LSE Research Online

Wolfgang Polonik and Qiwei Yao

Testing for multivariate volatility functions using minimum volume sets and inverse regression

Article (Authors version) (Unrefereed)

Original citation:
Polonik, Wolfgang and Yao, Qiwei (2008) Testing for multivariate volatility functions using minimum volume sets and inverse regression. Journal of econometrics, 147 (1). pp. 151-162. DOI: 10.1016/j.jeconom.2008.09.003

© 2008 Elsevier

This manuscript (last edited October 13, 2007) is a more mathematical version of W. Polonik and Q. Yao (2006) *Testing for multivariate volatility functions using minimum volume sets and inverse regression*, published in the Journal of econometrics.

This version available at: http://eprints.lse.ac.uk/24132/
Available in LSE Research Online: May 2009

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.

This document is an author's version of the journal article, prior to undergoing peer-review. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

http://eprints.lse.ac.uk

Testing for Multivariate Volatility Functions

Wolfgang Polonik*
Department of Statistics
University of California
Davis, CA95616-8705, USA
Email: polonik@wald.ucdavis.edu

Qiwei Yao†
Department of Statistics
London School of Economics
London, WC2A 2AE, UK
Email: q.yao@lse.ac.uk

June 17, 2006

Abstract

We propose two new types of nonparametric tests for investigating multivariate regression functions. The tests are based on cumulative sums coupled with either minimum volume sets or inverse regression ideas; involving no multivariate nonparametric regression estimation. The methods proposed facilitate the investigation for different features such as if a multivariate regression function is (i) constant, (ii) of a bathtub shape, and (iii) of a given parametric form. The inference based on those tests may be further enhanced through associated diagnostic plots. Although the potential use of those ideas is much wider, we focus on the inference for multivariate volatility functions in this paper, i.e. we test for (i) heteroscedasticity, (ii) the so-called "smiling effect", and (iii) some parametric volatility models.The limit behavior of the proposed tests is investigated, and practical feasibility is shown via simulation studies. We further illustrate our methods with some real financial data.

Keywords: Brownian bridge, empirical process, ARCH models, heteroscedasticity, integral stochastic order, level set, smiling effect.

This manuscript (last edited October 13, 2007) is a more mathematical version of W. Polonik and Q. Yao (2006): Testing for multivariate volatility functions using minimum volume sets and inverse regression. Under Review at *Journal of Econometrics*.

*Partially supported by NSF grants # 0103606 and # 0406431.
†Partially supported by an EPSRC grant.

1 Introduction

We propose and study two types of nonparametric tests for investigating if a multivariate regression function is, for instance, constant, of a bathtub shape, or of a particular parametric form. The methodology has the potential to be useful in various contexts, including regression analysis and the analysis of time series. All the procedures proposed are associated with diagnostic plots.

In terms of methodology, our approach may be seen as a generalization of the classical goodness-of-fit tests for distribution functions (such as the Kolmogorov-Smirnov test) to those for regression functions. While most the classical goodness-of-fit tests for one-dimensional distribution functions are asymptotically distribution-free under the null hypotheses, this nice property is typically lost in multivariate cases. This explains the difficulties in directly applying, for example, the Kolmogorov-Smirnov tests for multivariate distribution functions (Polonik 1999). We circumvent this problem by using either minimum volume (MV) sets or an inverse regression idea. With MV sets, we effectively test a multivariate function in term of a single-indexed empirical process. The tests based on inverse regression rely on several one-dimensional empirical processes. Therefore the asymptotic distribution-free properties may be restored. The idea of using MV sets was initially proposed by Polonik (1999) for the goodness-of-fit tests for multivariate distribution functions.

We illustrate the new methods in the context of testing various features of volatility functions, which is particularly relevant to analysing financial time series. Let $\{Y_t\}$ be a strictly stationary and ergodic time series defined by

$$Y_t = \sigma_t \varepsilon_t, \tag{1.1}$$

where $\sigma_t \geq 0$ is $\mathcal{F}_{t-1}$-measurable, $\mathcal{F}_t$ denotes the σ-algebra generated by $\{Y_{t-k},\, k \geq 0\}$, and $\{\varepsilon_t\}$ is a sequence of independent and identically distributed random variables with mean 0, $E(\varepsilon_t^2) < \infty$, $E|\varepsilon_t| = 1$. Furthermore, we assume that ε_t is independent of $\mathcal{F}_{t-1}$. Now it is easy to see that $E(Y_t|\mathcal{F}_{t-1}) = 0$, and $E(|Y_t|\big|\mathcal{F}_{t-1}) = \sigma_t$. Hence $\sigma_t \equiv \sigma(Y_{t-1}, Y_{t-2}, \cdots)$ is a regression function of $|Y_t|$ on $Y_{t-1}, Y_{t-2}, \cdots$, and is called a volatility function. In fact (1.1) is a standard setting for modelling volatilities of financial returns (see, e.g. Morgan, 1996, p.92), although the conventional assumption is $E(\varepsilon_t^2) = 1$. This implies $E(Y_t^2|\mathcal{F}_{t-1}) = \sigma_t^2$. We adopt the parameterization implied by the condition $E|\varepsilon_t| = 1$ instead, in order to conduct the inference based on the absolute returns $|Y_t|$, which significantly relaxes the moment conditions required for the inference based on the squared returns Y_t^2.

We will consider three types of null hypotheses on σ_t, i.e. homoscedasticity (against heteroscedasticity), a so-called "smile effect" and a specified parametric form such as ARCH models. For the first two cases, our tests are asymptotically distribution-free under the null hypotheses. However the tests for parametric models are no longer distribution-free due to the presence of the estimators for the parameters in the test statistics, for which we adopt two bootstrap methods to approximate the P-values of the tests.

The rest of the paper is organized as follows. Section 2 deals with the tests for homoscedasticity. Section 3 extends the ideas for testing parametric forms of volatility functions. The tests for a bathtub shaped volatility (i.e. "smiling factor") is discussed in section 4. Numerical illustration with both simulated and real data is presented in section 5. Theoretical properties of the test statistics are derived in section 6.

2 Tests for homoscedasticity

In this section we deal with the tests for the homoscedasticity hypothesis

$$H_0 \;:\; \sigma(\cdot) \equiv \nu_y, \qquad \nu_y > 0 \;\text{ is a constant.} \tag{2.1}$$

Conventional practice is to test the null hypothesis (2.1) against a specified parametric form such as ARCH models; see, e.g., section 4.2 of Fan and Yao (2003) and references therein. More recently, a nonparametric approach has been adopted for testing for conditional heteroscedasticity for univariate volatility functions; see Chen and An (1997) and Laïb (2003). In terms of methodology, the available tests may be classified into two categories: tests solely based on analyzing residuals (Engle 1982, Lee 1991, McLeod and Li 1993, and Horváth et al. 2001), and tests based on residual-regression (Chen and An, 1997, Stute, 1997, Koul and Stute, 1999, and Laïb, 2003). The latter is based on the fact that under the null-hypothesis (2.1), it holds that

$$E\{(\,|Y_t| - E|Y_t|\,)\, I(\,\mathbf{X}_t \le \mathbf{x}\,)\,\} = 0 \qquad\qquad \text{for all } \mathbf{x} \in \mathcal{R}^p, \tag{2.2}$$

where $\mathbf{X}_t = (Y_{t-1}, \cdots, Y_{t-p})'$ ($p \geq 1$), and $\mathbf{X}_t \leq \mathbf{x}$ denotes that each component of $\mathbf{X}_t$ is not greater than the corresponding component of $\mathbf{x}$. All the work mentioned above in the second category deals with univariate regressor only. The methods proposed in this paper may be viewed as an attempt to extend these methods from the second category to multivariate cases.

Our new tests are based on the observation that if hypothesis (2.1) holds then $F \equiv G$, where $F(\cdot)$ denotes the distribution function of $\mathbf{X}_t$, and G is a distribution function defined as

$$G(\mathbf{x}) = \nu_y^{-1} E\{|Y_t| I(\mathbf{X}_t \leq \mathbf{x})\}, \tag{2.3}$$

where $\nu_y = E|Y_t|$. It is easy to see that $G(\cdot)$ is a well-defined probability measure. Thus hypothesis (2.1) may be viewed as a hypothesis on two probability distributions. However we do not have observations directly from distribution G. Note that $F(\mathbf{x})$ and $G(\mathbf{x})$ may be estimated by, respectively,

$$F_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} I(\mathbf{X}_t \leq \mathbf{x}), \qquad G_n(\mathbf{x}) = \frac{1}{n\,\widehat{\nu}_y} \sum_{i=1}^{n} |Y_t| I(\mathbf{X}_t \leq \mathbf{x}),$$

where $\widehat{\nu}_y = n^{-1} \sum_{1 \leq t \leq n} |Y_t|$. Hence, we may test hypothesis (2.1) using the statistic

$$\sup_{\mathbf{x}} \widehat{\nu}_y \left| G_n(\mathbf{x}) - F_n(\mathbf{x}) \right| \;=\; \sup_{\mathbf{x}} \left| \frac{1}{n} \sum_{i=1}^{n} \big\{ |Y_t| - \widehat{\nu}_y \big\} I(\mathbf{X}_t \leq \mathbf{x}) \right|, \tag{2.4}$$

When $p = 1$, tests of this type have been extensively explored by, among others, Chen and An (1997), Stute (1997), Stute et al. (1998), Koul and Stute (1999) and Laïb (2003). For $p > 1$ the null-distribution of the test statistic (2.4) depends on the underlying distribution. Note that the lack of the (asymptotic) distribution-free property may cause non-trivial difficulties in determining the critical values of the tests since the null hypothesis (2.1) is not simple.

We construct the tests using minimum volume (MV) sets or an inverse regression idea. A remarkable gain for these new approaches is that the null-distributions of our test statistics are asymptotically distribution-free even when $p > 1$. Note that for testing the null hypothesis (2.1), one may simply let $p = 1$ (i.e. $\mathbf{X}_t = Y_{t-1}$) in (2.4). We argue that the tests with $p > 1$ are significantly more powerful than those with $p = 1$ when σ_t depends on several lagged values of Y_t. Numerical results in section 5 provide convincing evidence to support this argument. Note for GARCH(1,1) processes,

σ_t effectively depends on Y_{t-k} for all $1 \leq k < \infty$.

2.1 Tests based on MV sets

We first introduce the concept of minimum volume (MV) sets. A MV set under a distribution H on $\mathcal{R}^p$ indexed by $\alpha \in [0,1]$ is defined as

$$M_H(\alpha) := \arg \min_{A \subset \mathcal{R}^p} \{\text{Leb}(A) : \; H(A) \geq \alpha\}, \tag{2.5}$$

where Leb(A) denotes the Lebesgue measure of A. Obviously $M_H(\alpha)$ is a set of minimum Lebesgue measure among the sets of H-measure not smaller than α. When H possesses a probability density function h which has no flat parts (see (2.9) below), MV sets exist and are essentially unique (up to H-nullsets). In fact, in this case $M_H(\alpha) = \{\, \mathbf{x} : h(\mathbf{x}) \geq \lambda_\alpha \,\}$ for an appropriate level $\lambda_\alpha \geq 0$. In the following we assume that both F and G posses pdf's f and g, respectively.

One of the reasons to use MV sets for constructing our tests is that they are capable of discriminating different distributions. Suppose that both f and g do not have flat parts in the sense of (2.9). Polonik (1999) showed that $F = G$ if and only if $(F - G)\{M_F(\alpha)\} = 0$ and $(F - G)\{M_G(\alpha)\} = 0$ for all $\alpha \in [0,1]$.

Estimators for MV sets under F may be obtained by replacing F in (2.5) by the estimator F_n. In order to obtain a reasonable estimator, we need to restrict candidate sets A in (2.5) to be in a class of selected sets (to avoid "oversmoothing"). Let $\mathcal{C}$ be a set consisting of appropriate subsets of $\mathcal{R}^p$. An estimator for $M_F(\alpha)$ may be defined as

$$\widehat{M}_{\mathcal{C},F_n}(\alpha) := \arg \min_{A \in \mathcal{C}} \{\text{Leb}(A) : \; F_n(A) \geq \alpha\}, \tag{2.6}$$

which is called an empirical MV set in $\mathcal{C}$. We should choose $\mathcal{C}$ such that it contains all the MV sets under both F and G. Then hypothesis (2.1) holds if and only if $E\{|Y_t| - \nu_y \big| \mathbf{X}_t \in A\} = 0$ for all $A \in \mathcal{C}$; see also (2.2). The latter is equivalent to

$$(G - F)(A) = 0 \quad \text{for all } A \in \mathcal{C}. \tag{2.7}$$

Let $\widehat{M}_{\mathcal{C},G_n}(\alpha)$ be the empirical MV sets in $\mathcal{C}$ under G_n; see (2.6). Put

$$\widehat{\Delta}_n(C) = (F_n - G_n)(C). \tag{2.8}$$

Relation (2.7) suggests that we may define a test statistic

$$T_1 = \sup_{\alpha\in(0,1]} \big[\, |\widehat{\Delta}_n(\widehat{M}_{\mathcal{C},F_n}(\alpha))| + |\widehat{\Delta}_n(\widehat{M}_{\mathcal{C},G_n}(\alpha))| \,\big],$$

and reject hypothesis (2.1) for large values of T_1.

We will assume that $\mathcal{C}$ is chosen such that $M_F(\alpha) \in \mathcal{C}$ and $M_G(\alpha) \in \mathcal{C}$ for all α. This assumption may be interpreted as a correctly chosen model. In addition, for technical reasons we always assume that $\emptyset \in \mathcal{C}$. The general idea is that the class $\mathcal{C}$ should be rich enough to distinguish F from G when H_0 does not hold. On the other hand, $\mathcal{C}$ may not be too rich, so that all the MV sets may be consistently estimated by their empirical counterparts. Complexity of the class $\mathcal{C}$ may be measured by its metric entropy which leads to the covering integral $I_B(\mathcal{C})$ defined in (6.1) below. Most results below require a finite covering integral which restricts $\mathcal{C}$ from being too large; see further discussion in section 6. Another condition which restricts $\mathcal{C}$ from being too large is **(C2)** which is also introduced in section 6. In our applications, in order to make the computation attainable, we typically let $\mathcal{C}$ consist of all ellipsoids. This effectively imposes a shape constraint on the underlying distributions. We will use the following assumptions:

(A1) F and G have bounded and continuous Lebesgue densities f and g respectively.

