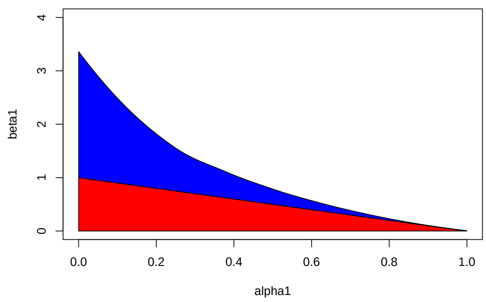
We summarize the parameters into $\theta(i) = (\alpha_0(i), \alpha_1(i), \beta_1(i))'$. In the case that the parameters are constant, i.e. $\theta(i) \equiv \theta = (\alpha_0', \alpha_1', \beta_1')'$, the top Lyapunov exponent associated with this model, according to Bougerol and Picard (1992b), is

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

It is worth noting that $\gamma(\theta) < 0$ allows (for instance) the IGARCH case, i.e. $\alpha_1 + \beta_1 = 1$. We illustrate in Figure 3 the region of parameters corresponding to the case that the volatility process is non-covariance stationary but still strictly stationary (integrated or mildly explosive).

The aim of this paper is to construct a test which is able to detect if there exists a period where the parameters of the GARCH model have changed. The theory developed in this paper is theoretically supported for null hypotheses in the regime of strict stationarity

Figure 3. The plot of the feasible parameter region with a standard normally distributed ζ_1, where the red region corresponds to covariance stationarity and the blue region corresponds to the strictly stationary mildly explosive region; X axis, α_1, Y axis, β_1.



(see the definition of Θ below). *However, the alternative hypotheses of our test includes GARCH processes which are not strictly stationary.* Moreover, due to the monotonicity of the test statistics developed in this paper, we conjecture that the test also works if the null hypothesis lies in the nonstationary regime. Formally, we would like to test whether there exists a period $\{n_1, ..., n_2\}$ (with $1 < n_1 < n_2 < n$), in which the parameter values in (2.2) change their values compared with $\{1, ..., n\}$. The task breaks into two parts: First, one has to check 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 equals $\theta^* = (\alpha_0^*, \alpha_1^*, \beta_1^*) \in \Theta$ at the beginning and possibly has a period of change in $\{\lfloor n\tau_1^* \rfloor + 1, ..., \lfloor n\tau_2^* \rfloor\}$ (where $\tau_1^*, \tau_2^* \in [0, 1]$, $\tau_1^* < \tau_2^*$) with magnitude $\Delta^* > 0$ and direction $H \in \mathbb{R}^3$. Namely,

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

Here $\lfloor x \rfloor$ denotes the flooring operator, i.e. the largest integer smaller than 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 $i = 1, ..., n$ versus the hypothesis that there exists a period of begin integrated or mild explosive in which $\alpha_1^*(i) + \beta_1^*(i) \geq 1$ for some i. $\alpha_1^*(i) + \beta_1^*(i)$ is referred to as the persistence parameter in our setting. Graphically, this corresponds to the question of whether or not there exists regions where the process leaves the variance-stationary regime (i.e. the variance explodes). The formulation that for $i > \lfloor n\tau_2^* \rfloor$, $\theta^*(i)$ returns to the original value θ^* is only assumed for simplicity. Our test can be modified in such a way that the values of $\theta^*(i)$ in this region are not used, cf. Remark 3.2.

It is therefore natural to formulate the hypotheses in the following way: Let $H = (0,1,1)'$. With some fixed $c := \alpha_1^* + \beta_1^* < 1$, we want to test

$$H_0 : \Delta^* < 0 \quad \text{v.s.} \quad H_1 : \Delta^* \geq 0. \tag{2.3}$$

To transfer the setting to the one of change point tests, we modify (2.3) as follows:

$$H_0 : \Delta^* = 0 \quad \text{v.s.} \quad H_1 : \Delta^* > 0. \tag{2.4}$$

We will discuss the connection between (2.3) and (2.4) in Remark 3.3. Our method is a general way to test the parameter constancy for GARCH processes. For example, in practice, a useful choice for c may be obtained from $c = \hat{\alpha}_1 + \hat{\beta}_1$, where $\hat{\alpha}_1$, $\hat{\beta}_1$ are obtained from fitting a global model with all observations. We illustrate this with VIX in our empirical study (c.f. Section 5). To construct a test, we first derive estimators for the parameters. For a fixed period τ_1^*, τ_2^* we can use a standard QMLE approach. It is not hard to see from (2.2), that in the case of the 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 $Y_i^c := (X_{i-1}^2, ..., X_1^2, 0, 0, ...)$ and

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

The estimated parameter with observations during any given period $\{\lfloor n\tau_1 \rfloor+1, ..., \lfloor n\tau_2 \rfloor\}$ 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 the covariance matrix

$$\Sigma = V(\theta^*)^{-1} I(\theta^*) V(\theta^*)^{-1}, \tag{2.5}$$

where

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

and

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

Our next step is 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

$$\hat{B}_n^{(\chi)}(\tau_1, \tau_2) := \sqrt{n}(\tau_2-\tau_1)^\chi ((0,1,1)\bar{\Sigma}_{n,\tau_1,\tau_2}(0,1,1)')^{-1/2}\{\hat{\alpha}_{1,n,\tau_1,\tau_2} + \hat{\beta}_{1,n,\tau_1,\tau_2} - c\}, \tag{2.6}$$

where $\chi \in (0,1)$ 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 components of $\hat{\theta}_{n,\tau_1,\tau_2}$, and $\bar{\Sigma}_{n,\tau_1,\tau_2}$ is an estimator of Σ using observations outside of $\{\lfloor n\tau_1 \rfloor, ..., \lfloor n\tau_2 \rfloor\}$.

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 (3.14) below) with observations outside $\{\lfloor n\tau_1 \rfloor, ..., \lfloor n\tau_2 \rfloor\}$. 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 the variance-covariance matrix of the estimated parameters 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 $\{\frac{B(\tau_2)-B(\tau_1)}{(\tau_2-\tau_1)^{1-\chi}}\}$, where $B(.)$ is a $1-$dimensional Brownian motion. We show this formally in Theorem 3.3 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 of our supreme test statistics. We summarize the text procedure in the following context. The procedure depends on some tuning parameters. We have experimented with various choices and make our suggestions as follows.

Step 0 Choose some $L > 0$ (the number of grid points associated with detection accuracy). The corresponding grid points are $\mathcal{G} = \{\frac{j}{L} : j = 0, ..., L\}$ on the time line. We suggest making L as large as possible so that the calculation on the machine is still done within an acceptable time. In principle, $L = n$ is optimal, but may lead to an infeasible duration of computation in practice.

Step 1 We denote H as a vector which helps to formulate different null hypothesis. Let $H = (0,1,1)'$.

Step 2 Choose values for $\kappa, \kappa' \in (0,1)$ and $\chi \in (0,1)$. The theoretical results hold for all fixed choices of κ, κ', χ. In practice, the choice of κ, κ' is a trade-off between a high power of the testing procedure to find a break point near the boundary and guaranteeing a good quality of the approximation of the asymptotic distribution of the test statistics. We suggest setting $\kappa = 0.1, \kappa' = 0.1$. We have seen in simulations that the selection of χ has no significant influence on the properties of the test. We therefore suggest setting $\chi = 0.5$.

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 (3.11) and calculate $\bar{\Sigma}_{n,\tau_1,\tau_2}$ as in (3.14). Then we can calculate the test statistic as

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

(cf. (3.15)). Figure 4 shows how one calculates the supreme test statistic over different windows associated with the grid points.

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

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

We define $\hat{q}_{W,\delta} := \hat{\mu}_{n,[\lfloor N \cdot \delta \rfloor]}$, where $\hat{\mu}_{n,[1]}, ..., \hat{\mu}_{n,[N]}$ are the order statistics of $\hat{\mu}_{n,1}, ..., \hat{\mu}_{n,N}$.

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

$$\hat{B}_n^{(\chi)} = \sup_{(\tau_1,\tau_2) \in R_{\kappa,\kappa'} \cap \mathcal{G}^2} \hat{B}_n^{(\chi)}(\tau_1, \tau_2) > \hat{q}_{W,\delta}, \tag{2.7}$$

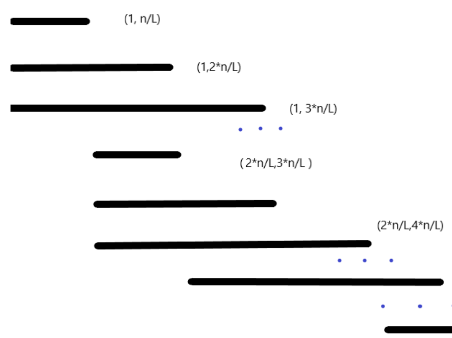
there is a significant shock in the parameter values. In this case, one can estimate the true shock period as $[\tau_1^*, \tau_2^*]$ by

$$(\hat{\tau}_{1,n}, \hat{\tau}_{2,n}) \in \text{argmax}_{(\tau_1,\tau_2)\in R_{\kappa,\kappa'}} \hat{B}_n^{(\chi)}(\tau_1, \tau_2).$$

If instead (2.7) 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 3.3.

Figure 4. The plot of the windows where the supreme is calculated.



We name this the GARCH Supreme Richter-Wang-Wu (GSRWW) test. Its limit distribution is $\sup_{\tau_1,\tau_2\in R_{\kappa,\kappa'}}\{\frac{B(\tau_2)-B(\tau_1)}{(\tau_2-\tau_1)^{1-\chi}}\}$ (with $B(.)$ as the standard Brownian motion).

3. A UNIFORM TEST FOR $GARCH(R, S)$

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

$$\begin{aligned} X_i^2 &= \zeta_i^2 \sigma_i^2, \\ \sigma_i^2 &= \alpha_0(i) + \sum_{j=1}^{r} \alpha_j(i) X_{i-j}^2 + \sum_{k=1}^{s} \beta_k(i) \sigma_{i-k}^2. \end{aligned} \tag{3.8}$$

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 of the parameters being constant, i.e. $\theta(i) \equiv \theta = (\alpha_0, \alpha_1, ..., \alpha_r, \beta_1, ..., \beta_s)'$. Note that we present our set of assumptions to ensure the existence of a unique stationary solution to our model in (3.8) . 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 the jth element being 1, $1 \le j \le r + s$. Define the $(r + s) \times (r + s)$-matrix as

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

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

$$\gamma(\theta) := \lim_{i\to\infty} \frac{1}{i} \log |A_i(\theta)A_{i-1}(\theta)\ldots A_1(\theta)|_2,$$

This exists if $\mathbb{E}|\zeta_0^2|^a < \infty$ for some $a > 0$ (cf. Francq and Zakoïan (2004)).