(A2) The densities f and g have no flat parts, i.e.

$$\sup_{\lambda>0} F\,\{\mathbf{x}\in\mathcal{R}^p :\, |f(\mathbf{x}) - \lambda\,| \le \eta\,\} \to 0 \qquad \text{as } \eta \to 0. \tag{2.9}$$

and the same holds with (F, f) replaced by (G, g).

Theorem 2.1 *Let the class $\mathcal{C}$ be such that $\mathbf{(C2)}_{F_n}$, $\mathbf{(C2)}_{G_n}$ and (6.1) hold. Assume further $M_F(\alpha) \in$*

$\mathcal{C}$ and $M_G(\alpha) \in \mathcal{C}$ for all α, and that **(A1)** *and* **(A2)** *hold. Then if $F = G$ we have as $n \to \infty$*

$$\sqrt{\frac{n\,\widehat{\nu}_y^2}{\widehat{\sigma}_y^2}}\,T_1 \;\xrightarrow{\mathcal{D}}\; 2 \sup_{\alpha \in [0,1]} |B(\alpha)|$$

where $\widehat{\sigma}_y^2 = n^{-1} \sum_{1 \le t \le n} Y_t^2 - \widehat{\nu}_y^2$, and B denotes a standard Brownian bridge process. On the other hand, if $F \neq G$, then $P(T_1 > c) \to 1$ as $n \to \infty$ for some constant $c > 0$.

Combining the theorem above and (9.39) of Billingsley (1999), we have the approximation:

$$P\Big(\sqrt{\frac{n\,\widehat{\nu}_y^2}{\widehat{\sigma}_y^2}}T_1 > z\Big) \;\approx\; 2\sum_{k=1}^{\infty}(-1)^{k-1} \exp\Big\{-\frac{k^2 z^2}{2}\Big\}. \tag{2.10}$$

Diagnostic plots. The test based on statistic T_1 naturally leads to two diagnostic plots $\alpha \to \Big(G_n\{\widehat{M}_{\mathcal{C},G_n}(\alpha)\},\; F_n\{\widehat{M}_{\mathcal{C},G_n}(\alpha)\}\Big)$ and $\alpha \to \Big(G_n\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\},\; F_n\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\}\Big)$ which are called CC-plots (Polonik 1999). They may be viewed as a generalization of the standard QQ-plots for univariate distributions to multivariate distributions. Under null hypothesis (2.1), both plots should be approximately a 45^o straight line.

2.2 Tests based on inverse regression

Although the test T_1 is based on an empirical process with single index, we still need to compute the MV sets in $\mathcal{R}^p$. In this subsection, we swap the roles of Y_t and $\mathbf{X}_t$; leading to a test based on one-dimensional sets only. The key idea is to use an inverse regression equation, i.e. under the null hypothesis (2.1),

$$E\big[(\mathbf{Z}_t - \boldsymbol{\nu}_y)\, I\{|Y_t| < y\}\big] = 0 \qquad\qquad \text{for all } y \ge 0, \tag{2.11}$$

where $\mathbf{Z}_t = (\,|Y_{t-1}|, \dots, |Y_{t-p}|\,)'$, and $\boldsymbol{\nu}_y$ is a $p \times 1$ vector with all elements equal to $\nu_y = E|Y_t|$. Hence we may define a test statistic

$$T_2 = \frac{1}{n} \max_{1 \le j \le p} \sup_x \Big| \sum_{t=1}^{n} \big(\,|Y_{t-j}| - \widehat{\nu}_{y,j}\,\big)\, I(|Y_t| \le x) \Big|,$$

and reject hypothesis (2.1) for large values of T_2, where $\widehat{\nu}_{y,j} = n^{-1} \sum_{1 \le t \le n} |Y_{t-j}|$.

Theorem 2.2 *Let p be fixed. Suppose $\sqrt{n}\,(\widehat{\nu}_{y,j} - \nu_y) = O_P(1)$, $j = 1, \ldots, p$. If (2.11) holds, then we have as $n \to \infty$ that*

$$\sqrt{\frac{n\,\widehat{\nu}_y^2}{\widehat{\sigma}_y^2}}\, T_2 \xrightarrow{\mathcal{D}} \max_{1\le j\le p}\ \sup_{0\le\alpha\le 1} |B_j(\alpha)|\,,$$

where $B_1, \cdots, B_p$ denote p independent standard Brownian bridge processes. On the other hand, if (2.11) does not hold, then $P(T_2 > c) \to 1$ for some constant $c > 0$.

Due to the independence of the limiting Brownian bridge processes, it holds that

$$P\{\max_{1\le j\le p}\ \sup_{0\le\alpha\le 1} |B_j(\alpha)| \ge z\} = 1 - \big[\,1 + 2\sum_{k=1}^{\infty}(-1)^k \exp\big\{-2\,k^2 z^2\big\}\,\big]^p,$$

see, e.g., (9.39) of Billingsley (1999).

Note that (2.11) in general does not imply $F = G$. Therefore, in principle the test T_2 is powerless to reject hypothesis (2.1) for the processes which fulfill (2.11) but not (2.1). We argue that such a situation is rare in practice. The advantage of using one-dimensional sets in T_2 brings in considerable convenience in practice, in spite of the fact that it is not an omnibus test for the conditional heteroscedasticity.

Diagnostic plots. It is easy to see that under the null hypothesis (3.1), the plots

$$y \;\to\; \Big(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} |Y_{t-j}| I(|Y_t| \le y),\ \frac{\widehat{\nu}_{y,j}}{\sqrt{n}}\sum_{t=1}^{n} I(|Y_t| \le y)\Big), \quad j = 1, \cdots, p$$

should all approximately be 45^o lines through the origin.

3 Tests for parametric heteroscedasticity

The methods presented in above may be formally extended to testing for the parametric heteroscedasticity hypothesis

$$H_0:\ \sigma_t = \sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0) \quad \text{for some } \boldsymbol{\theta}_0 \in \boldsymbol{\Theta}, \tag{3.1}$$

where $\mathbf{X}_t = (Y_{t-1}, \cdots, Y_{t-p})'$, Y_t is defined by (1.1), the form of function $\sigma_0 > 0$ is known, and $\boldsymbol{\Theta} \subset \mathcal{R}^q$, and $p, q \ge 1$ are integers. For example, for $q = p+1$ and $\sigma_0(\mathbf{x}, \boldsymbol{\theta})^2 = \theta_1 + \theta_2 x_1^2 + \cdots + \theta_{p+1} x_p^2$, we test the validation of ARCH(p) model.

3.1 Tests based on MV sets

Put $G_{\boldsymbol{\theta}}(A) = \frac{1}{\nu_\theta} \mathrm{E}\Big\{ \frac{|Y_t|}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta})} I(\mathbf{X}_t \in A) \Big\}$ where $\nu_\theta = \mathrm{E}\{|Y_t|/\sigma_0(\mathbf{X}_t, \boldsymbol{\theta})\}$ is a normalizing constant, and $E_{\boldsymbol{\theta}}$ denotes expectation taken with $\sigma(\cdot) = \sigma_0(\cdot, \boldsymbol{\theta})$ in (1.1). It is easy to see that $G_{\boldsymbol{\theta}}(\cdot)$ is a well-defined probability measure on $\mathcal{R}^p$. Furthermore, the null hypothesis (3.1) holds if and only if $G_{\boldsymbol{\theta}_0} \equiv F$. The latter is equivalent to $(G_{\boldsymbol{\theta}_0} - F)(A) = 0$ for all $A \in \mathcal{C}$, if, for example, $\mathcal{C}$ contains all the MV sets under F and $G_{\boldsymbol{\theta}_0}$. Hence we may construct a test statistic based on a sample version of the above expression. To this end, let $\widehat{\boldsymbol{\theta}}$ be an estimator for $\boldsymbol{\theta}_0$, and define

$$e_{t,\widehat{\boldsymbol{\theta}}} = |\, Y_t \,| / \sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}}), \tag{3.2}$$

$$G_{n,\widehat{\boldsymbol{\theta}}}(A) = \frac{1}{n\, \widehat{\nu}_{\widehat{\boldsymbol{\theta}}}} \sum_{t=1}^{n} e_{t,\widehat{\boldsymbol{\theta}}}\, I(\mathbf{X}_t \in A), \tag{3.3}$$

where $\widehat{\nu}_{\widehat{\boldsymbol{\theta}}} = \frac{1}{n}\sum_{t=1}^{n} e_{t,\widehat{\boldsymbol{\theta}}}$. Let $\widehat{M}_{\mathcal{C},F_n}(\alpha)$ and $\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)$ be the empirical MV set under, respectively, F_n and $G_{n,\widehat{\boldsymbol{\theta}}}$; see (2.6). For $\alpha \in [0,1]$, put $\widehat{\Delta}_{n,\widehat{\boldsymbol{\theta}}}(C) = (G_{n,\widehat{\boldsymbol{\theta}}} - F_n)(C)$, and define the test statistic as

$$T_3 = \sup_{\alpha \in (0,1]} \big\{ \big|\widehat{\Delta}_{n,\widehat{\boldsymbol{\theta}}}(\widehat{M}_{\mathcal{C},F_n}(\alpha))\big| + \big|\widehat{\Delta}_{n,\widehat{\boldsymbol{\theta}}}(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))\big| \big\}.$$

Due to the presence of the estimator $\widehat{\boldsymbol{\theta}}$, the asymptotic null-distribution of the statistic T_3 depends on the underlying processes in a rather implicit manner. This becomes clear from the following result about the processes $\widehat{\Delta}_{n,\widehat{\boldsymbol{\theta}}}(\widehat{M}_{\mathcal{C},F_n}(\alpha))$ and $\widehat{\Delta}_{n,\widehat{\boldsymbol{\theta}}}(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))$, respectively. To formulate this result we need the following assumptions on the volatility function.

(V1) $\sigma_0(\mathbf{x}, \cdot)$ is differentiable at $\boldsymbol{\theta}_0$ for every $\mathbf{x}$ with $E_{\boldsymbol{\theta}_0}\Big(\frac{\partial}{\partial \theta_k}\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)\Big)^2 < \infty$, and $|\sigma_0(\mathbf{x}, \boldsymbol{\theta}_0| \geq \epsilon$ for some $\epsilon > 0$.

(V2) With $B_{c/\sqrt{n}}(\boldsymbol{\theta}_0) := \big\{ \boldsymbol{\tau} \in \mathcal{R}^q : |\boldsymbol{\tau} - \boldsymbol{\theta}_0| \leq \frac{c}{\sqrt{n}} \big\}$ we have:

(a) there exists functions $b_1(\cdot), \ldots, b_p(\cdot)$ with $\mathrm{E}b_k^2(\mathbf{X}_t) < \infty$, $k = 1, \ldots, p$, such that for each $c > 0$

as $n \to \infty$

$$\sup_{1\le t\le n}\ \sup_{\boldsymbol{\tau}\in B_{c/\sqrt{n}}(\boldsymbol{\theta}_0)}\left|\frac{\frac{\partial}{\partial\theta_k}\sigma_0(\mathbf{X}_t,\boldsymbol{\tau})-\frac{\partial}{\partial\theta_k}\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{b_k(\mathbf{X}_t)}\right| = o_P(1) \qquad \text{for all } k=1,\dots,p.$$

(b) $$\sup_{1\le t\le n}\ \sup_{\boldsymbol{\tau}\in B_{c/\sqrt{n}}(\boldsymbol{\theta}_0)}\left|\frac{\sigma_0(\mathbf{X}_t,\boldsymbol{\tau})}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}-1\right| = o_P(1).$$

Further let

$$W_n(C) := \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\,|\varepsilon_t|-1\,)\,[\,I(\mathbf{X}_t\in C)-F(C)\,], \quad C\in\mathcal{C}, \tag{3.4}$$

and assume that

(C1) the class $\mathcal{C}$ is such that $\{\,W_n(C);\ C\in\mathcal{C}\,\}$ is asymptotically equicontinuous with respect to $d_F(C,D)=F(C\,\Delta\,D)$.

We need another condition related to the complexity of $\mathcal{C}$. For a (in general random) distribution H we define condition $\mathbf{(C2)}_H$ as

$$\mathbf{(C2)}_H \qquad \sup_{\alpha\in[0,1]}|H(M_H(\alpha))-\alpha| = o_P(1). \tag{3.5}$$

This condition for instance holds (for continuous underlying distributions F and G) with $H=F_n$ or $H=G_n$ for $\mathcal{C}$ the class of all balls or ellipsoids in $\mathcal{R}^d$.

Theorem 3.1 *Assume that* **(A1)** *and* **(A2)** *hold, and that* $\sigma_0(\mathbf{x},\cdot)$ *satisfies* **(V1)** *and* **(V2)**. *Suppose that* $\sqrt{n}\,(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)=O_P(1)$. *Let* $\mathcal{C}$ *be such that* $M_F(\alpha)\in\mathcal{C}$ *for all* $\alpha\in[0,1]$, *that* $I_B(\mathcal{C})<\infty$ *(cf. (6.1)) and that* **(C1)** *and* $\mathbf{(C2)}_{G_n,\widehat{\boldsymbol{\theta}}}$ *hold. Then, under the null hypothesis (3.1) we have the following:*

$$\sqrt{n}\,\widehat{\Delta}_{n,\widehat{\boldsymbol{\theta}}}(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) = W_n(M_F(\alpha)) \ + \ (\sqrt{n}\,(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0))'\,\mathbf{b}_{\boldsymbol{\theta}_0}(M_F(\alpha)) \ + \ o_P(1) \quad \text{as } n\to\infty, \tag{3.6}$$

where the $o_P(1)$*-term is uniform in* α, *and* $\mathbf{b}_{\boldsymbol{\theta}_0}(M_F(\alpha)) = E_{\boldsymbol{\theta}_0}\big[\frac{\dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}\,(\,I(\mathbf{X}_t\in M_F(\alpha))-\alpha\,)\big]$. *The same expansion (3.6) holds for* $\widehat{\Delta}_{n,\widehat{\boldsymbol{\theta}}}(\widehat{M}_{\mathcal{C},F_n}(\alpha))$ *if we assume* $\mathbf{(C2)}_{F_n}$ *(rather than* $\mathbf{(C2)}_{G_n,\widehat{\boldsymbol{\theta}}}$*). Moreover,* $\{W_n(M_F(\alpha)),\alpha\in[0,1]\}$, *considered as a process in* $\ell^\infty([0,1])$, *converges in distribution* *to* $\{\mathrm{Var}(|\varepsilon_t|)\, B(\alpha),\ \alpha \in [0,1]\}$, *where* $\{\, B(\alpha),\ \alpha \in [0,1]\,\}$ *denotes a standard Brownian bridge.*

This result shows that the asymptotic behavior of $\widehat{\Delta}_{n,\widehat{\boldsymbol\theta}}$ cannot be used to compute P-values of the corresponding test T_3 directly, because in general, under H_0 the asymptotic distribution of T_3 will depend on the underlying process as well as the particular estimator $\widehat{\boldsymbol\theta}$. This is due to the second term on the r.h.s. of (3.6), which would vanish if we would not need to estimate $\boldsymbol\theta$ (i.e. $\widehat{\boldsymbol\theta} = \boldsymbol\theta_0$).