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*

$$\tilde{\Theta} = \{\theta \in \mathbb{R}_{\geq 0}^{r+s+1} : \alpha_0 \geq \alpha_{min}, \gamma(\theta) < 0 \text{ a.s.}, \sum_{j=1}^{s} \beta_j < 1\}. \tag{3.9}$$

Let $\Theta \subset \tilde{\Theta}$ *be compact. Assume that* $\theta^* \in int(\Theta)$.
(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) $\gamma(\theta) < 0$ guarantees the strict stationarity of the GARCH process. Note that this includes parameter values corresponding to IGARCH or mildly explosive GARCH with $\sum_j \alpha_j + \sum_k \beta_k > 1$.

From Francq and Zakoïan (2004) and Proposition 1.1 in the appendix we see that Assumption 3.1 implies existence of a solution of (3.8) which has geometric decay of dependence.

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

In this subsection, we describe the QMLE and formulate a theorem which yields its uniform consistency. For estimation of the parameters $\theta \in \Theta$, we consider the following QML approach. We denote by $Y_i^c := (X_{i-1}^2, X_{i-2}^2, ..., X_1^2, 0, 0, ...)$ 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

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

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

$$\sigma(y,\theta)^2 = \alpha_0 + \sum_{j=1}^{r} \alpha_j y_j + \sum_{k=1}^{s} \beta_k \sigma((y_{k+1}, y_{k+2}, ...), \theta)^2. \tag{3.10}$$

The analytic definition of the recursion of $\sigma(y, \theta)^2$ is formulated in a forward way (using $y_1, y_2, ...$ instead of y_{-1},y_{-2},...) 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 (3.10) terminates after a finite number of steps due to zeros in Y_i^c. Morevoer, instead of using the truncated version $Y_i^c = (X_{i-1}^2, X_{i-2}^2, ..., X_1^2, 0, ..., 0)$

which corresponds to assuming that all initial values $X_0^2 = X_{-1}^2 = ... = 0$, one can also use different initial values like $X_0^2 = X_{-1}^2 = ... = \alpha_0$ or $X_0^2 = X_{-1}^2 = ... = X_1^2$ as investigated in Francq and Zakoïan (2004). For a discussion of different initial values, consider Bougerol and Picard (1992a) (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 $i = \lfloor n\tau_1 \rfloor + 1, ..., \lfloor n\tau_2 \rfloor$ is obtained by

$$\hat{\theta}_{n,\tau_1,\tau_2} := \operatorname{argmin}_{\theta \in \Theta} L^c_{n,\tau_1,\tau_2}(\theta). \tag{3.11}$$

With these definitions, we obtain the following uniform consistency under the null hypothesis of no parameter change.

THEOREM 3.1. (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 $\lfloor n\tau_1^* \rfloor + 1, ..., \lfloor n\tau_2^* \rfloor$, 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) mildly explosive behaviour even in the stationary case if $\alpha_1 + \beta_1$ is large. To model the mildly explosive behaviour 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 $\theta^*(i) = (\alpha_0^*(i), \alpha_1^*(i), ..., \alpha_r^*(i), \beta_1^*(i), ..., \beta_s^*(i))'$ denotes the true parameter in the baseline model, which possibly has a period of change in $\{\lfloor n\tau_1^* \rfloor + 1, ..., \lfloor n\tau_2^* \rfloor\}$ (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 (3.8) 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. Francq and Zakoïan (2012)).

In general, we fix some $H \in \mathbb{R}^{(r+s+1)\times 1}$ and formulate the hypotheses

$$H_0 : \Delta^* = 0 \quad vs. \quad H_1 : \Delta^* > 0. \tag{3.12}$$

We now give an important example.

Example 3.1. ($GARCH^{sh}(1,1)$) *Here, $\theta = (\alpha_0, \alpha_1, \beta_1)'$. Fix some $\bar{\alpha}_1$, for instance $\bar{\alpha}_1 = 0.99$, 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 $i = 1, ..., n$ against $\alpha_1^*(i) > \bar{\alpha}_1$ for some i. Thus it corresponds to testing (3.12) with $H = (0, 1, 0)'$.*

Motivated by the consistency result of Theorem 3.1, we propose a test based on the supreme distance between the estimated targeting parameter value and the value under the null,

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

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 3.1. (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 3.1,*

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

which ensures that our test has asymptotic power of 1.

Remark 3.2. *i) We conjecture that this result can be extended even to nonstationary alternatives where $\theta^* + H\Delta^* \notin \Theta$ as long as $H'(1, 0, ..., 0) = 0$ (Francq and Zakoïan (2012) discover that one cannot expect $\hat{\alpha}_0$ to be consistently estimated in the nonstationary regime).*

ii) We note that this method does not restrict to the case that the parameters change back to the same value afterwards. One can easily extend the following theory to the case where a different value is obtained after the change by using an estimate for Σ which is only based on the observations at time points $1, ..., \lfloor n\tau_1 \rfloor - 1$.

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

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

and the Fisher information matrix

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

Note that if $\mathbb{E}\xi_0^4 < \infty$, then $V(\theta), I(\theta)$ exist in a neighbourhood of θ^* and $I(\theta^*)$ is nonsingular with $I(\theta^*) = \frac{\mu_4 - 1}{2} V(\theta^*)$. Details can be found in Proposition 1.2 in the Appendix.

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.

THEOREM 3.2. (WEAK BAHADUR REPRESENTATION) *Let Assumption 3.1 hold. Assume that for some $a > 0$, $\mathbb{E}|\zeta_0|^{4+a} < \infty$. Then for each $\kappa > 0$,*

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

This linearization result allows to transfer the properties of the sum $\nabla_\theta L_{n,\tau_1,\tau_2}$ to the difference $\hat{\theta}_{n,\tau_1,\tau_2} - \theta^*$; especially we obtain a 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 Wu and Zhou (2011). The functional limit distribution then naturally implies the pointwise convergence results from Francq and Zakoïan (2004) but is much stronger, since it can be used as a starting point to apply theorems (such as the continuous mapping theorem) from empirical process theory. Let $\ell^\infty(T)$ denote the space of bounded functions $f : T \to \mathbb{R}$, (cf. van der Vaart (1998)), Section 18, Example 18.5. As a direct consequence of the uniform Bahadur representation, we can derive the distribution of the difference of the estimator and the true value θ^* under the null.

THEOREM 3.3. (ASYMPTOTIC DISTRIBUTION OF $\hat{\theta}_{n,\tau_1,\tau_2}$) *Suppose that Assumption 3.1 holds. Suppose that there exists $a' > 0$ such that $\mathbb{E}|\zeta_0|^{4+a'} < \infty$. Fix $\kappa > 0$ and that H_0 is true. Then on $\ell^\infty(R_\kappa)^{r+s+1}$,*

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

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

We now 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 more than one deviation from a "stable" state.

COROLLARY 3.1. (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 under H_0,*

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

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

$$\begin{aligned}\Sigma &= V(\theta^*)^{-1}I(\theta^*)V(\theta^*)^{-1} \\ &= \frac{\mu_4 - 1}{2}\cdot V(\theta^*)^{-1}.\end{aligned} \tag{3.13}$$

Here we restrict ourselves to the estimation of Σ via the representation of (3.13) to avoid estimating μ_4 separately. To get a test with high power, it seems reasonable to estimate Σ under the more restrictive situation in 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

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

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

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

Now put

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

Then the following intermediate result holds:

Proposition 3.1. (Estimation of $V(\theta^*)$, $I(\theta^*)$) *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

$$\bar{\Sigma}_{n,\tau_1,\tau_2} := \bar{V}_n(\bar{\theta}_{n,\tau_1,\tau_2})^{-1}\bar{I}_n(\bar{\theta}_{n,\tau_1,\tau_2})\bar{V}_n(\bar{\theta}_{n,\tau_1,\tau_2})^{-1} \tag{3.14}$$

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 3.1 and Proposition 3.1, we obtain that under H_0,

$$\begin{aligned}\hat{B}^{(\chi)}_{n,H} &:= \sqrt{n}\sup_{(\tau_1,\tau_2)\in R_{\kappa,\kappa'}} (\tau_2-\tau_1)^{\chi}\bar{\Sigma}^{-1/2}_{n,\tau_1,\tau_2,H}\big\{H'\hat{\theta}_{n,\tau_1,\tau_2} - H'\theta^*\big\}\\ &\xrightarrow{d} \sup_{(\tau_1,\tau_2)\in R_{\kappa,\kappa'}}\Big\{\frac{B(\tau_2)-B(\tau_1)}{(\tau_2-\tau_1)^{1-\chi}}\Big\} =: W\end{aligned} \tag{3.15}$$

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

Remark 3.3. (Modification of the test and the hypotheses) *i) Let* $q_{W,\delta}$ *denote the* $(1-\delta)$ *quantile of* W*. Then* $\mathbb{1}_{\{\hat{B}^{(\chi)}_{n,H} > q_{W,\delta}\}}$ *is also a level* δ *test for the extended hypotheses*

$$H_0 : \Delta^* \le 0 \quad vs. \quad H_1 : \Delta^* > 0. \tag{3.16}$$

The reason being that $\Delta^* < 0$ *in connection with the uniform consistency of Theorem 3.1 only produces smaller values of the test statistics* $\hat{B}^{(\chi)}_{n,H}$*.*

ii) For any fixed $\theta^* \in int(\Theta)$*, the power function* $\beta(\Delta^*) := \mathbb{P}_{\Delta^*}(\hat{B}^{(\chi)}_{n,H} > q_{W,\delta})$ *is continuous around* $\Delta^* = 0$ *since the process* X_i *from (3.8) depends continuously on* Δ^* *through* $\theta^*(i)$*. Therefore,* $\mathbb{1}_{\{\hat{B}^{(\chi)}_{n,H} > q_{W,\delta}\}}$ *is also a level* δ *test for*

$$H_0 : \Delta^* < 0 \quad vs. \quad H_1 : \Delta^* \ge 0. \tag{3.17}$$

Intuitively, one can argue that (3.17) is nearly the same as testing

$$H_0 : \Delta^* \le \epsilon \quad vs. \quad H_1 : \Delta^* > \epsilon$$

with some arbitrarily small $\epsilon > 0$*, which again is nearly the same as testing (3.16).*

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

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

which is motivated by (3.15). For the special case of a parameter change in a GARCH(1,1) model with hypotheses (2.4) and corresponding $H = (0,1,1)'$, we obtain the statement

$$\begin{aligned}&\sup_{(\tau_1,\tau_2)} \hat{B}^{(\chi)}_{n,H}(\tau_1,\tau_2)\\ &= \sqrt{n} \sup_{(\tau_1,\tau_2)\in R_{\kappa,\kappa'}} (\tau_2-\tau_1)^{\chi} (H\bar{\Sigma}_{n,\tau_1,\tau_2}H')^{-1/2} (\hat{\alpha}_1 + \hat{\beta}_1 - c) \xrightarrow{d} \sup_{(\tau_1,\tau_2)\in R_{\kappa,\kappa'}} \Big\{ \frac{B(\tau_2) - B(\tau_1)}{(\tau_2-\tau_1)^{1-\chi}} \Big\}\end{aligned}$$

for the test statistic defined in equation (2.6).

Example 3.2. (Example 3.1 continued) *From Corollary 3.1, we obtain that under* $H_0 : \alpha_1^* = 1$*:*

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

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 $\theta^*(i) = (\alpha_0^*(i), \alpha_1^*(i), \beta_1^*(i))'$,

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

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

i) $H = (0,1,1)'$ and $\alpha_0^* = 0.2$, $\alpha_1^* = 0.5$, $\beta_1^* = 0.25$,
ii) $H = (0,1,1)'$ and $\alpha_0^* = 0.3$, $\alpha_1^* = 0.39$, $\beta_1^* = 0.6$,
iii) $H = (0,1,1)'$ and $\alpha_0^* = 0.5$, $\alpha_1^* = 0.07$, $\beta_1^* = 0.91$,

and

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

We now check the behaviour 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 $\hat{q}_{W,0.90} \approx 3.031$, $\hat{q}_{W,0.95} \approx 3.285$ and the results given in Table 1 (cf. Step 4 in the algorithm in Section 2). We find that the performance of the test is quite unaffected by the choice of χ and, therefore, we do not present results for different values of χ here. We can see from the table that as the sample size increases, the coverage probabilities approach the nominal level for both $\delta = 0.90, 0.95$, which illustrates that the asymptotic results hold true already for relatively small sample sizes.

Table 1. The averaged acceptance rate under null hypothesis H_0. ($L = 30$).

	i)		ii)		iii)	
n / δ	0.90	0.95	0.90	0.95	0.90	0.95
500	0.867	0.901	0.862	0.904	0.861	0.907
1000	0.892	0.905	0.880	0.911	0.894	0.915
2000	0.899	0.923	0.887	0.915	0.881	0.943

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 exert a significant influence on the performance of the test; therefore we do not present the simulation results for different τ_1^* here. The test results of 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 drastically increases.

5. REAL-DATA APPLICATION

We now apply our test to real data. We first consider two commonly used financial risk indicators. One is the VIX, and the other is the Treasury-EuroDollar (TED) rate spread. The VIX is a weighted combination of prices for a range of options on the S&P 500 index, which reflects the market expectation of the volatility level. The TED spread is the difference between the 3-Month London Interbank Offered Rate (LIBOR) based on US dollars (https://fred.stlouisfed.org/series/USD3MTD156N) and the 3-Month Treasury Bill (https://fred.stlouisfed.org/series/DTB3),

Table 2. The 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	0.321	0.436	0.828	0.886	0.917	0.939
	1000	0.583	0.599	0.898	0.904	0.953	0.969
	2000	0.812	0.826	0.912	1.000	1.000	1.000
ii)	500	0.357	0.475	0.830	0.902	0.930	0.949
	1000	0.798	0.804	0.870	0.963	0.988	0.997
	2000	0.805	0.819	0.904	1.000	1.000	1.000
iii)	500	0.431	0.577	0.845	0.901	0.952	0.958
	1000	0.783	0.826	0.860	0.956	0.981	0.998
	2000	0.792	0.831	0.924	1.000	1.000	1.000

Table 3. The 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	0.388	0.506	0.834	0.922	0.937	0.950
	1000	0.657	0.693	0.901	0.913	0.996	0.997
	2000	0.883	0.894	0.923	1.000	1.000	1.000
ii)	500	0.401	0.701	0.875	0.932	0.951	0.962
	1000	0.801	0.837	0.900	0.994	0.999	0.999
	2000	0.938	0.972	0.995	1.000	1.000	1.000
iii)	500	0.401	0.704	0.873	0.934	0.951	0.977
	1000	0.800	0.854	0.904	0.997	0.998	0.999
	2000	0.943	0.979	0.995	1.000	1.000	1.000

which typically measures the liquidity among the inter-bank money market. The VIX is available from Yahoo Finance, and the TED spread is downloaded from the following address: https://fred.stlouisfed.org/series/TEDRATE. We adopt a daily frequency for both indices for the time span 01/07/2004 - 09/05/2018.

The VIX is often regarded as a measure of the market fear of stock investors, which is related to the cost of purchasing insurance against market downturns. We usually see that the VIX will be high in a bearish market and low in a bullish market. The TED spread represents the credit risk in the general economy. It signals how banks are willing to lend to each other, which is related to the liquidity of the markets. A high level of TED spread is a sign of low liquidity and high risk of default on inter-bank loans.

Both the VIX and the TED spread are often considered early-warning indicators of market stress. Namely, when market uncertainty is high, a temporary shock to the financial system lead to increased default or otherwise adverse effect to the global financial market, see for example as described in González-Hermosillo and Hesse (2009). Abrupt changes of the parameter values of the underlying processes are likely to be associated with this type of sudden changes of market conditions. The goal of our analysis is to discover the existence of periods of unstable behaviour of the underlying volatility pro-

Table 4. The detected significant break periods for the TED spread, the corresponding persistence parameter ($\hat{\alpha}_1 + \hat{\beta}_1$) and the test statistics; (***) means significant at both $0.95, 0.90$. Parameter estimation for $\hat{\alpha}_1$, $\hat{\beta}_1$ in brackets.

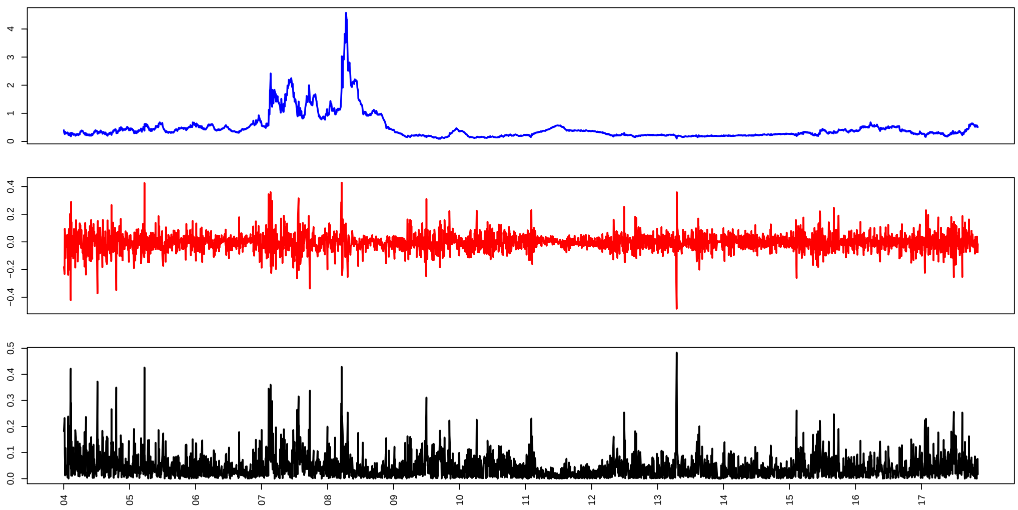
	$\hat{\tau}_1$	$\hat{\tau}_2$	in	out	test statistics
1	2007/05/09	2008/02/29	1.05 (0.15, 0.90)	0.99 (0.08, 0.91)	4.49(***)
2	2013/06/20	2014/04/11	1.07 (0.32, 0.73)	0.98 (0.04, 0.95)	4.76(***)

Table 5. The detected significant break periods for the VIX. Testing the corresponding persistence parameter ($\hat{\alpha}_1 + \hat{\beta}_1 = 0.88$) and the test statistics; (***) means significant at both $0.95, 0.90$. Parameter estimation for $\hat{\alpha}_1$, $\hat{\beta}_1$ in brackets.

	$\hat{\tau}_1$	$\hat{\tau}_2$	in	out	test statistics
1	2007/12/24	2011/03/28	0.96 (0.13, 0.83)	0.72 (0.22, 0.50)	13.01(***)
2	2011/12/12	2012/11/28	0.99 (0.05, 0.94)	0.82 (0.12, 0.70)	15.53(***)

cess. This can be helpful to decide if a government should perform an intervention based on the estimated underlying parameter values. Figure 5 shows a plot of the following adjusted series: the TED spread, the log returns and absolute log returns of the TED spread. From the plot, we observe that the returns fluctuate around the zero. During the years 2008-2009, there is a period of high volatility. We divide the data into a sequence of consecutive windows of 1000 observations each. 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.

Figure 5. The plot of TED spread in percentage (upper panel), the log difference of TED spread (middle panel) and the absolute value of TED spread (lower panel).



From the histogram and Q-Q plot of the time series in Figure 8 we observe a strong evidence of leptokurtic behaviour. We follow the suggestions as in Section 2 for the choices of tuning parameters and grid size L is chosen to be $L = 20$. In Table 4 we present the

Figure 6. QQ plot and the histogram for the daily TED spread.