Assumption **(C1)** holds, for instance, under appropriate mixing conditions together with the assumption of $\mathcal{C}$ having a finite bracketing integral (cf. Doukhan et al. 1995). However, there are certainly various types of assumptions that can be used to assure **(C1)** to hold, and this is the reason for making **(C1)** an assumption rather than to formulate one set of sufficient conditions.

3.2 Tests based on inverse regression

Based on the same idea as for T_2, we may test the null hypothesis (3.1) using the statistic

$$T_4 = \max_{1\le j\le p} \sup_x \frac{1}{n}\Big| \sum_{t=1}^{n} (|Y_{t-j}| - \widehat{\nu}_{y,j}) I(e_{t,\widehat{\boldsymbol\theta}_0} \le x)\Big|.$$

where $e_{t,\widehat{\boldsymbol\theta}}$ is defined in (3.2), and $\widehat{\nu}_{yj} = n^{-1}\sum_{1\le i\le n} |Y_{i-j}|$. Similar to Theorem 3.1 we have the following expansion for the process underlying T_4. Recall the definition of $B_{c/\sqrt{n}}(\boldsymbol\theta_0)$ given above.

Theorem 3.2 *Suppose that* $\sqrt{n}\,(\widehat{\boldsymbol\theta} - \boldsymbol\theta_0) = O_P(1)$, *and that* $\sqrt{n}\,(\widehat{\boldsymbol\nu}_y - \boldsymbol\nu_y) = O_P(1)$, *and assume that* $\sigma(\cdot)$ *satisfies* **(V1)** *and* **(V2)**. *Then, under the null hypothesis (3.1) we have as* $n \to \infty$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big\{\mathbf{Z}_t - \widehat{\nu}_y\big\} I(\, e_{t,\widehat{\boldsymbol\theta}} \le y\,) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big\{\mathbf{Z}_t - \nu_y\big\}\, (\, I(\, \varepsilon_t \le y) - F_\varepsilon(y)\,)$$
$$+\ y\, f(y)\, (\, \sqrt{n}\, (\, \widehat{\boldsymbol\theta} - \boldsymbol\theta_0\,)\,)'\, \mathbf{a}_0 \ +\ o_P(1), \qquad (3.7)$$

where $\mathbf{a}_0 = (a_{0,j},\ j = 1, \dots, p)'$, *and* $a_{0,j} = \mathrm{E}_{\boldsymbol\theta_0}\Big(\big\{|Y_{t-j}| - \nu_y\big\} \frac{\dot\sigma_0(\mathbf{X}_t, \boldsymbol\theta_0)}{\sigma_0(\mathbf{X}_t, \boldsymbol\theta_0)} \Big)$. *Moreover,*

$$[\,\mathrm{Cov}(Z_t)\,]^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \big\{\mathbf{Z}_t - \nu_y\big\}\, (\, I(\, \varepsilon_t \le y) - F_\varepsilon(y)\,) \to_{\mathcal{D}} (B_{1,F_\varepsilon}(\alpha), \dots, B_{p,F_\varepsilon}(\alpha))',$$

where $\{\, B_{j,F_\varepsilon}(\alpha),\ \alpha \in [0,1]\,\}$, $j = 1, \dots, p$ *denote* p *independent Brownian bridges with*

$\text{Cov}(B_{j,F_\varepsilon}(\alpha),\, B_{j,F_\varepsilon}(\beta)) = F_\varepsilon(\alpha \wedge \beta) - F_\varepsilon(\alpha)\, F_\varepsilon(\beta).$

Again, the term involving $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$ on the r.h.s. of (3.7) makes the asymptotic null-distribution of T_4 depend on the underlying process. Below we propose to approximate the null-distributions of both T_3 and T_4 by some bootstrap methods.

One alternative would be to apply a martingale transform to the (estimated) empirical processes to make their sampling distributions independent of underlying processes; see Khmaladze (1981, 1988). This method has been successfully applied in dealing with processes indexed by a real parameter; see, for example, Stute (1997), Stute *et al.* (1998) and Koul and Stute (1999). Although the asymptotic expansion presented in Theorems 3.1 and 3.2 indicate that an application of this method is feasible, there are several issues to be explored before this method can actually be considered seriously in the present context. One is the practical implementation of the method to multivariate cases, and another is to conduct a detailed theoretical power study for our tests, in order to make sure that the transformation method does not lead to a significant loss of power. These studies go beyond the scope of this manuscript and will be conducted elsewhere.

3.3 Bootstrap tests

For bootstrap tests for a composite null hypothesis, ideally the bootstrap sample should be drawn from the 'representative' distribution of the null hypothesis, which determines the significance levels for the tests. This may be achieved easily if the null-distribution of a test statistic is distribution-free, as then the bootstrap sample may be draw from any distribution under the null hypothesis. In general, we typically replace the 'representative' distribution by the distribution under the null hypothesis which is the 'closest' to the observations (Hinkley 1988, Hall and Wilson 1991). As we have pointed out above, the asymptotic distributions of both T_3 and T_4 are not distribution-free. We outline below the two bootstrap methods to estimate the P-values of the tests.

Parametric bootstrap test. Under some circumstances we may assume that the distribution of innovations ε_t in model (1.1) is known, say, F_1. The bootstrap sample may be drawn from the equation

$$Y_t^* = \sigma_0(\mathbf{X}_t^*, \widehat{\boldsymbol{\theta}})\varepsilon_t^*, \tag{3.8}$$

where $\mathbf{X}_t^* = (Y_{t-1}^*, \cdots, Y_{t-p}^*)'$, $\widehat{\boldsymbol{\theta}}$ is an estimator for $\boldsymbol{\theta}$, and $\{\varepsilon_t^*\} \sim_{i.i.d.} F_1$.

Nonparametric bootstrap test. If the distribution of ε_t in model (1.1) is unknown, we may adopt a nonparametric bootstrap method as follows: define the residuals $\widehat{\varepsilon}_t = Y_t/\sigma(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})$, and draw bootstrap sample from (3.8) but now $\{\varepsilon_t^*\}$ are independent drawn from the residuals $\{\widehat{\varepsilon}_t\}$.

With a bootstrap sample, T_3^* (or T_4^*) is computed in the same manner as T_3 (or T_4) with $\{Y_t\}$ replaced by $\{Y_t^*\}$. The bootstrap estimate for the P-value is the relative frequency of the occurrence of the event $T_3 > T_3^*$ (or $T_4 > T_4^*$) in a repeated bootstrap sampling with B times, where $B > 0$ is a large integer.

The nonparametric bootstrap test outlined above is more general than the parametric one. It may still apply when the innovation distribution F_1 is known. However some power-loss may be expected then, since the residuals $\widehat{\varepsilon}_t$, and therefore also bootstrap innovations ε_t^*, will not behave like a random sample from F_1 *if* the null hypothesis (3.1) does not holds; see the numerical examples in table 5.1 in section 5.

The following result justifies the parametric bootstrap procedure. For a given $\boldsymbol{\theta}$ we denote by $Y_{t,\boldsymbol{\theta}}$ a stationary solution to

$$Y_t = \sigma_0(\mathbf{X}_t, \boldsymbol{\theta})\, \varepsilon_t.$$

Theorem 3.3 *Suppose that the assumptions of Theorem 3.1 with $F = G_{\boldsymbol{\theta}_0}$ and $G = G_{\boldsymbol{\theta}}$ hold uniformly in $\boldsymbol{\theta} \in \mathcal{U}_{c/\sqrt{n}}(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \le c\, n^{-1/2}\}$ for any choice of $c > 0$. In addition we assume that uniformly in $\boldsymbol{\theta} \in \mathcal{U}_{c/\sqrt{n}}(\boldsymbol{\theta}_0)$*

(i) there exist realizations $Y_{t,\boldsymbol{\theta}}$ and $Y_{t,\boldsymbol{\theta}_0}$ on the same probability space such that

$$\sup_{1 \le t \le n} \big|\, |Y_{t,\boldsymbol{\theta}}| - |Y_{t,\boldsymbol{\theta}_0}|\, \big| = o_P(1) \qquad \text{as } n \to \infty; \tag{3.9}$$

(ii) the estimator $\widehat{\theta}$ is such that

$$\sqrt{n}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} h(\varepsilon_t)\, \boldsymbol{\psi}_{\boldsymbol{\theta}}(\mathbf{X}_{t,\boldsymbol{\theta}}) + o_P(1) \quad \text{as } n \to \infty, \tag{3.10}$$

with $\mathrm{E}h(\varepsilon_t) = 0$, $\sigma_h^2 = \mathrm{E}h^2(\varepsilon_t) < \infty$, and $\boldsymbol{\psi}_{\boldsymbol{\theta}}$ is a Lipschitz continuous q-vector of measurable functions satisfying $\mathrm{E}\big[\boldsymbol{\psi}_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_0})\big]^2 < \infty$ and $\big\|\boldsymbol{\psi}_{\boldsymbol{\theta}}(\mathbf{x}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{x})\big\| \le K(\mathbf{x})\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|$ for a *measurable function K with* $\sup_{\boldsymbol\theta\in\mathcal{U}_{c/\sqrt{n}}(\boldsymbol\theta_0)} \mathrm{E}K^2(\mathbf{X}_{t,\boldsymbol\theta}) < \infty$;

(iii) the pdf's $g_{\boldsymbol\theta}(\cdot)$ *are Lipschitz continuous uniformly in* $\boldsymbol\theta \in \mathcal{U}_{c/\sqrt{n}}(\boldsymbol\theta_0)$, *and* $g_{\boldsymbol\theta}(\mathbf{x})$ *is continuous in* $\boldsymbol\theta_0$ *for all* $\mathbf{x}$*;*

(iv) $\dot\sigma_0(\mathbf{x},\boldsymbol\theta)$ *is continuous in both* $\mathbf{x}$ *and* $\boldsymbol\theta \in \mathcal{U}_{c/\sqrt{n}}(\boldsymbol\theta_0)$, *and there exist a q-vector of measurable functions* $\mathbf{a}$ *with* $|\dot\sigma_0(\mathbf{x},\boldsymbol\theta)| \le \mathbf{a}(x)$ *and* $\sup_{\boldsymbol\theta\in\mathcal{U}_{c/\sqrt{n}}(\boldsymbol\theta)} \mathrm{E}\mathbf{a}(X_{t,\boldsymbol\theta}) < \infty$.

Then we have as $n \to \infty$ *that*

$$\sup_{x\in\mathcal{R}} |P(T_3^* \le x|X_1,\ldots,X_n) - P(T_3 \le x)| \to 0 \ \textit{ in probability.}$$

Remark: An estimator $\widehat{\boldsymbol\theta}$ that allows for the required expansion (3.10) with $h(\varepsilon_t) = \varepsilon_t^2$ can be found in Hall and Yao (2003), p. 304, for the quasi-maximum-likelihood estimator of a GARCH-model (with of course the ARCH-model being a special case). One of the crucial assumptions is (i). It holds, for instance, for power-ARCH(p)-models with index 1. This can be seen by observing that if $|Y_{t,\boldsymbol\theta}|$ is a solution to $|Y_t| = \sigma_0(\mathbf{X}_t,\boldsymbol\theta)|\varepsilon_t|$ with $\sigma_0(\mathbf{X}_t,\boldsymbol\theta) = a + \sum_{j=1}^p b_j|Y_{t-j}|$ with $\boldsymbol\theta = (a, b_1,\ldots,b_p)$ then, if $\sum_{j=1}^p b_j < 1$ with $b_j \ge 0$ we have

$$\begin{aligned}|Y_{t,\boldsymbol\theta}| &= a|\varepsilon_t| + a\sum_{\ell=1}^{\infty}\sum_{1\le j_1,\ldots,j_\ell\le p} b_{j_1}\cdots b_{j_\ell}|\varepsilon_t|\,|\varepsilon_{t-j_1}|\cdots|\varepsilon_{t-j_1-\cdots j_\ell}|\\ &:= a|\varepsilon_t| + a\Phi(b_1,\ldots,b_p,|\varepsilon_t|,|\varepsilon_{t-1}|,|\varepsilon_{t-2}|,\ldots).\end{aligned}$$

Using the same sequence of random variables ε_t and by changing the parameter $\boldsymbol\theta$ to $\boldsymbol\theta_n$ we can see that the corresponding $Y_{\boldsymbol\theta_n}$ satisfies

$$\begin{aligned}|Y_{t,\boldsymbol\theta}| - |Y_{t,\boldsymbol\theta_n}| &= (a-a_n)\,|\varepsilon_t| \ + \ (a-a_n)\,\Phi(b_1,\ldots,b_p,|\varepsilon_t|,|\varepsilon_{t-1}|,|\varepsilon_{t-2}|,\ldots)\\ &\quad + \ a_n\left[\Phi(b_1,\ldots,b_p,|\varepsilon_t|,|\varepsilon_{t-1}|,|\varepsilon_{t-2}|,\ldots) - \Phi(b_{1,n},\ldots,b_{p,n},|\varepsilon_t|,|\varepsilon_{t-1}|,|\varepsilon_{t-2}|,\ldots)\right].\end{aligned}$$

Clearly, the first two summands are $O_P(n^{-1/2})$ if $\widehat{\boldsymbol\theta} - \boldsymbol\theta = O_P(n^{-1/2})$. Observe further that

$$\Phi(b_1,\dots,b_p,|\varepsilon_{t-1}|,|\varepsilon_{t-2}|,\dots) - \Phi(b_{n,1},\dots,b_{n,p},|\varepsilon_{t-1}|,|\varepsilon_{t-2}|,\dots) = \sum_{\ell=1}^{\infty} \sum_{1\le j_1,\dots,j_\ell\le p} (b_{j_1}\cdots b_{j_\ell} - b_{n,j_1}\cdots b_{n,j_\ell})\,|\varepsilon_{t-j_1}|\cdots|\varepsilon_{t-j_1-\dots-j_\ell}|.$$

Next write $b_{j_1}\cdots b_{j_\ell} - b_{n,j_1}\cdots b_{n,j_\ell} = c^\ell_{n,1}\,(b_1 - b_{n,1}) + \cdots + c^\ell_{n,p}\,(b_p - b_{n,p})$ where $c^\ell_{n,i}$ either equals zero, or $c^\ell_{n,i} = c_{n,1}\cdots c_{n,\ell-1}$ with $c_{n,i} \in \{b_1,\dots,b_p,b_{n,1},\dots,b_{n,p}\}$. We can now conclude that condition (i) holds if we assume that for all $c > 0$ we have for n large enough that $\mathrm{E}\big(\Phi(\widetilde{b}_1,\dots,\widetilde{b}_p,|\varepsilon_t|,|\varepsilon_{t-1}|,\dots)\big)^{2+\delta} < \infty$ uniformly in $\|\widetilde{\mathbf b} - \mathbf b\| < c/\sqrt{n}$ for some $\delta > 0$.