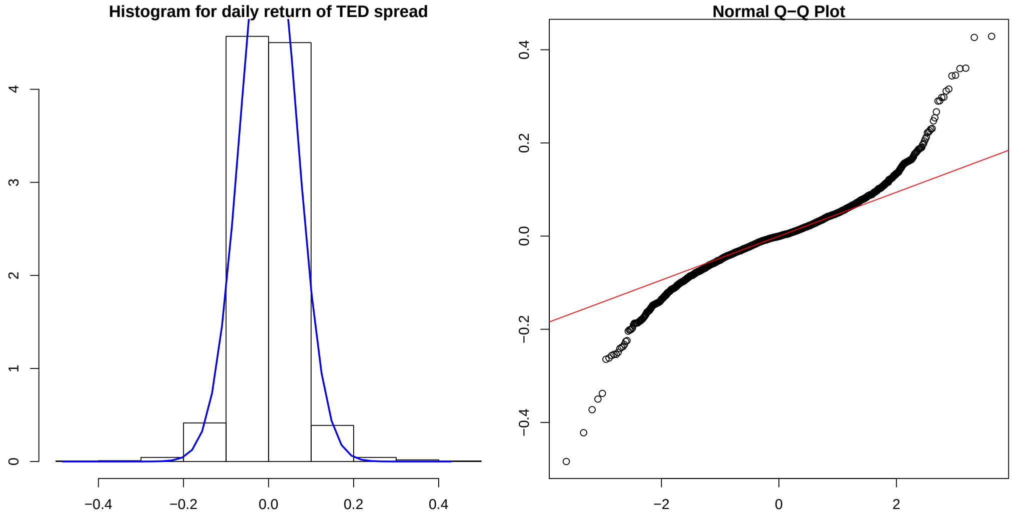
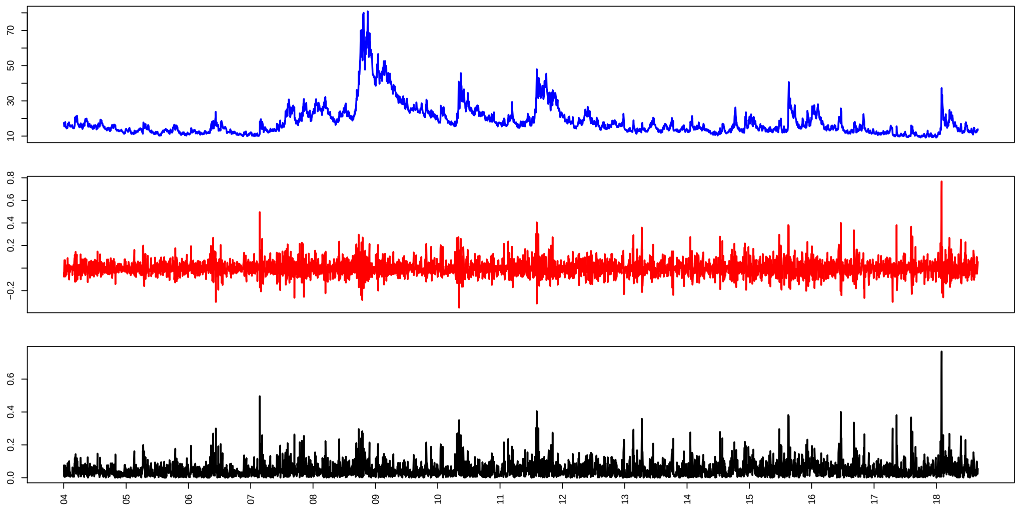


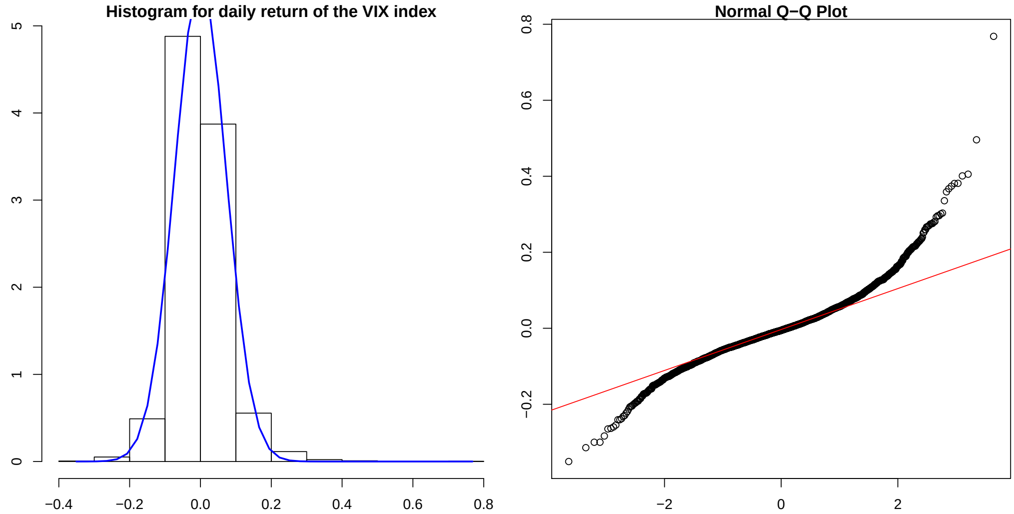
Figure 7. The plot of VIX (upper panel), the log difference of VIX (middle panel), and the absolute value of VIX (lower panel).



detected periods of the mildly explosive behaviour. The GSRWW test identifies the major financial crises such as the US subprime mortgage crisis as early as May 2007, and lasts until Feb. 2008. Furthermore, the test can detect some short-lived instability early, such as in Oct 2013 the TED spread dropped due to the worries of a potential default on the US debt.

The corresponding time series of the VIX is plotted in Figure 7. We observe that the index value increases sharply during the subprime crisis. A similar leptokurtic behaviour of the series can be found in Figure 8. We cannot detect any significant intervals against the null hypothesis of $H_0 : \alpha_1 + \beta_1 = 1$. Instead we fit a global model first using the whole

Figure 8. Q-Q plot and the histogram for the daily VIX index.



sample and test against the fitted value of the global model, i.e. $\alpha_1 + \beta_1 = c := 0.88$ (cf. the discussion after (2.4) in Section 2). We have detected two intervals of change-points, as listed in Table 5. In particular, the period starting in Dec 2007 signifies the early warning of the subprime mortgage crisis. The period starting on 2011/12/12 corresponds to the Euro debt crisis. In sum, our test can pick up the critical periods of financial crises early for both the VIX and the TED spread. Besides that, it can also successfully signify small periods of turbulence in the volatility processes of the the early-warning indicators.

Next, we test our methodology on the recent emerging Fintech markets. We gather the Bitcoin price series July 23, 2015- August 21, 2018 at a daily frequency. The data source is `https://coinmarketcap.com/currencies/bitcoin/historical-data/`. We show the returns and the absolute returns for the Bitcoin price series in Figure 9. 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 Q-Q plots and histograms in Figure 10 indicate the heavy-tailedness of the underlying distribution. We present the test results with a window of 1000 observations in Table 6. Again the log returns are stationary in all the windows (by results of ADF tests) and serial correlation is taken out by fitting an ARMA process in advance. We apply our test to the obtained residuals, which 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. We have chosen Bitcoin as an important additional study as the Fintech markets are known to behave independently with respect to the conventional financial market. Bitcoin is not controlled by any government, and speculators can use our test results for abnormal regimes of Bitcoin as indicators of the market sentiment.

Figure 9. The plot of Bitcoin price (upper panel), the log difference of Bitcoin (middle panel) and the absolute returns (lower panel).

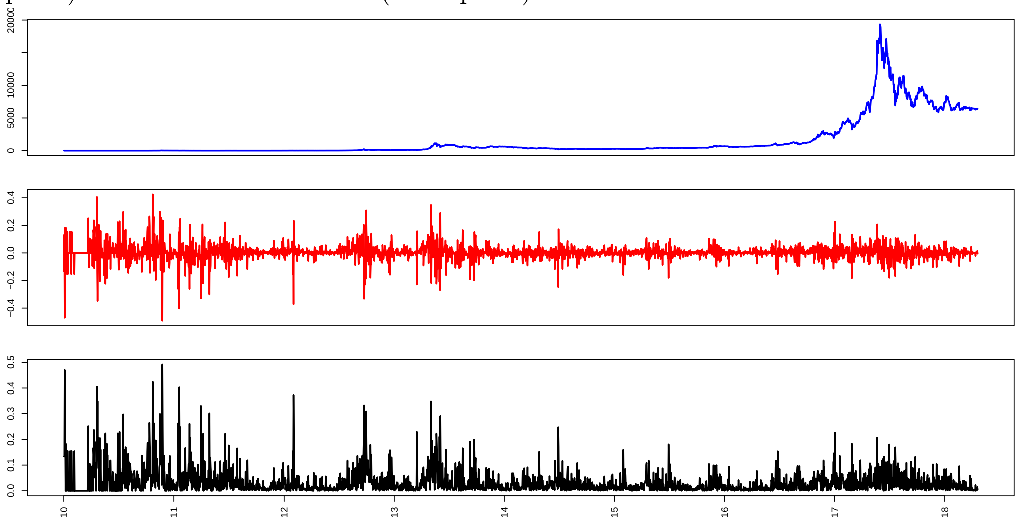


Figure 10. QQ plot and the histogram for the Bitcoin returns

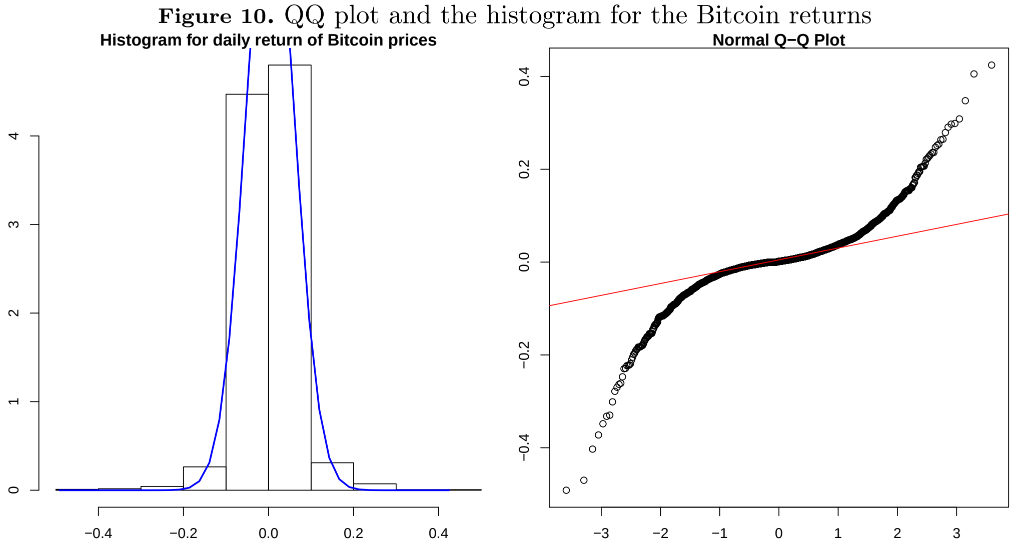


Table 6. The detected significant break periods for the Bitcoin log returns, the corresponding persistence parameter $(\hat{\alpha}_1 + \hat{\beta}_1)$ and the test statistics. (***) means significant at both $0.95, 0.90$. Parameter estimation for $\hat{\alpha}_1$, $\hat{\beta}_1$ in brackets.

	$\hat{\tau}_1$	$\hat{\tau}_2$	in	out	test statistics
1	2011/03/26	2011/08/23	1.14 (0.69, 0.45)	0.98 (0.19, 0.79)	11.68(***)
2	2013/04/14	2013/10/31	1.18 (0.55, 0.68)	0.99 (0.04, 0.95)	8.13(***)
3	2016/01/09	2017/12/09	1.05 (0.29, 0.76)	0.99 (0.04, 0.95)	19.04(***)

6. CONCLUSION

In this paper, we propose a uniform test for a mildly explosive GARCH process with double-supreme statistics. Theoretical results about the uniform parameter consistency and asymptotic distribution of the test statistics are provided. Our test is easy to implement and can help to effectively identify mildly explosive periods with good sizes and power. The quality of the test is discussed via a simulation study. We applied our procedure to real data time series as the TED spread, the VIX and the Bitcoin price series, and tracked their corresponding volatile periods. Further work may extend the algorithm to online procedures, allowing for real time detection of breaks.

References

Bardet, J.-M., W. Kengne, O. Wintenberger, et al. (2012). Multiple breaks detection in general causal time series using penalized quasi-likelihood. Electronic Journal of Statistics 6, 435–477.

Berkes, I., L. Horváth, and P. Kokoszka (2003). GARCH processes: structure and estimation. Bernoulli 9(2), 201–227.

Bloom, N. (2007). Uncertainty and the dynamics of R&D. American Economic Review 97(2), 250–255.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11(2), 143–172.

Bougerol, P. and N. Picard (1992a). Stationarity of GARCH processes and of some nonnegative time series. Journal of Econometrics 52(1-2), 115–127.

Bougerol, P. and N. Picard (1992b). Strict stationarity of generalized autoregressive processes. Ann. Probab. 20(4), 1714–1730.

Davis, R. A., T. C. Lee, and G. A. Rodriguez-Yam (2008). Break detection for a class of nonlinear time series models. Journal of Time Series Analysis 29(5), 834–867.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica: Journal of the Econometric Society, 987–1007.

Fan, J., L. Qi, and D. Xiu (2014). Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. Journal of Business & Economic Statistics 32(2), 178–191.

Francq, C. and J.-M. Zakoïan (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. Bernoulli 10(4), 605–637.

Francq, C. and J.-M. Zakoïan (2012). Strict stationarity testing and estimation of explosive and stationary generalized autoregressive conditional heteroscedasticity models. Econometrica 80(2), 821–861.

Fryzlewicz, P. and S. Subba Rao (2014). Multiple-change-point detection for autoregressive conditional heteroscedastic processes. Journal of the Royal Statistical Society: series B (statistical methodology) 76(5), 903–924.

González-Hermosillo, M. B. and H. Hesse (2009). Global Market Conditions and Systemic Risk. Number 9-230. International Monetary Fund.

Hafner, C. (2018). Testing for bubbles in cryptocurrencies with time-varying volatility.

Harvey, D. I., S. J. Leybourne, and Y. Zu (2018). Testing explosive bubbles with time-varying volatility. Econometric Reviews.

Jensen, S. T. and A. Rahbek (2004). Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. Econometrica 72(2), 641–646.

Jurado, K., S. C. Ludvigson, and S. Ng (2015). Measuring uncertainty. American Economic Review 105(3), 1177–1216.

Lee, S.-W. and B. E. Hansen (1994). Asymptotic theory for the GARCH (1, 1) quasi-maximum likelihood estimator. Econometric theory 10(1), 29–52.

Nelson, D. B. (1990). Stationarity and persistence in the GARCH (1, 1) model. Econometric theory 6(3), 318–334.

Paolella, M. S. (2018). Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH. Chichester.

Peng, L. and Q. Yao (2003). Least absolute deviations estimation for arch and garch models. Biometrika 90(4), 967–975.

Phillips, P. C., Y. Wu, and J. Yu (2011). Explosive behavior in the 1990s Nasdaq: When did exuberance escalate asset values? International economic review 52(1), 201–226.

van der Vaart, A. W. (1998). Asymptotic statistics, Volume 3 of Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

Wu, W. B. and W. Min (2005). On linear processes with dependent innovations. Stochastic Processes and their Applications 115(6), 939 – 958.

Wu, W. B. and X. Shao (2004). Limit theorems for iterated random functions. Journal Applied Probabability 41(2), 425–436.

Wu, W. B. and Z. Zhou (2011). Gaussian approximations for non-stationary multiple time series. Statist. Sinica 21(3), 1397–1413.

Zhang, D. and W. B. Wu (2017). Gaussian Approximation for High Dimensional Time Series. Ann. Statist. 45(5), 1895–1919.

A. TECHNICAL PROOFS AND LEMMATA

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, we define the weighted seminorm

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

For some random variable Z, we 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 ζ_i, $i\in\mathbb{Z}$ used in the model definition (3.8), let $\mathcal{F}_i := (\zeta_i,\zeta_{i-1},...)$. 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. Wu and Shao (2004)) as

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

where $X_i^* = h(\mathcal{F}_i^*)$ and $\mathcal{F}_i^* := (\zeta_i,...,\zeta_1,\zeta_0^*,\zeta_{-1},\zeta_{-2},...)$ with ζ_0^* being an independent copy of ζ_0. It is worth noting that in the following context we add the superscript X as $\delta_q^X(i)$ for the dependence measure of the process X.

A.1. Existence of GARCH models

PROPOSITION 1.1. (EXISTENCE OF THE GARCH MODEL) *If Assumption 3.1 holds, then the following statements hold true.*

(i) *(3.8) has a unique stationary solution* $X_i^2 = H(\mathcal{F}_i)$, $i\in\mathbb{Z}$.
(ii) *There exists* $q>0$ *with* $\|X_0^2\|_q \le D$ *and* $\delta_q^{X^2}(k) = O(c^k)$ *(recall that* $\delta_q^{X^2}(k)$ *is the functional dependence measure of* X^2 *at the lag* k*) for some* $0<c<1$.
(iii) $\lambda_{max}(B(\theta)) < 1$, *where*

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

PROOF. (PROOF OF PROPOSITION 1.1:) Following the proof of Lemma 2.3 in Berkes et al. (2003), 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 $[0,\delta)\to[0,\infty)$, $s(t,\theta) = \mathbb{E}|A_m(\theta)A_{m-1}(\theta)\dots A_1(\theta)|_2^t$ 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

$$\mathbb{E}|A_m(\theta)A_{m-1}(\theta)\dots A_1(\theta)|_2^q = s(q,\theta) < 1. \tag{1.18}$$

Define

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

Following Section 3.1 in Wu and Min (2005), the model (3.8) admits the representation

$$P_i(\theta) = A_i(\theta)P_{i-1}(\theta) + a_i(\theta). \tag{1.19}$$

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 $W_n(y,\theta) := G_{\zeta_n,\theta}\circ G_{\zeta_{n-1},\theta}\circ ...\circ G_{\zeta_1,\theta}(y)$. Then we have

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

Using the submultiplicativity of $|\cdot|_2$, we therefore have with (1.18) and some suitable constant $C > 0$:

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

By Theorem 2 in Wu and Shao (2004), 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 Francq and Zakoïan (2004).

A.2. Proofs for asymptotic theory

Observe that

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

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. Francq and Zakoïan (2004), Proof of Theorem 2.1).

To prove Theorem 3.1, we introduce some notation. For some sequence of real-valued random variables W_n, we write $W_n \xrightarrow{p} \infty$ if for each $M\in\mathbb{N}$, $\mathbb{P}(W_n < M)\to 0$ $(n\to\infty)$.

Lemma 1.1. *Let Assumption 3.1 hold. Let $\kappa > 0$, and $R := \{(\tau_1,\tau_2)\in[0,1]^2 : \tau_1 < \tau_2, |\tau_1-\tau_2|\ge\kappa\}$. 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*

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

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

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

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

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

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

$$\sup_{\tau_2\in[\kappa,1]}\Big|\frac{S_{\lfloor n\tau_2\rfloor}}{\lfloor n\tau_2\rfloor}-\mathbb{E}S_1\Big|\to 0\quad a.s. \tag{1.20}$$

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

$$\sup_{\tau_1\in[0,1]}\Big|\tau_1\cdot\Big(\frac{S_{\lfloor n\tau_1\rfloor}}{\lfloor n\tau_1\rfloor}-\mathbb{E}S_1\Big)\Big|\to 0 \quad a.s. \tag{1.21}$$

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

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

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 $\le\varepsilon$.]

With the decomposition

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

and (1.20), (1.21), $\sup_{\tau_2\in[\kappa,1]}|\frac{\lfloor n\tau_2\rfloor}{n}-\tau_2|\le n^{-1}$ we obtain

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

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

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

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

Proof. (Proof of Theorem 3.1) We make use of some results obtained in the proof of Theorem 2.1. in Francq and Zakoïan (2004). It was shown therein that $L(\theta):=\mathbb{E}\ell(X_i^2,Y_i,\theta)$ fulfills

$$\mathbb{E}|\ell(X_i^2,Y_i,\theta^*)|<\infty, \qquad \forall\theta\ne\theta^*: \quad L(\theta)>L(\theta^*). \tag{1.22}$$

Let $k\in\mathbb{N}$. Use the notation from Lemma 1.1. Let $\theta\ne\theta^*$. By Beppo-Levi's theorem, we have

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

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

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

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

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

$$\begin{aligned}
0 &\geq \inf_{(\tau_1,\tau_2)\in R} \frac{1}{\tau_2 - \tau_1}\big\{L^c_{n,\tau_1,\tau_2}(\hat{\theta}_{n,\tau_1,\tau_2}) - L^c_{n,\tau_1,\tau_2}(\theta^*)\big\} \\
&= \inf_{i=1,\ldots,l} \inf_{(\tau_1,\tau_2)\in R} \inf_{\theta'\in V_{k(\theta_i)}(\theta_i)} \frac{1}{\tau_2-\tau_1}\{L^c_{n,\tau_1,\tau_2}(\theta') - L^c_{n,\tau_1,\tau_2}(\theta^*)\} \\
&\geq \inf_{i=1,\ldots,l} \inf_{(\tau_1,\tau_2)\in R} \inf_{\theta'\in V_{k(\theta_i)}(\theta_i)} \frac{L^c_{n,\tau_1,\tau_2}(\theta')}{\tau_2-\tau_1} - \sup_{(\tau_1,\tau_2)\in R} \frac{L^c_{n,\tau_1,\tau_2}(\theta^*)}{\tau_2-\tau_1} \qquad (1.23)
\end{aligned}$$