To assure the assumed equicontinuity of $W_n(C), C \in \mathcal{C}$ we need the index class $\mathcal{C}$ to be not too rich, in the sense that the so-called 'uniform bracketing integral' is finite. Without going into further detail here, we can say that this assumption is fulfilled for the class of ellipsoids that we are proposing to use in the applications. An estimator $\widehat{\boldsymbol\theta}$ that allows for the required expansion (3.10) with $h(\varepsilon_t) = \varepsilon_t^2$ can be found in Hall and Yao (2003), p. 304, for the quasi-maximum-likelihood estimator of a GARCH-model (with the ARCH-model being a special case). Besides assumptions on the complexity of $\mathcal{C}$, validity of (ii) in general requires assumptions on the dependence of the Y_t. Explicit conditions can be found in Doukhan et al. (1995), for instance.

4 Is higher volatility associated with rarer events?

In this section, we continue to explore the relationship between the two probability measures F and G in order to investigate some qualitative characteristics of the volatility function σ_t under the general framework (1.1). We will provide a statistical test to check whether the financial market is more volatile at occurrence of rarer events, which is reflected in GARCH models. For example, a simple ARCH(1) specifies volatility function as $\sigma^2_{t+1} = a + bY_t^2$ $(a, b \ge 0)$. Therefore large positive or large negative values of Y_t lead to large values of σ^2_{t+1}, while the chance of having excessive returns (i.e. $|Y_t|$ is excessively large) is small.

We use the same notation as in section 2. To facilitate our discussion, we assume $\sigma_t = \sigma(\mathbf{X}_t)$, where $\sigma(\cdot)$ is an unknown function. Recall $f(\cdot)$ is the density function of $\mathbf{X}_t$ and $M_F(\alpha)$ is the MV set of F; see (2.5). The measure G admits the density function $g(\cdot) = \nu_y^{-1}\sigma(\cdot)f(\cdot)$; see the definition of $G(\cdot)$ in (2.3). We consider to test the null hypothesis

$$H_0: \; F\{M_F(\alpha)\} = G\{M_F(\alpha)\} \quad \text{for all } \alpha \in (0,1) \tag{4.1}$$

against the one-sided alternative

$$H_1: F\{M_F(\alpha)\} \geq G\{M_F(\alpha)\} \quad \text{for all } \alpha \in (0,1), \tag{4.2}$$

$$\textit{and the inequality holds} \text{ strictly } \textit{for some } \alpha \in (0,1).$$

When the above H_0 is rejected, it indicates that the probability mass under G is more spreading-out than that under F. Note that $g(\cdot) \propto \sigma(\cdot)f(\cdot)$. Hence the G-measure is more spreading-out than the F-measure when $\sigma(\cdot)$ makes the probability mass thinner where $f(\cdot)$ is large, and thicker where $f(\cdot)$ is small. This phenomenon will occur when, for example, f is unimodal and decays to 0 at boundaries and $\sigma(\cdot)$ is a U-shape curve. The latter feature is termed "smiling effect" in volatility literature; see, for example, Härdle and Tsybakov (1997). It reflects the stylized feature that a financial market is more volatile when returns are large, either positively or negatively. Note that we only compare the two measures over the MV sets under F since we only look for the evidence that G is more spread-out than F in relation to the central areas of the data, which is reflected by the MV sets under F.

Hypothesis (4.2) (ignoring the strict inequalities) defines an integral stochastic order. In fact, by interpreting the probabilities in (4.2) as expected values of indicator functions of the MV sets, and taking linear combinations, one can see that (4.2) implies inequalities of the form $\int h(x)\, dG(x) \leq \int h(x) dF(x)$ for functions h whose level sets (i.e. the sets $\{\, x : h(x) \geq \lambda \,\}$, $\lambda \geq 0$) are MV sets of f. This means that (4.2) defines in integral stochastic order with generator consisting of all functions with level sets in $\{\, M_F(\alpha),\, \alpha \in [0,1] \,\}$. Recall that level sets of f are MV sets, and hence every (positive) function with each if its level set also being a level set of f is a member of the class of generators. This implies many integral relations between f and g, as for instance (by taking $h = f^k$) we obtain $\int f^k g \leq \int f^{k+1}$. See, e.g. Müller and Stoyan (2002) for details on integral stochastic orders.

We should reject the null hypothesis (4.2) when $(F - G)\{M_F(\alpha)\}$ takes large positive values for some $\alpha \in (0,1)$. Therefore we may define a test statistic as

$$T_5 = \int_0^1 \{ \widehat{\Delta}_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) \}^+ d\alpha$$

where $\widehat{\Delta}_n$ is defined as in (2.8), and we reject (4.2) for some large values of T_5.

Theorem 4.1 *Suppose that the conditions of Theorem 2.1 hold, and $F \equiv G$. Then as $n \to \infty$, it holds that*

$$\sqrt{\frac{n}{\widehat{\sigma}_y^2/\widehat{\nu}_y^2}}\, T_5 \xrightarrow{\mathcal{D}} \int_0^1 (\, B(\alpha)\,)^+ \, d\alpha,$$

where $B(\alpha)$ is the standard Brownian bridge, and $\widehat{\sigma}_y$ and $\widehat{\nu}_y$ are the same as in (2.10).

We tabulate the high quantiles of the random variable $\int_0^1 (\, B(\alpha)\,)^+ \, d\alpha$ below, which were obtained from a Monte Carlo simulation.

α	0.900	0.925	0.950	0.975	0.990	0.995
quantile	0.383	0.388	0.478	0.563	0.670	0.717

Remark 4.2 The setting (4.1) and (4.2) implicitly implies that we are concerned with the processes for which $F\{M_F(\alpha)\} \geq G\{M_F(\alpha)\}$ for all α. In practice, we may apply a pre-test with the test statistic $T_6 = \int_0^1 \{ \widehat{\Delta}_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) \}^- d\alpha$, where $x^- = \max(-x, 0)$. Obviously, T_6 shares the same asymptotic distribution as T_5. We should not proceed with the test T_5 if the pre-test with T_6 is significant (i.e., for example, $T_6 > 0.478$; see the table above).

Diagnostic plots. The CC-plots discussed in section 2.1 also serve as a diagnostic plot associated with the test T_5. Now, however, we are only interested in one-sided deviations.

5 Numerical illustration

We illustrate the proposed tests with numerical examples. For tests T_1 and T_5, we always choose $\widehat{M}_{\mathcal{C},F_n}(\alpha)$ among the sets with balls as their images under the mapping $\mathbf{X}_t \to \mathbf{S}^{-1/2}(\mathbf{X}_t - \bar{\mathbf{X}})$, where $\bar{\mathbf{X}}$ and $\mathbf{S}$ denote, respectively, the sample mean and the sample covariance matrix of $\{\mathbf{X}_t\}$; see (2.6). We choose $\widehat{M}_{\mathcal{C},G_n}(\alpha)$ among the sets also with balls as their images but under a different mapping $\mathbf{X}_t \to \mathbf{S}_g^{-1/2}(\mathbf{X}_t - \bar{\mathbf{X}}_g)$, where $\bar{\mathbf{X}}_g$ and $\mathbf{S}_g$ denote the mean and the covariance matrix of the distribution G_n. For test T_3, $\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)$ are defined in the same manner as $\widehat{M}_{\mathcal{C},G_n}(\alpha)$ with G_n replaced by $G_{n,\widehat{\boldsymbol{\theta}}}$ defined in (3.3).

Before we proceed to numerical experiements, we would like to point out that the tests proposed in this paper do not facilitate a direct comparison with the existing methods. As indicated in the beginning of section 2, the existing methods based on the residual-regression idea are for univariate $\mathbf{X}_t$ for which there is no point to adopt the MV-set or inverse regression. On the other hand, those parametric tests based on the residuals obtained from fitting ARCH/GARCH models with Gaussian innovation cannot apply to heavy-tailed Models I and II in (5.1) below, and would be in disadvantage when applying to a 'wrong' model such as Model III.

5.1 Simulated examples

First we deal with tests T_1, T_2 and T_5 with the data generated from three different models:

$$\begin{array}{ll} \text{I.} & Y_t = e_t, \\ \text{II.} & Y_t = \sigma_t e_t, \quad \sigma_t^2 = 0.5 + 0.1 Y_{t-1}^2 + 0.8 \sigma_{t-1}^2, \\ \text{III.} & Y_t = \sigma_t \varepsilon_t, \quad \sigma_t = 0.5 + 0.2|Y_{t-1}| + 0.75 \sigma_{t-1}, \end{array} \tag{5.1}$$

where ε_t be independent $N(0,1)$ random variables, and e_t be independent t_3-distributed random variables. The processes defined by model I is a sequences of i.i.d. random variables with $E|Y_t|^3 = \infty$, which roles out the use of conventional parametric testing methods; see section 4.2.6 of Fan and Yao (2003). Model II is a GARCH(1,1) with heavy tailed innovations. Model III defines a power GARCH(1,1) model with power index equal to 1.

For sample sizes 50, 100, 200 and 500, we applied the tests with $p = 1, 2, 4$ and 6. For each setting, we replicated the simulation 500 times. The relative frequencies of rejecting the null hypotheses at level 10%, 5% and 1% are listed in Tables 1 − 3.

Table 1 contains the results from applying tests to the independent t_3 observations. It indicates that the asymptotic approximations (2.10), (2.12) and Theorem 4.1 are adequate. The asymptotic approximations (2.10) and Theorem 4.1 remain the same for different values of p. With the range of sample sizes specified in the simulation, Table 1 suggests that overall those two approximations work

Table 1: Relative frequencies of rejecting null hypotheses; the three numbers in each entry corresponding to, respectively, the significance level 10%, 5% and 1%. The observations are drawn from Model I.

p	n	T_1			T_2			T_5		
1	50	.04	.00	.00	.05	.02	.00	.08	.03	.00
	100	.04	.01	.00	.04	.01	.00	.11	.05	.01
	200	.07	.04	.01	.07	.03	.00	.09	.05	.02
	500	.07	.03	.00	.06	.03	.00	.10	.05	.01
2	50	.03	.01	.00	.05	.02	.01	.07	.07	.00
	100	.04	.02	.00	.04	.01	.00	.09	.04	.01
	200	.04	.01	.00	.07	.03	.00	.10	.05	.01
	500	.06	.03	.00	.08	.03	.00	.10	.05	.01
4	50	.04	.01	.00	.04	.02	.01	.07	.03	.00
	100	.05	.03	.00	.04	.03	.01	.08	.03	.00
	200	.05	.02	.00	.04	.03	.00	.07	.04	.01
	500	.06	.03	.01	.07	.04	.01	.10	.05	.01
6	50	.09	.04	.01	.04	.02	.01	.05	.02	.00
	100	.06	.03	.00	.03	.02	.01	.05	.01	.00
	200	.06	.02	.00	.04	.02	.00	.07	.04	.01
	500	.07	.03	.01	.06	.04	.01	.09	.05	.01

Table 2: Legend is as for Table 1. The observations are drawn from Model II: GARCH(1,1) model with t_3-distributed innovations.

p	n	T_1			T_2			T_5		
1	50	.11	.05	.01	.23	.16	.05	.34	.27	.13
	100	.43	.33	.16	.57	.48	.33	.66	.60	.44
	200	.80	.74	.60	.89	.82	.74	.92	.89	.81
	500	1.00	1.00	.98	1.00	1.00	.99	1.00	1.00	1.00
2	50	.14	.07	.01	.25	.18	.06	.46	.35	.17
	100	.55	.45	.30	.62	.41	.35	.80	.73	.58
	200	.92	.88	.79	.92	.89	.77	.97	.96	.93
	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
4	50	.16	.10	.01	.20	.14	.04	.50	.40	.22
	100	.62	.56	.38	.60	.51	.34	.87	.79	.65
	200	.95	.93	.86	.92	.89	.78	.98	.97	.95
	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
6	50	.14	.07	.01	.17	.12	.06	.46	.36	.19
	100	.64	.55	.37	.58	.48	.32	.86	.81	.70
	200	.96	.94	.88	.95	.90	.79	.99	.98	.96
	500	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 3: Legend is as for Table 1. The observations are generated from Model III: power-GARCH(1,1) model with Gaussian innovations.

p	n	T_1			T_2			T_5		
1	50	.10	.05	.01	.14	.08	.03	.32	.24	.09
	100	.28	.20	.08	.34	.24	.11	.60	.49	.28
	200	.61	.49	.30	.61	.52	.32	.84	.74	.58
	500	.96	.93	.84	.96	.92	.80	.99	.99	.95
2	50	.10	.06	.01	.14	.08	.02	.39	.28	.12
	100	.37	.24	.10	.31	.20	.11	.71	.60	.35
	200	.73	.65	.46	.65	.55	.37	.92	.87	.73
	500	.99	.99	.95	.97	.95	.84	1.00	1.00	.99
4	50	.12	.07	.02	.14	.07	.01	.40	.29	.13
	100	.41	.29	.15	.33	.23	.10	.73	.64	.41
	200	.80	.71	.53	.61	.51	.31	.93	.88	.79
	500	1.00	0.99	.98	.97	.95	.82	1.00	1.00	1.00
6	50	.10	.06	.01	.14	.09	.03	.36	.25	.12
	100	.38	.28	.14	.29	.21	.09	.70	.60	.40
	200	.75	.69	.52	.62	.52	.32	.93	.89	.76
	500	.99	.99	.97	.96	.91	.83	1.00	1.00	0.99

about equally well for p between 1 and 6. Note that the reported relative frequencies are almost always smaller than the nominal levels. This will underplay the potential power of the tests.

With the data generated by the two heteroscedastic models, i.e. Models II and III, our tests demonstrate the power in rejecting both the null hypotheses (2.1) and (4.2); see Tables 2 and 3. In fact the power of rejection increases as the sample size n increases. Note that for both Models II and III, σ_t depends on infinite number of lagged values of Y_t, Tables 2 and 3 show that the tests with $p = 6$ and 4 are more powerful than those with $p = 1$ and 2 in most cases.

The diagnostic plots associated with the tests T_1 and T_5 are presented in Figure 1, and those associated with the test T_2 are displayed in Figure 2. To save the space we only presented the plots with $n = 200$ and $p = 2$. For each of Models I – III, five samples were randomly selected. Under the null hypothesis (2.1), those plots are closely around the diagonal line $y = x$; see the top row in both Figures 1 and 2. When there exists heteroscedasticity, some parts of the curves drifted away from the diagonal line; see Rows 2 and 3 in Figures 1 and 2. The departure is more pronounced for GARCH(1,1) model with heavy tailed innovations.

For testing the null hypothesis (4.2) against one-sided alternative (4.1) with statistic T_5, we look

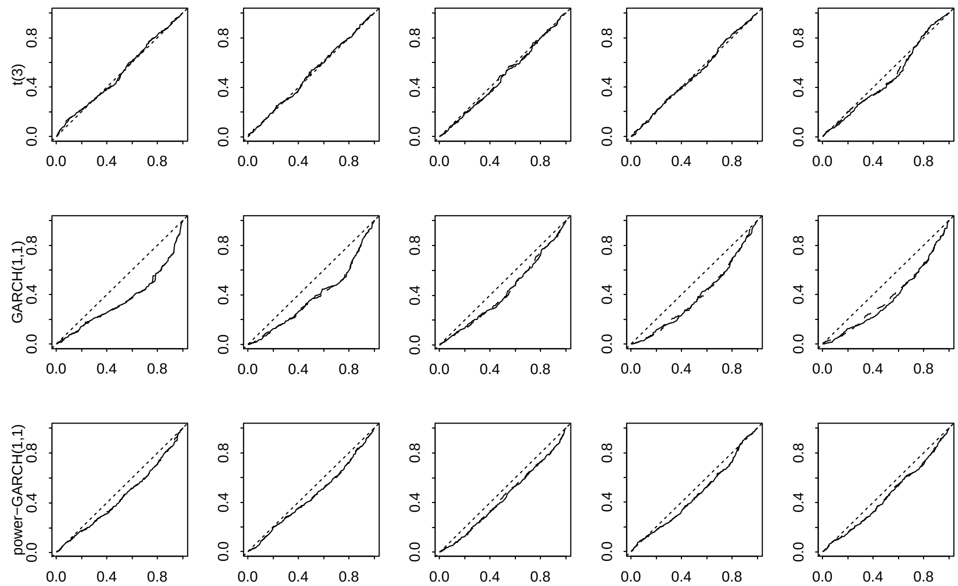

Figure 1: *Diagnostic plots associated with tests* T_1 *and* T_5 *(with* $p = 2$*): solid lines –* $G_n\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\}$ *against* $F_n\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\}$ *for* $\alpha \in (0, 1)$*; dashed lines –* $G_n\{\widehat{M}_{\mathcal{C},G_n}(\alpha)\}$ *against* $F_n\{\widehat{M}_{\mathcal{C},G_n}(\alpha)\}$ *for* $\alpha \in (0, 1)$*; dotted lines – straight line* $y = x$*. Each row represents 5 randomly selected samples (with* $n = 200$*) from, from top to bottom, each of Models I – III.*

for the one-sided departure of the solid curves under the diagonal $y = x$ in Figure 1. This is evident in most plots in Rows 2-3 there. This indicates that "smiling effect" may well exist for all the processes defined by Models II and III, and again such an effect is more pronounced for the GARCH(1,1) model with t_3 innovations.

Table 4: Simulation results for parametric bootstrap (PB) and nonparametric bootstrap (nonPB) tests based on T_3 and T_4. The three numbers in each entry are the relative frequencies of rejecting the ARCH(2) null hypothesis, corresponding to, respectively, the level $\alpha = .10, .05$ and $.01$.

Model	n	T_3 (PB)			T_3 (nonPB)			T_4 (PB)			T_4 (nonPB)		
ARCH(2)	200	.11	.05	.01	.10	.05	.01	.10	.04	.01	.12	.05	.01
(with t_3-innovation)	200	.00	.00	.00	.11	.04	.01	.09	.03	.00	.11	.04	.01
ARCH(3)	200	.21	.11	.04	.12	.06	.02	.09	.04	.02	.07	.03	.01
	500	.30	.24	.12	.19	.12	.04	.13	.06	.01	.08	.04	.01
GARCH(1,1)	200	.97	.96	.87	.96	.94	.86	.45	.22	.06	.40	.21	.05
	500	1.00	1.00	1.00	1.00	1.00	1.00	.94	.80	.35	.92	.73	.26
PGARCH(1,1)	200	1.00	1.00	1.00	.99	.99	.99	.62	.12	.00	.68	.20	.00
	500	1.00	1.00	1.00	1.00	1.00	1.00	.96	.63	.00	.97	.72	.00

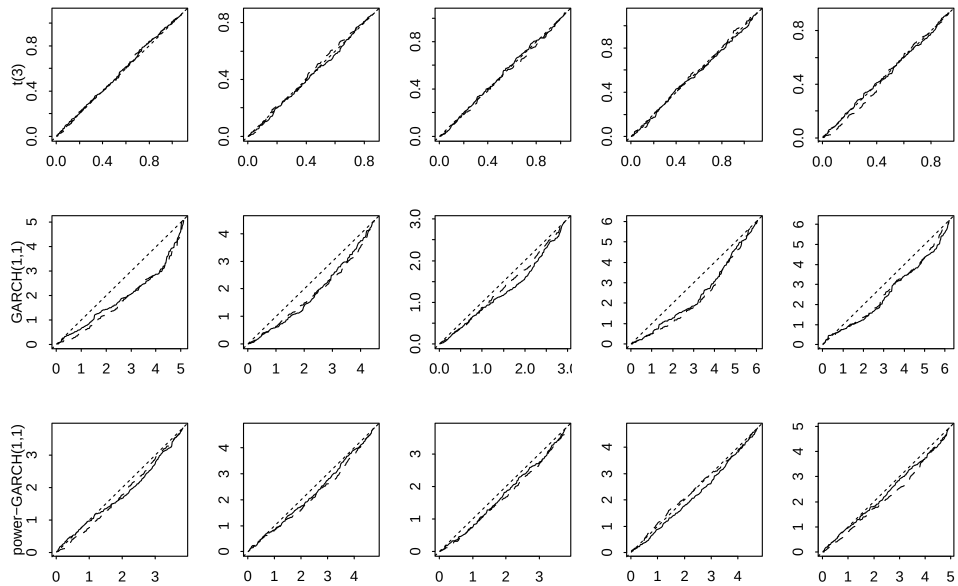

Figure 2: *Diagnostic plots associated with test* T_2 *(with* $p = 2$*):* $\frac{1}{\sqrt{n}} \sum_t |Y_{t-j}| I(|Y_t| \leq y)$ *against* $\frac{\widehat{\nu}_{y,j}}{\sqrt{n}} \sum_t I(|Y_t| \leq y)$ *for* $y > 0$*. Solid lines –* $j = 1$*, dashed lines –* $j = 2$*, dotted lines – straight line* $y = x$*. Each row represents 5 randomly selected samples (with* $n = 200$*) from, from top to bottom, each of Models I-III.*

We also conducted a simulation study on bootstrap tests with test statistics T_3 and T_4 for testing a specified null hypothesis (3.1) with $\sigma_t^2 = c + a_1 Y_{t-1}^2 + a_2 Y_{t-2}^2$, i.e. Y_t is a ARCH(2) model. We applied both the parametric bootstrap test with normal distribution F_1, and the nonparametric bootstrap test outlined in section 4.3. Samples of size $n = 200$ or 500 were generated from model (1.1) with different forms of σ_t. For each setting, we drew 400 samples, and repeated bootstrap sampling also $B = 400$ times.

Table 4 reports the relative frequencies of rejecting the null hypothesis of ARCH(2) model in the 400 replications at the significance levels 10%, 5% and 1%. The first row contains the results for an ARCH(2) process with Gaussian innovation and the coefficients $(c, a_1, a_2) = (0.5, 0.4, 0.5)$. Now the null hypothesis H_0 is true. Both parametric and nonparametric bootstrap methods provide very accurate estimates for the significance levels. The results for the same model but with t_3-innovations are reported in the second row of the table. As expected, the nonparametric bootstrap method still provide accurate estimates for the significance levels. However the parametric bootstrap method failed with statistic T_3. This was due to the use of a wrong innovation distribution in bootstrapping.

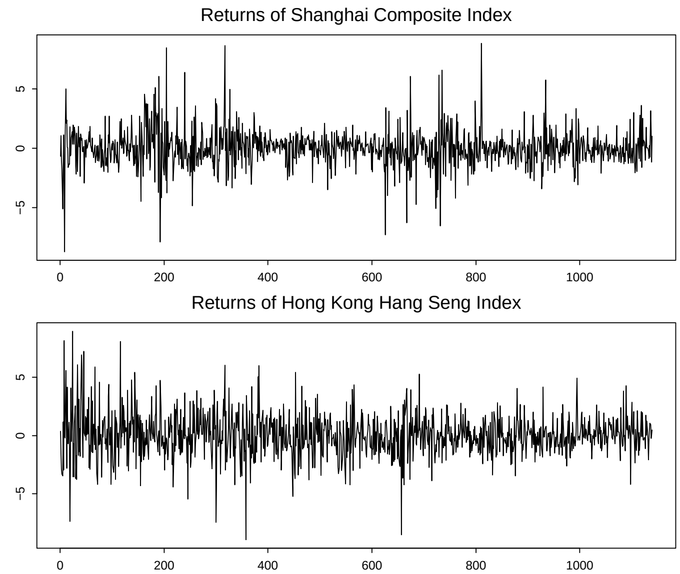

Figure 3: *Time plots of daily returns of Shanghai Composite Index and Hong Kong Hang Seng Index in 3 August 1998 – 30 December 2003.*

In Table 4, Rows 3–4 list the results for an ARCH(3) model with Gaussian innovations and the coefficients $(c, a_1, a_2, a_3) = (0.5, 0.3, 0.2, 0.4)$, Rows 5-6 for for GARCH(1,1) model IV but with Gaussian innovations, and the last two rows for power-GARCH(1,1) model V. Even with sample size $n = 500$, all the tests lack the power to tell the 'subtle' difference between ARCH(3) and ARCH(2). On the other hand, the tests based on MV sets (i.e. T_3) are considerably more powerful than the tests based inverse regression (i.e. T_4). Furthermore, the nonparametric bootstrap tests are almost always less powerful than the parametric bootstrap tests which used the correct innovation distribution in bootstrap samplings.

5.2 Real data examples

Now we illustrate the tests with two real data sets: the returns of daily close prices of Shanghai Composite Index and Hong Kong Hang Seng Index in 3 August 1998 – 30 December 2003; see Figure 3. We applied the tests to the whole series ($n = 1139$), as well as to the three subseries (for each of the two data sets) in 3 August 1998 – 30 June 2000, 3 July 2000 – 31 May 2002 and 3 June 2002 – 30 Dec 2003 with sample sizes, respectively, 400, 395 and 344.

For the returns of Shanghai Composite Index, the tests T_1 and T_2 with $p = 1, 2$ and 4 are all

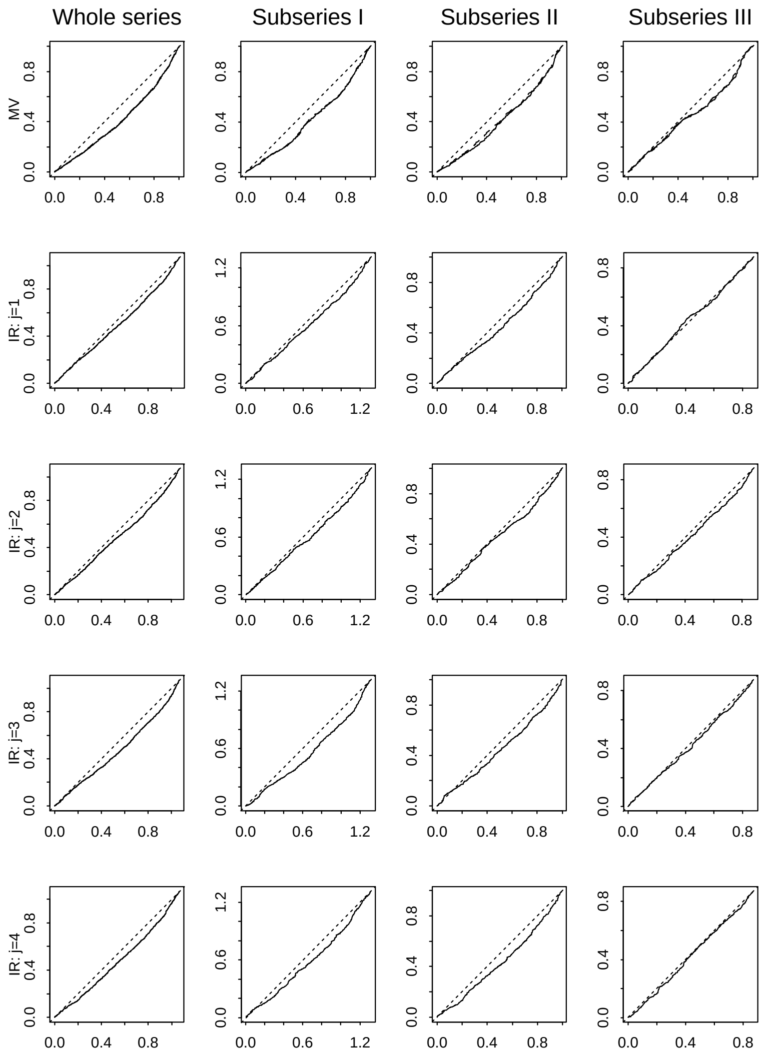

Figure 4: *Diagnostic plots associated with* T_1 *and* T_2 *(with* $p = 4$*) for returns of Shanghai Composite Index. Row I:* $G_n\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\}$ *against* $F_n\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\}$ *for* $\alpha \in (0,1)$ *– solid lines;* $G_n\{\widehat{M}_{\mathcal{C},G_n}(\alpha)\}$ *against* $F_n\{\widehat{M}_{\mathcal{C},G_n}(\alpha)\}$ *for* $\alpha \in (0,1)$ *– dashed lines. Rows II-V:* $\frac{1}{\sqrt{n}} \sum_t |Y_{t-j}| I(|Y_t| \le y)$ $(y > 0)$ *against* $\frac{\widehat{\nu}_{y,j}}{\sqrt{n}} \sum_t I(|Y_t| \le y)$ *for* $j = 1, 2, 3, 4$*. Dotted lines are diagonal* $y = x$*.*

significant at the level 1% for the whole series and the first subseries, and T_1 and T_2 with $p = 2$ and 4 are significant at the level 1% for the second and the third subseries. The test T_1 with $p = 1$ is not significant (with P-value greater than 0.1) for both the second and the third subseries while T_2 with $p = 1$ is significant at level 5% for both the subseries. This indicates that there is overwhelming evidence to reject the null hypothesis of a constant volatility function. It also echoes the observation in the simulation study that the tests with $p > 1$ are more powerful than those with $p = 1$ for