We furthermore have:

$$\begin{aligned}
&\inf_{(\tau_1,\tau_2)\in R} \inf_{\theta'\in V_{k(\theta_i)}(\theta_i)} \frac{L^c_{n,\tau_1,\tau_2}(\theta')}{\tau_2-\tau_1} \\
&\geq \inf_{(\tau_1,\tau_2)\in R} \inf_{\theta'\in V_{k(\theta_i)}(\theta_i)} \frac{L_{n,\tau_1,\tau_2}(\theta')}{\tau_2-\tau_1} - \kappa\cdot \sup_{(\tau_1,\tau_2)\in R} \sup_{\theta'\in\Theta} |L^c_{n,\tau_1,\tau_2}(\theta') - L_{n,\tau_1,\tau_2}(\theta')| \\
&\geq \inf_{(\tau_1,\tau_2)\in R} \frac{1}{n(\tau_2-\tau_1)} \sum_{i=\lfloor n\tau_1\rfloor}^{\lfloor n\tau_2\rfloor} W_i^{(k(\theta_i))}(\theta_i) \\
&\qquad -\kappa\cdot \sup_{(\tau_1,\tau_2)\in R} \sup_{\theta'\in\Theta} |L^c_{n,\tau_1,\tau_2}(\theta') - L_{n,\tau_1,\tau_2}(\theta')|. \qquad (1.24)
\end{aligned}$$

By Lemma 1.2(i),

$$\mathbb{P}\Big(\sup_{(\tau_1,\tau_2)\in R} \sup_{\theta'\in\Theta} |L^c_{n,\tau_1,\tau_2}(\theta') - L_{n,\tau_1,\tau_2}(\theta')| > \frac{\delta}{8}\Big) = o(1). \qquad (1.25)$$

By Lemma 1.1(i),

$$\mathbb{P}\Big(\inf_{i=1,\ldots,l} \inf_{(\tau_1,\tau_2)\in R} \frac{1}{n(\tau_2-\tau_1)} \sum_{i=\lfloor n\tau_1\rfloor}^{\lfloor n\tau_2\rfloor} W_i^{(k(\theta_i))}(\theta_i) \leq L(\theta^*) + \frac{\delta}{2}\Big) = o(1). \qquad (1.26)$$

By Lemma 1.1(ii),

$$\mathbb{P}\Big(\sup_{(\tau_1,\tau_2)\in R} \frac{L^c_{n,\tau_1,\tau_2}(\theta^*)}{\tau_2-\tau_1} \geq L(\theta^*) + \frac{\delta}{8}\Big) = o(1). \qquad (1.27)$$

Inserting (1.25), (1.26) into (1.24) and afterwards using (1.27), we have

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

showing the assertion.

Proof. (Proof of Theorem 3.2) 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 3.1, we have

$$\sup_{(\tau_1,\tau_2)\in R} |\hat{\theta}_{n,\tau_1,\tau_2} - \theta^*|_1 \to 0 \quad a.s. \qquad (1.28)$$

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

$$\begin{aligned}\hat{\theta}_{n,\tau_1,\tau_2} - \theta^* &= -[\nabla_\theta^2 L^c_{n,\tau_1,\tau_2}(\bar{\theta})]^{-1}\nabla_\theta L^c_{n,\tau_1,\tau_2}(\theta^*)\\ &= -[(\tau_2-\tau_1)V(\theta^*) + T_{n,\tau_1,\tau_2}(\bar{\theta}_{n,\tau_1,\tau_2})]^{-1}\nabla_\theta L^c_{n,\tau_1,\tau_2}(\theta^*),\end{aligned} \tag{1.29}$$

where $T_{n,\tau_1,\tau_2}(\bar{\theta}_n) = \nabla_\theta^2 L^c_{n,\tau_1,\tau_2}(\bar{\theta}_n) - (\tau_2-\tau_1)V(\theta^*)$ and $\bar{\theta}_{n,\tau_1,\tau_2} \in \Theta$ with $|\bar{\theta}_{n,\tau_1,\tau_2} - \theta^*|_1 \le |\hat{\theta}_{n,\tau_1,\tau_2} - \theta^*|_1$. By Lemma 1.2 and Lemma 1.4 and $|\frac{\lfloor n\tau_2\rfloor - \lfloor n\tau_1\rfloor}{n} - (\tau_1-\tau_2)| \le 2n^{-1}$, we have

$$\sup_{(\tau_1,\tau_2)\in R} |T_{n,\tau_1,\tau_2}(\bar{\theta}_{n,\tau_1,\tau_2})|_1 \le \sup_{(\tau_1,\tau_2)\in R} |V(\bar{\theta}_{n,\tau_1,\tau_2}) - V(\theta^*)|_1 + O_p(\log(n)^{3/2}n^{-1/2}). \tag{1.30}$$

By Lemma 1.5(ii) applied to $p = q$ and $\bar{\Theta}$, we obtain $\iota > 0$, $C > 0$ and $\rho \in (0,1)$ such that for $|\theta - \theta^*|_1 < \iota$,

$$\begin{aligned}|V(\theta) - V(\theta^*)|_1 &\le C(1 + \||Y_i|^q_{(\rho^j)_j,q}\|_1)(1 + \mathbb{E}\zeta_0^2)\cdot|\theta-\theta^*|_1\\ &\le C(1 + \frac{D^q}{1-\rho})(1+\mathbb{E}\zeta_0^2)\cdot|\theta-\theta^*|_1 =: \tilde{C}\cdot|\theta-\theta^*|_1.\end{aligned} \tag{1.31}$$

Since $\mathbb{E}\nabla_\theta \ell(Y_i, X_i^2, \theta^*) = 0$ (cf. Proposition 1.2(iii)), Lemma 1.2 and Lemma 1.4, we have

$$\sup_{(\tau_1,\tau_2)\in R} |\nabla_\theta L^c_{n,\tau_1,\tau_2}(\theta^*)|_1 = O_p(\log(n)^{3/2}n^{-1/2}). \tag{1.32}$$

Inserting (1.28) into (1.31), 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 (1.29) and (1.32) we obtain

$$\sup_{(\tau_1,\tau_2)\in R} |\hat{\theta}_{n,\tau_1,\tau_2} - \theta^*|_1 = O_p(\log(n)^{3/2}n^{-1/2}). \tag{1.33}$$

By (1.29), we have

$$\begin{aligned}&\left|\hat{\theta}_{n,\tau_1,\tau_2} - \theta^* + ((\tau_2-\tau_1)V(\theta^*))^{-1}\cdot\nabla_\theta L^c_{n,\tau_1,\tau_2}(\theta^*)\right|_2\\ &\le |((\tau_2-\tau_1)V(\theta^*) + T_{n,\tau_1,\tau_2}(\bar{\theta}_{n,\tau_1,\tau_2}))^{-1}|_1\cdot|T_{n,\tau_1,\tau_2}(\bar{\theta}_{n,\tau_1,\tau_2})|_1\\ &\qquad\times|((\tau_2-\tau_1)V(\theta^*))^{-1}\nabla_\theta L^c_{n,\tau_1,\tau_2}(\theta^*)|_1.\end{aligned} \tag{1.34}$$

Using (1.30), (1.31) and (1.33), 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 (1.32) into (1.34), we obtain

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

Since $\sup_{(\tau_1,\tau_2)\in R}|\nabla_\theta L^c_{n,\tau_1,\tau_2}(\theta^*) - \nabla_\theta L_{n,\tau_1,\tau_2}(\theta^*)|_1 = O_p(n^{-1})$ by Lemma 1.2, the proof is finished.

Proof. (Proof of Theorem 3.3) Let $R := \{(\tau_1,\tau_2)\in[0,1]^2 : \tau_1 < \tau_2, |\tau_1-\tau_2| \ge \kappa\}$. By Lemma 1.3 (applied with $M = 2 + \frac{a'}{4}$, $a = \frac{a'}{4}$), we obtain $C > 0$, $\rho \in (0,1)$ and $\iota > 0$ such that

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

and thus (component-wise) $\Delta_M^{\nabla_\theta\ell(Z,\theta^*)}(k) \le \frac{C}{1-\rho}\rho^k$. 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 1.2(iii), $\mathbb{E}W_i = 0$ and

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

By Corollary 1 in Wu and Zhou (2011), there exists a richer probability space and i.i.d. $V_1, V_2, \ldots \sim N(0, I_{(r+s+1)\times(r+s+1)})$, a process $(\hat{S}(i))_{i=1,\ldots,n}$ and $S^0(i) = \sum_{j=1}^{i} V_j$ such that $(S(i))_{i=1,\ldots,n} \overset{d}{=} (\hat{S}(i))_{i=1,\ldots,n}$ and

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

With Theorem 3.2 we obtain:

$$\sup_{(\tau_1,\tau_2)\in R} \big|\sqrt{n}(\tau_2 - \tau_1)(\hat{\theta}_{n,\tau_1,\tau_2} - \theta^*) - n^{-1/2}(S(\lfloor n\tau_2\rfloor) - S(\lfloor n\tau_1\rfloor))\big| = O_p(\log(n)^3 n^{-1/2}). \tag{1.35}$$

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

$$n^{-1/2}(S(\lfloor n\tau_2\rfloor) - S(\lfloor n\tau_1\rfloor)) \overset{d}{=} n^{-1/2}(\hat{S}(\lfloor n\tau_2\rfloor) - \hat{S}(\lfloor n\tau_1\rfloor)) \tag{1.36}$$

and

$$\begin{aligned}&\sup_{(r_1,r_2)\in R} \big|n^{-1/2}(\hat{S}(\lfloor n\tau_2\rfloor) - \hat{S}(\lfloor n\tau_1\rfloor)) - \Sigma^{1/2}\cdot n^{-1/2}(S^0(\lfloor n\tau_2\rfloor) - S^0(\lfloor n\tau_1\rfloor))\big|\\ &= O_p(n^{\frac{1}{\min\{M,4\}} - \frac{1}{2}} \log(n)^{3/2}).\end{aligned} \tag{1.37}$$

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

$$n^{-1/2}\Sigma^{1/2}\big\{S^0(\lfloor n\tau_2\rfloor) - S^0(\lfloor n\tau_1\rfloor)\big\} \overset{d}{\to} \Sigma^{1/2}\big\{B(\tau_2) - B(\tau_1)\big\}. \tag{1.38}$$

Combining (1.35), (1.36), (1.37) and (1.38), we obtain the result.