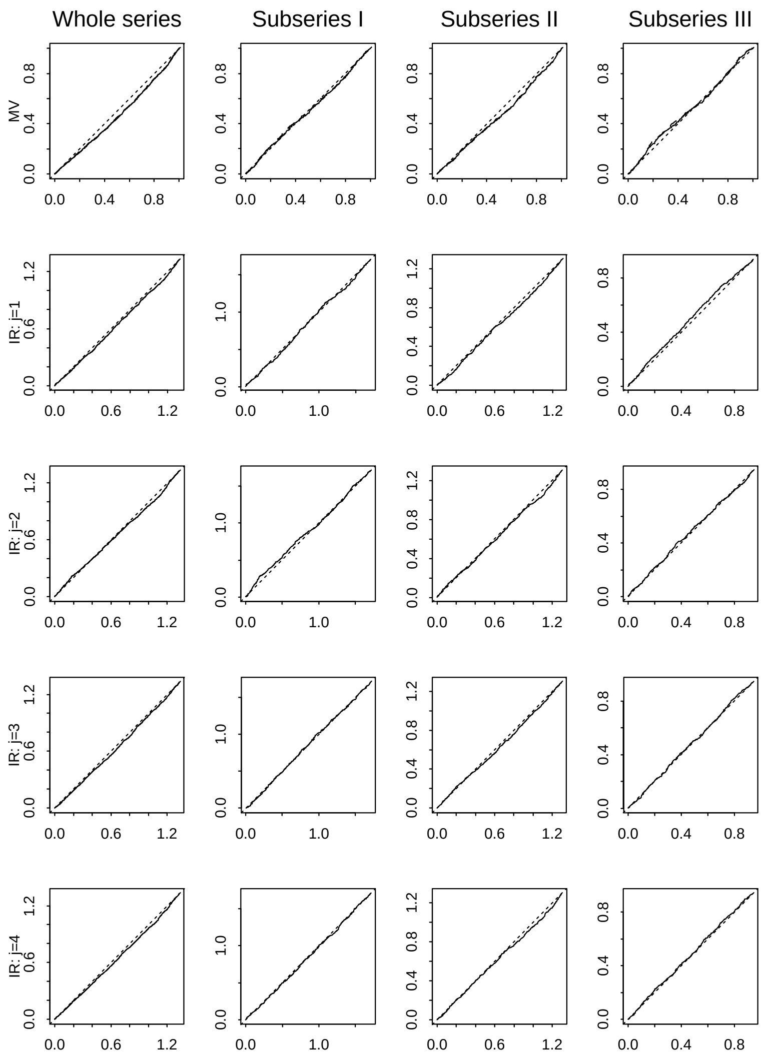

Figure 5: *Diagnostic plots associated with T_1 and T_2 (with $p = 4$) for returns of Hang Kong Hang Seng Index. Row I: $G_n\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\}$ against $F_n\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\}$ for $\alpha \in (0,1)$. Rows II-V: $\frac{1}{\sqrt{n}}\sum_t |Y_{t-j}| I(|Y_t| \leq y)$ against $\frac{\widehat{\nu}_{y,j}}{\sqrt{n}}\sum_t I(|Y_t| \leq y)$ $(y > 0)$ for $j = 1, 2, 3, 4$. Dotted lines are diagonal $y = x$.*

rejecting (2.1) when $\sigma(\cdot)$ depends on more than one lagged values. Figure 4 displayed the associated diagnostic plots. It clearly indicates the evidence of the departure for the null hypothesis for the whole series as well as all the three subseries.

We also applied the tests T_3 and T_4 for testing the null hypothesis of an ARCH(2) model $\sigma_t^2 = a_0 + a_1 Y_{t-1}^2 + a_2 Y_{t-2}^2$. The quasi-MLE for (a_0, a_1, a_2) for the three subseries were, respectively (1.906, 0.405, 0.128), (1.380, 0.118, 0.325) and (0.848, 0.239, 0.335), which were very different from each other. The nonparametric bootstrap tests with both T_3 and T_4 and the parametric bootstrap test

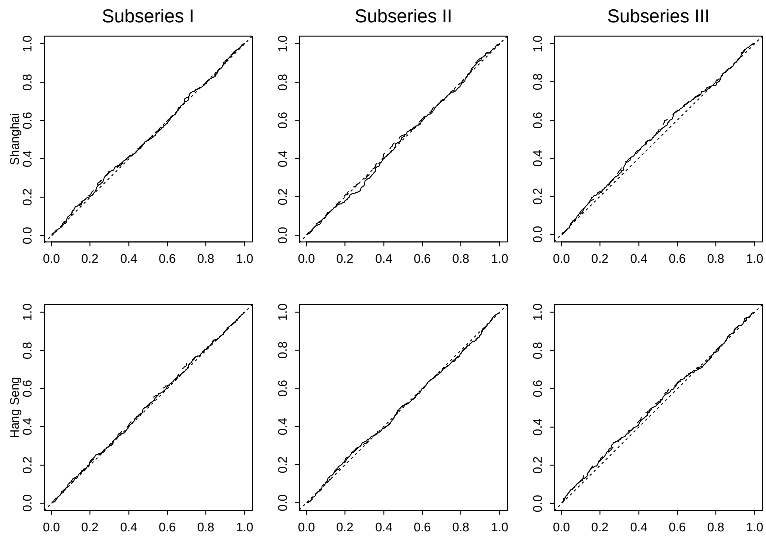

Figure 6: *Diagnostic plots associated with T_3 for subseries of both Shanghai Composite returns and Hang Kong Hang Seng returns. Solid lines – $G_{n,\widehat{\boldsymbol{\theta}}}\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\}$ against $F_n\{\widehat{M}_{\mathcal{C},F_n}(\alpha)\}$ for $\alpha \in (0,1)$; dashed lines – $G_{n,\widehat{\boldsymbol{\theta}}}\{\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\}$ against $F_n\{\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\}$ for $\alpha \in (0,1)$; dotted lines – $y = x$.*

with T_4 using Gaussian innovation distribution do not reject the null hypothesis of ARCH(2) model for Subseries I and II with minimum P-value 0.27. The parametric bootstrap test with T_3 yields the P-values 0.08 and 0.41 for Subseries I and II with Gaussian innovation distribution, and P-values 0.24 and 0.57 with t_7 innovation distribution. Overall there is no significant evidence to against ARCH(2) model for the first two subseries. For subseries III, the parametric bootstrap test with T_3 using either Gaussian or t_7 innovation distributions yields the P-value 0.00 while the nonparametric bootstrap test with T_3 yields the P-value 0.02; indicating the null hypothesis of ARCH(2) should be rejected. The associated diagnostic plots are displayed in Figure 6. It indeed shows a certain degree of the inadequacy of ARCH(2) model for Subseries III.

Interestingly the returns of Hong Kong Hang Seng Index showed drastically different behavior. For example, the test T_1 was not significant with $p = 1, 2$ for all three subseries, and was only significant with $p = 4$ for Subseries II at level 5% and Subseries III at level 10% only. The test T_2 was significant for all the three subseries with $p = 2, 4$ at the level 5% or 10%, was not significant with $p = 1$. For the whole series, both T_1 and T_2 were significant at level 5% with $p = 1$, and were significant at level 1% with $p = 2, 4$. One may argue that the substantial increase in significance might be due the non-stationarity rather than a genuine conditional heteroscedasticity. The associated diagnostic plots are displayed in Figure 5. In contrast to Figure 4, Figure 5 indicates little evidence of the departure from the null hypothesis, especially for the three subseries.

We further fitted ARCH(2) model to the three subseries. The estimated coefficients for (a_0, a_1, a_2) were, respectively (5.279, 0.012, 0.000), (2.438, 0.109, 0.047) and (1.589, 0.000, 0.000). This also supports a constant conditional variance model for Subseries I and III. Not surprisingly both the parametric (with Gaussian innovation distribution) and nonparametric bootstrap tests with both T_3 and T_4 are not significant; indicating that the null hypothesis of an ARCH(2) model could not be rejected for all three subseries. The diagnostic plots in Figure 6 reinforces this argument.

Finally we applied test T_5 to explore the existence of 'smiling effect' in the subseries of those two data sets. We first applied the pre-test T_6 to examine the evidence against the prerequisite $F\{M_F(\alpha)\} \geq G\{M_F(\alpha)\}$ for any α; see Remark 4.2. The P-values of the pre-test with $p = 1, 2$ and 4 are always greater than 0.1. Hence we may proceed with the test T_5 now.

For Shanghai composite returns, T_5 is siginificant with $p = 2, 4$ at level 1% for all the three subseries. It is significant with $p = 1$ at level 1% for Subseries I, at level 5% for Subseries II, and not significant for Subseries III; see also the plots in Row I in Figure 4. Overall there is evidence to indicate that the 'smiling effect' exists with all the three subseries.

Again the Hang Kong Hang Seng returns show different behavior. The tests T_5 with $p = 1, 2$ and 4 are all not significant even at the level 10% for Subseries I and III, are significant at level 5% for Subseries II. The diagnostic plots in Figure 5 (Row I) show that there might be some evidence for $F\{M_F(\alpha)\} > G\{M_F(\alpha)\}$ for some α for Subseries II. However the departure from the diagonal is much less pronounced than that for Shanghai composite returns; see Figure 4.

6 Asymptotic theory

In this section we present the proofs of the theoretical result formulated above. We first introduce the notion of metric entropy and covering integral which are used to control richness of $\mathcal{C}$. Assume that for a given $\varepsilon > 0$ there exists a partition $\{\mathcal{C}_1, \ldots, \mathcal{C}_N\}$ of $\mathcal{C}$ such that for each $k = 1, \ldots, N$ there exist sets $C_{*,k}$ and C_k^* with $\mathcal{C}_k = \{C \in \mathcal{C} : C_{*,k} \subset C \subset C_k^*\}$ and $\sqrt{F(C_k^* \setminus C_{*,k})} < \varepsilon$, and let $N_B(\varepsilon, \mathcal{C}, L_2(F))$ denote the smallest such N. Then the metric entropy is defined as the logarithm of $N_B(\varepsilon, \mathcal{C}, L_2(F))$.

Further let $H(\varepsilon) = H_B(\varepsilon, \mathcal{C}, L_2(F))$ be such that $\log N_B(\varepsilon, \mathcal{C}, L_2(F)) \leq H_B(\varepsilon, \mathcal{C}, L_2(F))$. The covering integral then is

$$I_B(\mathcal{C}) := \int_0^1 \sqrt{H_B(\varepsilon, \mathcal{C}, L_2(F))}\, d\varepsilon \; < \; \infty. \tag{6.1}$$

When $F = G_{\boldsymbol{\theta}}$ we write $I_B(\mathcal{C}, \boldsymbol{\theta})$. Observe that with this notation we have $I_B(\mathcal{C}) = I_B(\mathcal{C}, \boldsymbol{\theta}_0)$. There exist many classes $\mathcal{C}$ having a finite covering integral, as for instance intervals, balls, ellipsoids, rectangles, certain classes of sets with smooth boundaries, the class of convex sets in $\mathcal{R}^2$ etc. Notice, however, that the condition of a finite covering integral might depend on the underlying distribution F. We refer to standard textbooks on empirical process theory for further details.

Proof of Theorem 2.1. We first prove the second part. We will show below that for any $\alpha \in [0, 1]$ we have

$$G_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) - G(M_F(\alpha)) = o_P(1), \tag{6.2}$$

and the same result holds if in (6.2) we replace G_n, G by F_n, F, respectively, or if we replace F_n, F by G_n, G, respectively. Suppose for the moment that (6.2) actually holds. Then under H_1 there exist $\alpha_0 > 0$ with either $F(\,M_F(\alpha_0)\,) \neq G(\,M_F(\alpha_0)\,)$ or $F(\,M_G(\alpha_0)\,) \neq G(\,M_G(\alpha_0)\,)$. It follows that as $n \to \infty$ we have

$$\begin{aligned} T_1 &\geq |(G_n - F_n)(\,\widehat{M}_{\mathcal{C},F_n}(\alpha_0)\,)\,| + |(G_n - F_n)(\,\widehat{M}_{\mathcal{C},G_n}(\alpha_0)\,)\,| \\ &\to_p \; |\,G(\,M_F(\alpha_0)\,) - F(\,M_f(\alpha_0)\,)\,| + |\,G(\,M_G(\alpha_0)\,) - F(\,M_g(\alpha_0)\,)\,| > 0. \end{aligned}$$

It remains to prove (6.2). To this end first notice that $G_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) - G(M_F(\alpha)) = (\,G_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) - G(\widehat{M}_{\mathcal{C},F_n}(\alpha))\,) + (\,G(\widehat{M}_{\mathcal{C},F_n}(\alpha)) - G(M_F(\alpha))\,)$. The latter summand is $o_P(1)$ since we have $|\,G(\widehat{M}_{\mathcal{C},F_n}(\alpha)) - G(M_F(\alpha))\,| \leq G(\,\widehat{M}_{\mathcal{C},F_n}(\alpha) \Delta M_F(\alpha)\,) = o_P(1)$ (by Proposition 6.1). Further notice that for any set A the assumed ergodicity of Y_t implies that $G_n(A) \to G(A)$ almost surely. Using the standard bracketing trick (see e.g. Pollard 1984, or van der Vaart and Wellner 1995) we can conclude that $\sup_{C \in \mathcal{C}} |\,(\,G_n - G\,)(A)\,| = o_P(1)$. In particular this implies that $\sup_{0 \leq \alpha \leq 1} |(\,G_n - G\,)(\widehat{M}_{\mathcal{C},F_n}(\alpha))| = o_P(1)$. This shows (6.2). The proof of (6.2) with G_n, G replaced by F_n, F (and with F_n, F replaced by G_n, G) is analogous.

Now we prove the first assertion. We have

$$\sqrt{n}\,\widehat{\Delta}_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) = \frac{1}{\sqrt{n}\,\widehat{\nu}_y}\sum_{i=1}^{n}(\,|Y_t| - \nu_y\,)\left(I(\,\mathbf{X}_t \in \widehat{M}_{\mathcal{C},F_n}(\alpha)\,) - F(\widehat{M}_{\mathcal{C},F_n}(\alpha))\,\right)$$
$$+\ \frac{1}{\widehat{\nu}_y}\sqrt{n}\,(\widehat{\nu}_y - \nu_y)\,(F_n - F)(\widehat{M}_{\mathcal{C},F_n}(\alpha)). \tag{6.3}$$

Notice that under the null hypothesis $\sigma(\cdot) = \nu_y$ and hence $|Y_t| = \nu_y\,|\varepsilon_t|$. It follows by the law of large numbers that $F_n(A) \to F(A)$ almost surely. Using the standard bracketing trick (notice that our assumptions imply a finite bracketing number) we can conclude that $\sup_{C\in\mathcal{C}}|\,(\,F_n - F\,)(A)\,| = o_P(1)$. Hence, under H_0 (6.3) is $o_P(1)$ uniformly in α and $\sqrt{n}\,\frac{\widehat{\nu}_y}{\nu_y}\,\widehat{\Delta}_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) = B_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) + o_P(1)$, where

$$B_n(C) = \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\,|\varepsilon_t| - 1\,)\,[\,\mathbf{1}\{\nu_y\cdot(\varepsilon_{t-1},\ldots,\varepsilon_{t-p})' \in C\} - F(C)\,].$$

Since $B_n(C)$ is a sum of m-dependent random variables, stochastic equicontinuity of the $\mathcal{C}$-indexed process B_n (with respect to $d_F(C,D) = F(C\,\Delta D)$) follows from Doukhan et al.(1995). This together with uniform consistency of $\widehat{M}_{\mathcal{C},F_n}(\alpha)$ (Proposition 6.1) implies that $B_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) = B_n(M_F(\alpha)) + o_P(1)$. Further, since under H_0 we have $M_F(\alpha) = M_G(\alpha)$ for all α we hence can approximate both $\sqrt{n}\,\widehat{\Delta}_n(\widehat{M}_{\mathcal{C},F_n}(\alpha))$ and $\sqrt{n}\,\widehat{\Delta}_n(\widehat{M}_{\mathcal{C},G_n}(\alpha))$ by the same process $B_n(M_F(\alpha))$. Now notice that the summands in $B_n(M_F(\alpha))$ have mean zero, are uncorrelated and p-dependent. Hence, the finite dimensional distributions of B_n converge to mean zero Gaussian distributions. The covariance of $B_n(M_F(\alpha))$ and $B_n(M_f(\beta))$ can easily be calculated as $\gamma(\alpha,\beta) = \mathrm{Var}(|\varepsilon_t|)\ \Big(F(M_F(\alpha)\cap M_f(\beta)) - F(M_F(\alpha))F(M_f(\beta))\Big) = \mathrm{Var}(|\varepsilon_t|)\ (\,\min(\alpha,\,\beta) - \alpha\,\beta\,)$. Up to the constant $\mathrm{Var}(|\varepsilon_t|)$ this is the covariance functions of a Brownian Bridge. The asserted limit of T_1 now follows by applying a continuous mapping theorem. □

Proof of Theorem 2.2. Notice that T_2 is the maximum of the suprema of the components of the vector-valued process $\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\mathbf{Z}_t - \widehat{\boldsymbol{\nu}}_y\right)I(\,|Y_t| \le y\,)$. Since $F = G$ (under H_0) we have $\sigma_0(\cdot,\boldsymbol{\theta}) = \nu_y$.

It follows that

$$\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\big(\mathbf{Z}_t-\widehat{\boldsymbol{\nu}}_y\big)\,I(|Y_t|\le y)=\ \frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\,\mathbf{Z}_t-\boldsymbol{\nu}_y\,)\,\Big[I\big(|\varepsilon_t|<\frac{y}{\nu_y}\big)-F_{|\varepsilon|}\Big(\frac{y}{\nu_y}\Big)\Big] \tag{6.4}$$

$$+\ \sqrt{n}\,(\widehat{\boldsymbol{\nu}}_y-\boldsymbol{\nu}_y\,)\,\frac{1}{n}\sum_{t=1}^{n}\Big(I(\,|\varepsilon_t|\le\frac{y}{\nu_y}\,)-F_{|\varepsilon|}(\frac{y}{\nu_y})\Big). \tag{6.5}$$

The centered sum of indicator functions in (6.5) tends to zero in probability uniformly in y by Glivenko-Cantelli theorem. Since by assumption $\sqrt{n}\,(\widehat{\nu}_{y,j}-\nu_y\,)=O_P(1)$, the term (6.5) is $o_P(1)$ uniformly in α. Now, $\mathbf{Z}_t=\nu_y\,(\,|\varepsilon|_{t-1},\dots,|\varepsilon|_{t-p})'$. Hence, using Doukhan et al (1994, 1995), we can conclude that each component of the vector on the r.h.s. in (6.4) converges to a Gaussian process. The components can easily be seen to be uncorrelated, leading to independent Gaussian processes in the limit. Calculation of the covariance structure is straightforward, and an application of the continuous mapping theorem completes the proof. □

Proof of Theorem 4.1. In the proof of Theorem 2.1 we showed that if $F=G$ then both the processes $\sqrt{n}\,\widehat{\Delta}_n(\widehat{M}_{\mathcal{C},F_n}(\alpha))$ and $\sqrt{n}\,\widehat{\Delta}_n(\widehat{M}_{\mathcal{C},G_n}(\alpha))$ converge to the same process, a standard Brownian Bridge times the constant $\mathrm{Var}(|\varepsilon_t|)$. An application of continuous mapping theorem concludes the proof. □

Proof of Theorem 3.1. We have

$$\begin{aligned}
\sqrt{n}\,\widehat{\Delta}_{n,\widehat{\boldsymbol{\theta}}}(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) &=\ \sqrt{n}\,(G_{n,\widehat{\boldsymbol{\theta}}}-F_n)(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))\\
&=\ \frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}\sqrt{n}}\sum_{t=1}^{n}\Big(\frac{|Y_t|}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})}-\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}\Big)\,I(\,\mathbf{X}_t\in\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\,)\\
&=\ \frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}\sqrt{n}}\sum_{t=1}^{n}\Big(\frac{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)\,|\varepsilon_t|}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})}-1\Big)\,[\,I(\,\mathbf{X}_t\in\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\,)-F_n(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))\,]
\end{aligned}$$

$$=\ \frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}\sqrt{n}}\sum_{t=1}^{n}(\,|\varepsilon_t|-1\,)\,[\,I(\,\mathbf{X}_t\in\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\,)-F(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))\,] \tag{6.6}$$

$$+\ \ \frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}\sqrt{n}}\sum_{t=1}^{n}\Big(\frac{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})}-1\Big)\,|\varepsilon_t|\,[\,I(\,\mathbf{X}_t\in\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\,)-F(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))\,]+r_n, \tag{6.7}$$

where $r_n=(F_n-F)(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))\,\frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}\sqrt{n}}\sum_{i=1}^{n}\Big(\frac{\sigma(\mathbf{X}_t)\,|\varepsilon_t|}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})}-1\Big)$. We will see below, that $r_n=o_P(1)$. Write (6.6) as $\frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}}Z_n(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))$ where $Z_n(C)=\frac{1}{\sqrt{n}}\,\sum_{t=1}^{n}(\,|\varepsilon_t|-1\,)\,[\,I(\,\mathbf{X}_t\in C\,)-F(C)\,]$. It is straightforward to see that $\widehat{\nu}_{\widehat{\boldsymbol{\theta}}} \to 1$ in probability (cf. proof of Proposition 6.1). By assumption $\{ Z_n(C), C \in \mathcal{C} \}$ is asymptotically equicontinuous with respect to the (pseudo) metric $d_F(C,D) = F(C \,\Delta\, D)$. Since we also have $\sup_\alpha F(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha) \,\Delta\, M_F(\alpha)) = o_P(1)$, (see Proposition 6.1) we obtain $Z_n(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) = Z_n(M_F(\alpha)) + o_P(1)$ (uniformly in α), and $Z_n(M_F(\alpha))$ is the first term in the asserted expansion. It hence remains to consider (6.7). By performing a one-term Taylor expansion of $\sigma_0(x,\cdot)$ we obtain that

$$
\begin{aligned}
\frac{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})} - 1 \;&=\; (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\,)' \;\frac{\dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})} \\
&\qquad + \;\frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\,)'}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})} \int_0^1 \big(\,\dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0 + \alpha\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\,) - \dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)\,\big)\, d\alpha \\
&= \Bigg[\; (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\,)' \;\frac{\dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)} \\
&\qquad + \;\frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\,)'}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)} \int_0^1 \big(\,\dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0 + \alpha\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\,) - \dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)\,\big)\, d\alpha\Bigg] \cdot (1 + o_P(1)) \\
&= \Bigg[\; (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\,)' \;\frac{\dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)} \;+\; \frac{(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\,)'}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}\; b(\mathbf{X}_t)\cdot o_P(1)\Bigg] \cdot (1 + o_P(1))
\end{aligned}
$$

where the last two equality utilize **(V2)**(b) and **(V1)**, respectively, and where $b(\cdot) = (b_1(\cdot), \ldots, b_p(\cdot))$ (cf. **(V1)**).

Plugging this into (6.7) we get

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=1}^n \Big(\frac{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})} - 1 \Big)\, |\varepsilon_t|\, I(\,\mathbf{X}_t \in \widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\,) \\
&\quad = \Bigg[\big(\sqrt{n}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big)' \Big(\frac{1}{n} \sum_{t=1}^n \frac{\dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}\, |\varepsilon_t|\; I(\,\mathbf{X}_t \in \widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\,)\Big) \qquad (6.8) \\
&\qquad + \;\; \big(\sqrt{n}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\big)' \,\frac{1}{n} \sum_{t=1}^n |\varepsilon_t|\, \frac{b(\mathbf{X}_t)}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}\, I(\,\mathbf{X}_t \in \widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\,) \;\cdot\; o_P(1)\Bigg]\; (1 + o_P(1)). \qquad (6.9)
\end{aligned}
$$

First observe that our assumptions assure that the sum in (6.9) in absolute value is less than or equal to $c\,\frac{1}{n}\sum_{t=1}^n |\varepsilon_t|\,|b(\mathbf{X}_t)| \;\to\; c\,\mathrm{E}|b(\mathbf{X}_t| \;<\; \infty$ for some $c > 0$. In particular this implies that the sum in (6.9) is $O_P(1)$ uniformly in α. Next we consider (6.8). We write the sum in (6.8) as $\mathbf{S}_n(\,\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)\,)$ where $\mathbf{S}_n(C) = \frac{1}{n}\sum_{t=1}^n \frac{\dot\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}\,|\varepsilon_t|\;\; I(\,\mathbf{X}_t \in C\,)$. It remains to show that

$\mathbf{S}_n(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) \to \mathbf{b}_0(M_F(\alpha))$. This can be seen as follows. First notice that by definition of $\mathbf{b}_{\boldsymbol{\theta}_0}(C)$ we have $E(\mathbf{S}_n(C)) = \mathbf{b}_{\boldsymbol{\theta}_0}(C)$. Ergodicity implies that $\|\mathbf{S}_n(C) - \mathbf{b}_{\boldsymbol{\theta}_0}(C)\| = o_P(1)$, and a standard bracketing argument shows that this convergence holds uniform over $C \in \mathcal{C}$. Now we write

$$\begin{aligned}
\|\mathbf{S}_n(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) - \mathbf{b}_{\boldsymbol{\theta}_0}(M_F(\alpha))\| &\leq \|\mathbf{S}_n(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) - \mathbf{b}_{\boldsymbol{\theta}_0}(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))\| + \|\mathbf{b}_{\boldsymbol{\theta}_0}(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) - \mathbf{b}_{\boldsymbol{\theta}_0}(M_F(\alpha))\| \\
&\leq \sup_{C\in\mathcal{C}} \|\mathbf{S}_n(C) - \mathbf{b}_{\boldsymbol{\theta}_0}(C)\| + \|\mathbf{b}_{\boldsymbol{\theta}_0}(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) - \mathbf{b}_{\boldsymbol{\theta}_0}(M_F(\alpha))\| \\
&= o_P(1) + \|\mathbf{b}_{\boldsymbol{\theta}_0}(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) - \mathbf{b}_{\boldsymbol{\theta}_0}(M_F(\alpha))\|.
\end{aligned}$$

The map $\mathcal{C} \ni C \to \mathbf{b}_{\boldsymbol{\theta}_0}(C)$ is continuous with respect to $F(C\Delta\, D)$, because an application of Cauchy-Schwatrz inequality gives $\|\mathbf{b}_{\boldsymbol{\theta}_0}(C) - \mathbf{b}_{\boldsymbol{\theta}_0}(D)\| \leq c\sqrt{F(C\Delta\, D)}$ for some $c > 0$. By again using (uniform) consistency of $\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)$ as estimators of $M_F(\alpha)$ we obtain that also $\sup_\alpha \|\mathbf{b}_{\boldsymbol{\theta}_0}(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)) - \mathbf{b}_{\boldsymbol{\theta}_0}(M_F(\alpha))\| = o_P(1)$.