Proof. (Proof of Proposition 3.1) Using similar arguments as in the proof of Theorem 3.1 (now with $1 - (\tau_2 - \tau_1) \geq \kappa$ instead of $\tau_2 - \tau_1 \geq \kappa$), we obtain

$$\sup_{(\tau_1,\tau_2)\in \bar{R}_\kappa} |\bar{\theta}_{n,\tau_1,\tau_2} - \theta^*| \overset{p}{\to} 0. \tag{1.39}$$

(i) By Lemma 1.2 and Lemma 1.4 and $|\frac{1-(\lfloor n\tau_2\rfloor - \lfloor n\tau_1\rfloor)}{1-(\tau_2-\tau_1)} - 1| \leq n^{-1}$, we have

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

By Lipschitz continuity of $V(\cdot)$ (cf. (1.31)) and (1.39), 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 $\bar{\Theta} \subset int(\Theta)$ be some compact set. Using Lemma 1.5(ii) and (1.58), 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),

$$\begin{aligned}&\sup_{\theta,\tilde{\theta}\in\bar{\Theta}, |\theta-\tilde{\theta}|_1<\iota} |\tilde{g}_{\tilde{\theta}}(\zeta,y,\theta) - \tilde{g}_{\tilde{\theta}}(\zeta,y',\theta)|\\ &\leq C(1 + |y|_{(\rho^j)_j,2p}^{2p} + |y'|_{(\rho^j)_j,2p}^{2p})|y - y'|_{(\rho^j)_j,p}^{p}(1+\zeta^2)^2.\end{aligned} \tag{1.40}$$

and

$$\sup_{\theta,\theta',\tilde{\theta}\in\bar{\Theta},|\theta-\tilde{\theta}|_1<\iota,|\theta'-\tilde{\theta}|_1<\iota} \frac{|\tilde{g}_{\tilde{\theta}}(\zeta,y,\theta)-\tilde{g}_{\tilde{\theta}}(\zeta,y,\theta')|}{|\theta-\theta'|_1} \le C(1+|y|^p_{(\rho^j)_j,p})(1+\zeta^2)^2. \quad (1.41)$$

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) $\sup_{\theta\in\Theta}|\nabla_\theta\ell(x,0,\theta)|\le C(1+|x|)$. With Lemma 1.5(iii) we conclude that (component-wise) $\sup_{\theta\in\Theta}|\nabla_\theta\ell(x,y,\theta)|\le C(1+|y|^2_{(\rho^j)_j,1})(1+|x|)\cdot|y|_{(\rho^j)_j,1}$. Using again Lemma 1.5(iii), we obtain

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

This shows that g has similar properties as $\nabla^2_\theta\ell$, but with factors $(1+\zeta^2)^2$ in (1.40), (1.41) instead of $(1+\zeta^2)$. Therefore, we obtain the same result as in (i) under the stated moment condition.

A.3. Technical lemmata

From Francq and Zakoïan (2004) (the proof of Theorem 2.2, part (ii) therein), we directly obtain (ii),(iii) of the following Proposition.

PROPOSITION 1.2. *(Properties of $I(\theta)$, $V(\theta)$) Let Assumption 3.1 hold. Assume that $\mu_4:=\mathbb{E}\zeta_0^4<\infty$. Then the following statements hold true.*

(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) $\mathbb{E}\nabla_\theta\ell(X_i^2,Y_i,\theta^*)=0$.

PROOF. (PROOF OF PROPOSITION 1.2) (i) By Proposition 1.1, there exists $q>0$ with $\|X_0^2\|_q<\infty$. From the bounds (1.58) (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 Francq and Zakoïan (2004), see the proof step (ii) of Theorem 2.2 (the missing $\frac{1}{2}$ is due to the different formulation of the likelihood).

LEMMA 1.2. (NEGLIGIBILITY OF TRUNCATION) *Let Assumption 3.1 hold. Then for $g=\nabla^l_\theta\ell$, $l=0,1,2$ it holds that*

(i)

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

(ii)

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

PROOF. (PROOF OF LEMMA 1.2) Note that for arbitrary $0<\tilde{q}\le\min\{q,1\}$ and random variables $Z_i = (Z_{i1}, Z_{i2}, ...)$ with $\|Z_i\|_q \le D$ it holds that

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

Let

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

By Lemma 1.5(iii), we have with Hölder's inequality for $0 < q' \le q$ chosen such that $0 < q'(l+3) \le 1$:

$$\begin{aligned}
&\|W_i\|_{q'} \\
&= \Big\| \sup_{\theta\in\Theta} |g(X_i^2, Y_i^c, \theta) - g(X_i^2, Y_i, \theta)| \Big\|_{q'} \\
&\le C(1 + \||Y_i|_{(\rho^j)_j,1}\|_{q'(l+3)}^{l+1} + \||Y_i^c|_{(\rho^j)_j,1}\|_{q'(l+3)}^{l+1}) \cdot (1 + \|X_i^2\|_{q'(l+3)}) \||y - y'|_{(\rho^j)_j,1}\|_{q'(l+3)} \\
&\le C(1 + 2\tilde{D}(q')^{l+1})(1 + D) \cdot \Big(\sum_{j=i}^{\infty} \rho^{q'(l+3)j}\|X_j^2\|_{q'(l+3)}^{q'(l+3)}\Big)^{1/(q'(l+3))} \\
&\le C(1 + 2\tilde{D}(q')^{l+1})(1 + D)\tilde{D}(q'(l+3))\rho^i =: \tilde{C} \cdot \rho^i. \qquad (1.42)
\end{aligned}$$

Therefore, we have

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

giving the result.
(ii) It holds that

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

In the following we show that $W_i \to 0$ $(i \to \infty)$ a.s.. Then the assertion follows with a Cesaro sum argument. Let $\varepsilon > 0$ be arbitrary. Then with Markov's inequality and (1.42),

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

showing $W_i \to 0$ 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 1.3. (DEPENDENCE MEASURES OF $\nabla_\theta \ell$, $\nabla_\theta^2 \ell$) *Let Assumption 3.1 hold. Let* $M \ge 1$. *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*

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

PROOF. (PROOF OF LEMMA 1.3) 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 $\frac{(M+3\kappa)}{\kappa}p \le q$. By Hoelder's inequality ($\frac{M}{M+3\kappa} + \frac{\kappa}{M+3\kappa} + \frac{2\kappa}{M+3\kappa} = 1$) and Lemma 1.5(ii) there exists $\iota > 0$, $C > 0$, $\rho \in (0,1)$ such that

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

where $\tilde{C} := C(1 + 2\frac{D^{2p}}{1-\rho})(1 + \|\zeta^2\|_{M+3\kappa})$. By Proposition 1.1(ii), it holds that $\delta_q^{X^2}(k) = O(c^k)$, which finishes the proof.

In the following we make use of results from Zhang and Wu (2017). Therefore we have to define $\Delta_q^Z(m) := \sum_{k=m}^{\infty} \delta_q^Z(k)$ and $\|Z\|_{q,\alpha} := \sup_{m\ge 0}(m+1)^{\alpha}\Delta_q^Z(m)$.

LEMMA 1.4. *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*

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

PROOF. (PROOF OF LEMMA 1.4) Let $\iota > 0$ (is chosen below). Let $S_j(\theta) := \sum_{i=1}^{j} \big\{ g(X_i^2, Y_i, \theta) - \mathbb{E}g(X_i^2, Y_i, \theta) \big\}$, $j = 1, ..., n$. For fixed $n \in \mathbb{N}$, choose $d \in \mathbb{N}$ such that $2^{d-1} \le n \le 2^d$. For $i = 0, 1, ..., d-1$, define

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

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

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

Note that

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

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

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{|\theta-\theta^*|_1<\iota}\sup_{r\in[0,1]}\Big|\sum_{i=1}^{\lfloor nr\rfloor}\big\{g(X_i^2,Y_i,\theta)-\mathbb{E}g(X_i^2,Y_i,\theta)\big\}\Big|>Q(n\log(n)^3)^{1/2}\Big)\\
&\le\sum_{i=0}^{d-1}\mathbb{P}\Big(\sup_{|\theta-\theta^*|_1<\iota}\Phi_i(\theta)>\frac{Q(n\log(n)^3)^{1/2}}{d}\Big)\\
&\le\sum_{i=0}^{d-1}2^{d-i}\cdot\mathbb{P}\Big(\sup_{|\theta-\theta^*|_1<\iota}|S_{2^i}(\theta)|>\frac{Q(n\log(n)^3)^{1/2}}{d}\Big).
\end{aligned}\tag{1.43}$$

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

Apply Lemma 1.5(ii) with $p=q$ and Lemma 1.3 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 $|\theta-\theta'|_1\le n^{-1}$.

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

$$\begin{aligned}
&\mathbb{P}\Big(\sup_{|\theta-\theta^*|_1<\iota}|S_m(\theta)|>\frac{Q(n\log(n)^3)^{1/2}}{d}\Big)\\
&\le\mathbb{P}\Big(\sup_{\theta\in\Theta_n,|\theta-\theta^*|_1<\iota}|S_m(\theta)|>\frac{Q(n\log(n)^3)^{1/2}}{2d}\Big)\\
&\qquad+\mathbb{P}\Big(\sup_{\theta,\theta'\in\Theta,|\theta-\theta'|_1\le n^{-1},|\theta-\theta^*|_1<\iota,|\theta'-\theta^*|_1<\iota}|S_m(\theta)-S_m(\theta')|>\frac{Q(n\log(n)^3)^{1/2}}{2d}\Big)\\
&=:I_n+II_n.
\end{aligned}\tag{1.44}$$

By Lemma 1.3 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

$$W_{M,\alpha}:=\|\sup_{|\theta-\theta^*|_1<\iota}|g(Z,\theta)|\|_{M,\alpha}=\sup_{m\ge0}(m+1)^\alpha\Delta_M^{\sup_{|\theta-\theta^*|_1<\iota}|g(Z,\theta)|}(m)<\infty,$$

and

$$W_{2,\alpha}:=\sup_{|\theta-\theta^*|_1<\iota}\|g(Z,\theta)\|_{2,\alpha}=\sup_{m\ge0}(m+1)^\alpha\sup_{|\theta-\theta^*|_1<\iota}\Delta_2^{g(Z,\theta)}(m)<\infty.$$

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

By applying Theorem 6.2 of Zhang and Wu (2017) with $q=M$ to $(g(Z_i,\theta))_{\theta\in\tilde{\Theta}_n,|\theta-\theta^*|_1<\iota}$, we have with some constants $C_\alpha > 0$:

$$
\begin{aligned}
I_n &= \mathbb{P}\Big(\sup_{\theta\in\Theta_n, |\theta-\theta^*|_1<\iota} |S_m(\theta)| > \frac{Q(n\log(n)^3)^{1/2}}{2d}\Big)\\
&\le \frac{C_\alpha m\cdot l^{M/2} W_{M,\alpha}^M}{(Q/2d)^M (n^{1/2}\log(n)^{3/2})^M} + C_\alpha \exp\Big(- \frac{C_\alpha (Q/2d)^2 n \log(n)^3}{m W_{2,\alpha}^2}\Big)\\
&\le O(m\cdot n^{-\frac{M}{2}} + n^{-2}), \qquad (1.45)
\end{aligned}
$$

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

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 1.5(ii):

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

Thus

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

With Markov's inequality, we therefore obtain

$$
II_n \le \frac{2\tilde{C} m}{(Q/2d) n^{3/2}\log(n)^{3/2}}. \qquad (1.46)
$$

Inserting (1.45) and (1.46) into (1.44) and then into (1.43), we obtain with some constant $\tilde{C} > 0$:

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

showing the assertion.

A.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

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

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 1.5(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 1.5. *$\theta \mapsto \sigma(y, \theta)$ and $\theta \mapsto \ell(x, y, \theta)$ are three times continuously differentiable. Let $\bar{\Theta} \subset int(\Theta)$ 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 $l = 0, 1, 2, 3$:

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

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

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

and

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

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

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

Proof. (Proof of Lemma 1.5) (i) From Proposition (1.1)(iii) we obtain that the following explicit representation holds, where $F(y, \theta) := (\alpha_0 + \sum_{j=1}^r \alpha_j y_j, 0, ..., 0)'$:

$$\sigma(y, \theta)^2 = \sum_{k=0}^\infty \big(B(\theta)^k F(y_{k\to}, \theta)\big)_1, \tag{1.47}$$

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

$$\begin{aligned}\sigma(y, \theta)^2 &= \alpha_0 \sum_{k=0}^\infty (B(\theta))_{11} + \sum_{j=1}^r \alpha_j \sum_{k=0}^\infty (B(\theta)^k)_{11} y_{k+j} \\ &\overset{k'=k+j}{=} \alpha_0 \sum_{k=0}^\infty (B(\theta))_{11} + \sum_{k'=1}^\infty \Big(\sum_{j=1}^r \alpha_j (B(\theta)^{k'-j})_{11} \mathbb{1}_{k'\ge j}\Big) y_{k'} \\ &=: c_0(\theta) + \sum_{k'=1}^\infty c_{k'}(\theta) y_j.\end{aligned} \tag{1.48}$$

From Proposition (1.1)(iii) we obtain that $c_j(\theta) \geq 0$ satisfies

$$\sup_{\theta\in\Theta} |c_k(\theta)| \leq C \cdot \rho^k \tag{1.49}$$

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

$$\nabla_\theta^k(\sigma(y,\theta)^2) = \nabla_\theta^k c_0(\theta) + \sum_{k=1}^{\infty} \nabla_\theta^k c_k(\theta) \cdot y_k, \qquad k \in \{0,1,2,3\}, \tag{1.50}$$

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 Francq and Zakoïan (2004). Note that for $j = 0, ..., r$, we have $\partial_{\alpha_j} F(y,\theta) \leq \frac{1}{\alpha_j} F(y,\theta)$ and thus

$$\partial_{\alpha_j} c_k(\theta) \leq \frac{1}{\alpha_j} c_k(\theta). \tag{1.51}$$

For $j = 1, ..., s$, 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

$$\partial_{\beta_j} c_k(\theta) \leq \frac{1}{\beta_j} k \cdot c_k(\theta). \tag{1.52}$$

From (1.51) and (1.52) 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 Francq and Zakoïan (2004)):

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

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

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

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

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

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

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

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

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

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

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

(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

$$\ell(x,y,\theta) = \frac{1}{2}\Big(\frac{x}{\sigma(y,\theta)^2} + \log(\sigma(y,\theta)^2)\Big), \tag{1.53}$$

$$\nabla_\theta \ell(x,y,\theta) = \frac{\nabla_\theta(\sigma(y,\theta)^2)}{2\sigma(y,\theta)^2}\Big(1 - \frac{x}{\sigma(y,\theta)^2}\Big), \tag{1.54}$$

$$\begin{aligned} \nabla^2_\theta \ell(x,y,\theta) = \Big[&- \frac{\nabla_\theta(\sigma(y,\theta)^2)\nabla_\theta(\sigma(y,\theta)^2)'}{2\sigma(y,\theta)^4} + \frac{\nabla^2_\theta(\sigma(y,\theta)^2)}{2\sigma(y,\theta)^2}\Big]\Big(1 - \frac{x}{\sigma(y,\theta)^2}\Big) \\ &+ \frac{\nabla_\theta(\sigma(y,\theta)^2)\nabla_\theta(\sigma(y,\theta)^2)'}{2\sigma(y,\theta)^4} \cdot \frac{x}{\sigma(y,\theta)^2}. \end{aligned} \tag{1.55}$$

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

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

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):

$$\sup_{|\theta - \tilde{\theta}|_1 < \iota} |\nabla_\theta \tilde{\ell}_{\tilde{\theta}}(\zeta, y, \theta)| \leq C_2(1 + |y|^{p/2}_{(\rho_2^j)_j,p/2})(1 + (1 + |y|^{p/2}_{(\rho_2^j)_j,p/2})\zeta^2). \tag{1.56}$$

By using

$$|y|^{p/2}_{(\rho_2^j)_j,p/2} \le \sum_{j=1}^{\infty} \rho_2^{j/2} \cdot \rho_2^{j/2} y_j^p \le (\sum_{j=1}^{\infty} \rho_2^j)^{1/2} (\sum_{j=1}^{\infty} \rho_2^j y_j^p)^{1/2}$$
$$= (1-\rho_2)^{-1/2} |y|^{p/2}_{(\rho_2^j)_j,p}, \tag{1.57}$$

we can obtain the more compact form

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

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

$$\sup_{|\theta-\tilde{\theta}|_1<\iota} |\nabla^l_\theta \tilde{\ell}_{\tilde{\theta}}(\zeta,y,\theta)| \le C_3(1+|y|^p_{(\rho_2^j)_j,p})(1+\zeta^2) =: M_p(y,\zeta), \qquad l=1,2,3. \tag{1.58}$$

From (1.50) we deduce that (component-wise) for $l = 0,1,2$ with some constant $C_4 > 0$, uniformly in $\theta,\theta' \in \bar{\Theta}$:

$$|\nabla^l_\theta(\sigma(y,\theta)^2) - \nabla^l_\theta(\sigma(y',\theta)^2)| \le C_4|y-y'|_{(\rho^j)_j,1}, \tag{1.59}$$

By using $|\frac{1}{\sigma(y,\theta)^2} - \frac{1}{\sigma(y,\theta)^2}| \le \frac{1}{\sigma^4_{min}}|\sigma(y,\theta)^2 - \sigma(y',\theta)^2|$ and the very rough bounds $\sigma(y,\theta)^2 \ge \sigma^2_{min}$, (1.59) and (1.50), we obtain (component-wise) with some constant $C_5 > 0$:

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

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

$$\sup_{\theta\in\bar{\Theta}} |\nabla^l_\theta \tilde{\ell}_{\tilde{\theta}}(\zeta,y,\theta) - \nabla_\theta \tilde{\ell}_{\tilde{\theta}}(\zeta,y',\theta)|$$
$$\le C_5(1+|y|_{(\rho^j)_j,1} + |y'|_{(\rho^j)_j,1})^{1+l} |y-y'|_{(\rho^j)_j,1}(1+\zeta^2) =: N_l(y,y',\zeta). \tag{1.60}$$

Using (1.58) and (1.60), we have for $l = 1,2$ and arbitrary small $p' > 0$ (use $\min\{1,x\} \le x^{p'}$):

$$\begin{aligned}
&\sup_{\theta\in\bar{\Theta}} |\nabla^l_\theta \tilde{\ell}_{\tilde{\theta}}(\zeta,y,\theta) - \nabla_\theta \tilde{\ell}_{\tilde{\theta}}(\zeta,y',\theta)| \\
&\le \min\{M_p(y,\zeta) + M_p(y',\zeta), N_l(y,y',\zeta)\} \\
&= \{M_p(y,\zeta) + M_p(y',\zeta)\} \min\{1, \frac{N_l(y,y',\zeta)}{M_p(y,\zeta)+M_p(y',\zeta)}\} \\
&\le \{M_p(y,\zeta) + M_p(y',\zeta)\} \Big(\frac{N_l(y,\zeta)}{M_p(y,\zeta)+M_p(y',\zeta)}\Big)^{p'} \\
&= \{M_p(y,\zeta) + M_p(y',\zeta)\}^{1-p'} N_l(y,\zeta)^{p'}.
\end{aligned}$$

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

$$\{M_p(y,\zeta) + M_p(y',\zeta)\}^{1-p'} \le C_3^{1-p'}(1+|y|^p_{(\rho_2^j)_j,p} + |y'|^p_{(\rho_2^j)_j,p})(1+\zeta^2)^{1-p'},$$
$$N_l(y,y',\zeta)^{p'} \le C_5^{p'}(1+|y|^p_{(\rho^{pj}),p} + |y'|^p_{(\rho^{pj}),p})|y-y'|^p_{(\rho^{pj})_j,p}(1+\zeta^2)^{p'}.$$

With (1.57), $\rho_3 := \max\{\rho_2, \rho^p\}$ and some constant $C_6 > 0$

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

By using (1.58) and the mean value theorem, we obtain for $l = 1, 2$:

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

giving the result.
(iii) Using the representations (1.54), (1.55) and the inequalities (1.59), (1.50) and $\sigma(y, \theta)^2 \geq \sigma_{min}^2$, this is an immediate consequence.