It remains to show that $r_n = o_P(1)$ (cf. (6.7)). This follows from $\frac{1}{\sqrt{n}}\sum_{i=1}^n(\,|\varepsilon_t| - 1) = O_P(1)$, and $|(F_n - F)(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha))| \leq \sup_{C\in\mathcal{C}}|(F_n - F)(C)| = o_P(1)$. The latter again follows by using a standard bracketing argument together with the assumed ergodicity of X_t.

The same proof holds for $\sqrt{n}\,\widehat{\Delta}_n(\widehat{M}_{\mathcal{C},F_n}(\alpha))$ by replacing $\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha)$ through $\widehat{M}_{\mathcal{C},F_n}(\alpha)$. □

Proof of Theorem 3.2. We have

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^n (\,|\,Y_{t-j}\,| - \widehat{\nu}_y\,)\; I(\,e_{t,\widehat{\boldsymbol{\theta}}} < y\,) \\
&\quad = \frac{1}{\sqrt{n}}\sum_{t=1}^n (\,|\,Y_{t-j}\,| - \nu_y\,)\;\Big[I\Big(\,\varepsilon_t < x\,\frac{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}\,\Big) - P(\,e_{t,\widehat{\boldsymbol{\theta}}} < y\,)\Big] \\
&\qquad + \sqrt{n}\,(\,\widehat{\nu}_y - \nu_y)\,\frac{1}{n}\sum_{t=1}^n \Big[\,I(\,e_{t,\widehat{\boldsymbol{\theta}}} < y\,) - P(\,e_{t,\widehat{\boldsymbol{\theta}}} < y\,)\Big],
\end{aligned}$$

and for the main term on the right-hand side we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\,|\,Y_{t-j}\,| - \nu_y\,) \Big[I\Big(\varepsilon_t < y\, \frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)} \Big) - P(\, e_{t,\widehat{\boldsymbol{\theta}}} < y\,) \Big]$$

$$= \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\,|\,Y_{t-j}\,| - \nu_y\,) \Big[I\Big(\varepsilon_t < y\, \frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)} \Big) - F_\varepsilon\Big(y\, \frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)} \Big) \Big] \tag{6.10}$$

$$+ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\,|\,Y_{t-j}\,| - \nu_y\,) \Big[F_\varepsilon\Big(y\, \frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)} \Big) - F_\varepsilon(y) \Big] \tag{6.11}$$

$$+ (\, F_\varepsilon(y) - P(\, e_{t,\widehat{\boldsymbol{\theta}}} < y\,)\,) \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\,|\,Y_{t-j}\,| - \nu_y\,)\,. \tag{6.12}$$

First notice that (6.12) is $o_P(1)$ uniformly in y. This can be seen as follows. We have

$$|\, F_\varepsilon(y) - P(\, e_{t,\widehat{\boldsymbol{\theta}}} < y\,)\,| = \left| F_\varepsilon(y) - \mathrm{E}\, F_\varepsilon\Big(y\, \frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)} \Big) \right|$$

$$\le \mathrm{E} \left| F_\varepsilon(y) - F_\varepsilon\Big(y\, \frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)} \Big) \right|. \tag{6.13}$$

Our assumptions imply that for each fixed y we have $y\, \frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)} \to y$ in probability, and hence an appropriate version of the dominated convergence theorem implies that (6.13) converges to zero for each fixed y. Uniformity of this convergence follows by using a monotonicity argument. Since by assumption $\widehat{\nu}_y - \nu_y = O_P(\,1/\sqrt{n}\,)$ the assertion follows.

Next we show that the term in (6.11) converges (uniformly in y) to $y\, f(y)\, (\,\sqrt{n}\,(\,\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\,)\,)'\, A$. To see this apply the mean value theorem to see that the term in (6.11) equals

$$y\, f(y)\, \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\,|\,Y_{t-j}\,| - \nu_y\,) \Big(\frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)} - 1 \Big) \tag{6.14}$$

$$+ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\,|\,Y_{t-j}\,| - \nu_y\,)\, (\, f(\xi_t) - f(y)\,)\, y \Big(\frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)} - 1 \Big) \tag{6.15}$$

with ξ_t between $y\, \frac{\sigma_0(\mathbf{X}_t, \widehat{\boldsymbol{\theta}})}{\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)}$ and y. The same arguments as used to analyze (6.7) in the proof of

Theorem 3.1 show that we can write (6.14) as

$$y\,f(y)\;\big(\sqrt{n}\,(\widehat{\boldsymbol\theta}-\boldsymbol\theta_0)\big)'\,\frac{1}{n}\sum_{t=1}^{n}(\,|\,Y_{t-j}\,|-\nu_y\,)\;\frac{\dot\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol\theta})}{\sigma_0(\mathbf{X}_t,\boldsymbol\theta_0)} \tag{6.16}$$

$$+\;\;y\,f(y)\;\big(\sqrt{n}\,(\widehat{\boldsymbol\theta}-\boldsymbol\theta_0)\big)'\,\frac{1}{n}\sum_{t=1}^{n}(\,|\,Y_{t-j}\,|-\nu_y\,)\;b(X_t)\quad\cdot\quad o_P(1) \tag{6.17}$$

By the ergodic theorem the sum in (6.16) converges to its (finite) expectation, and hence (6.17) is $o_P(1)$. Using the ergodic theorem again we obtain the difference of the term in (6.16) and the second term in the expansion (3.7) converges to zero. It remains to show that the term in (6.15) is $o_P(1)$. Using similar arguments as above this term can be written as

$$\big(\sqrt{n}\,(\widehat{\boldsymbol\theta}-\boldsymbol\theta_0)\big)'\,\frac{1}{n}\sum_{t=1}^{n}(\,f(\xi_t)-f(y)\,)\,y\;(\,|\,Y_{t-j}\,|-\nu_y\,)\;\frac{\dot\sigma_0(\mathbf{X}_t,\boldsymbol\theta_0)}{\sigma_0(\mathbf{X}_t,\boldsymbol\theta_0)} \tag{6.18}$$

$$+\;\;y\,f(y)\;\big(\sqrt{n}\,(\widehat{\boldsymbol\theta}-\boldsymbol\theta_0)\big)'\,\frac{1}{n}\sum_{t=1}^{n}(\,f(\xi_t)-f(y)\,)\,y\;(\,|\,Y_{t-j}\,|-\nu_y\,)\;b(X_t)\quad\cdot\quad o_P(1). \tag{6.19}$$

We now show that both of these terms are $o_P(1)$. First, using the fact that $\frac{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol\theta})}{\sigma_0(\mathbf{X}_t,\boldsymbol\theta)}-1=o_P(1)$ we can argue as in Horváth et al. (2001) to show that $\sup_{1\le t\le n}\sup_y|\,(\,f(\xi_t)-f(y)\,)\,y\,|=o_P(1)$. (See proof of (4.5) in Horváth et al.) As above it follows that (6.19) is $o_P(1)$. It remains to show that $\frac{1}{n}\sum_{t=1}^{n}\Big|\;(\,|\,Y_{t-j}\,|-\nu_y\,)\;\frac{\dot\sigma_0(\mathbf{X}_t,\boldsymbol\theta_0)}{\sigma_0(\mathbf{X}_t,\boldsymbol\theta_0)}\;\Big|=O_P(1)$. This, however, follows from an application of the ergodic theorem since $\mathrm{E}\big|\frac{\dot\sigma_0(\mathbf{X}_t,\boldsymbol\theta_0)}{\sigma_0(\mathbf{X}_t,\boldsymbol\theta_0)}\big|\,|\,(|Y_t|-\nu_y)\,|<\infty$ (by Cauchy-Schwartz's inequality). We have shown that (6.15) is $o_P(1)$.

Finally we turn to (6.10). The idea here is the following. Let

$$Z_n(y,\boldsymbol\tau)=\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\,|\,Y_{t-j}\,|-\nu_y\,)\Big[\,I\Big(\,\varepsilon_t<y\,\frac{\sigma_0(\mathbf{X}_t,\boldsymbol\tau)}{\sigma_0(\mathbf{X}_t,\boldsymbol\theta_0)}\Big)-F_\varepsilon\Big(\,y\,\frac{\sigma_0(\mathbf{X}_t,\boldsymbol\tau)}{\sigma_0(\mathbf{X}_t,\boldsymbol\theta_0)}\,\Big)\Big]. \tag{6.20}$$

Notice that the expression in (6.10) equals $Z_n(y,\widehat{\boldsymbol\theta})$. We will show that for each $C>0$ we have

$$\sup_y\;\sup_{\boldsymbol\tau\in T_{n,C}}\;|\,Z_n(y,\boldsymbol\tau)-Z_n(y,\boldsymbol\theta_0)\,|=o_P(1). \tag{6.21}$$

Since $\widehat{\boldsymbol\theta}$ is a $\sqrt{n}$-consistent estimator this implies that (6.10) as a process has the same limit as the process $Z_n(y, \boldsymbol{\theta}_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n (\,|\,Y_{t-j}\,| - \nu_y\,)\,(\,I(\,\varepsilon_t\, <\, y\,) - F_\varepsilon(y)\,)$, which has the asserted limit distribution (as we will see below).

In order to show (6.21) we consider $\{\, Z_n(y, \boldsymbol{\tau}),\, y \in \mathbf{R},\, \boldsymbol{\tau} \in T_{n,C}\,\}$ and construct a sequence of nested partitions $\mathcal{P}_k = \{\, \mathcal{P}_{1,k}, \ldots, \mathcal{P}_{N_k,k}\,\}$ of the index space $\mathbf{R} \times T_{n,C}$ such that as $n, k \to \infty$

$$\sup_y \sup_{\boldsymbol{\tau}\in T_{n,C}} |\, Z_n(y,\boldsymbol{\tau}) - Z_n(y,\boldsymbol{\theta}_0)\,| \quad \leq \max_{1\leq j\leq N(2^{-k})} \sup_{(y_1,\boldsymbol{\tau}_1),\,(y_2,\boldsymbol{\tau}_2)\in\mathcal{P}_{j,k}} |\, Z_n(y_1,\boldsymbol{\tau}_1) - Z_n(y_2,\boldsymbol{\tau}_2)\,| = o_P(1), \tag{6.22}$$

which of course implies (6.21). We now construct the partition and show the two properties asserted in (6.22). To this end we use results of Nishiyama (2000) (Theorem 4.2, Corollary 4.3; see also Nishiyama 1996).

Write $Z_n(y, \boldsymbol{\tau}) = \sum_{t=1}^n \xi_t(y, \boldsymbol{\tau})$ and observe that the random variables $\xi_t(y, \boldsymbol{\tau}) = \frac{1}{\sqrt{n}}(\,|\,Y_{t-j}\,| - \nu_y\,)\left[\, I\Big(\,\varepsilon_t < y\, \frac{\sigma_0(\mathbf{X}_t,\boldsymbol{\tau})}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}\,\Big) - F_\varepsilon\Big(\,y\, \frac{\sigma_0(\mathbf{X}_t,\boldsymbol{\tau})}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}\,\Big)\,\right]$ form a sequence of martingale differences with respect to the filtration $\mathcal{F}_t = \sigma(\varepsilon_t,\, \varepsilon_{t-1} \ldots)$. We now construct the sequence of partitions mentioned above. Fix $C, \eta > 0$, and with $r(\mathbf{X}_t, \boldsymbol{\tau}) = \sigma_0(\mathbf{X}_t, \boldsymbol{\tau})/\sigma_0(\mathbf{X}_t, \boldsymbol{\theta}_0)$ let $A_n = \{\, m < \sup_{1\leq t\leq n} \sup_{\boldsymbol{\tau}\in T_{n,C}} r(\mathbf{X}_t, \boldsymbol{\tau}) < M\,\}$ where $m, M > 0$ are chosen such that $P(\,A_n) > 1 - \eta$ for n large enough. Further recall that f is Lipschitz continuous. Let $c > 0$ denote the Lipschitz constant. Then we find the partition $\mathcal{P}_k$ as follows. Let $y_1 = F_\varepsilon^{-1}(2^{-k})/M$ and define $y_j = y_1 + j\, 2^{-k}/M\, c$. Let N be the largest integer such that $y_N < F_\varepsilon^{-1}(1 - 2^{-k})/m$, and let $y_{N+1} = \frac{1}{m} F_\varepsilon^{-1}(1 - 2^{-k})$. Then $y_0 = 0 < F_\varepsilon^{-1}(2^{-k})/M = y_1 < y_2 < \cdots < y_N < y_{N+1} = F_\varepsilon^{-1}(1 - 2^{-k})/m < y_{N+2} = \infty$. It follows that $\mathcal{P}_{j,k} = (y_j,\, y_{j+1}] \times T_{n,C}$ defines a partition of the index space $\mathbf{R} \times T_{n,C}$ such that on the set A_n we have $F_\varepsilon(y_{j+1}\, r(\mathbf{X}_t, \boldsymbol{\tau})) - F_\varepsilon(y_j\, r(\mathbf{X}_t, \boldsymbol{\tau})) < c\,(\,y_{j+1} - y_j\,)\, r(\mathbf{X}_t, \boldsymbol{\tau}) < 2^{-k}$ for $j = 1, \ldots, N$ and $F_\varepsilon(y_1\, r(\mathbf{X}_t, \boldsymbol{\tau})) - F_\varepsilon(y_0\, r(\mathbf{X}_t, \boldsymbol{\tau})) < F_\varepsilon(\,F_\varepsilon^{-1}(2^{-k})\, r(\mathbf{X}_t, \boldsymbol{\tau})/M) < 2^{-k}$, and similarly we have $F_\varepsilon(y_{N+2}\, r(\mathbf{X}_t, \boldsymbol{\tau})) - F_\varepsilon(y_{N+1}\, r(\mathbf{X}_t, \boldsymbol{\tau})) < 2^{-k}$. Since $\mathrm{E}\varepsilon_t < \infty$ we have $\max(\,F_\varepsilon^{-1}(2^{-k}),\, F_\varepsilon^{-1}(1 - 2^{-k})\,) = O(1/2^{-k})$. Hence, $\log N = \log N(2^{-k}) \leq C(\,\log(1/2^{-k})\,)^2 =: H(2^{-k})$ for some constant $C > 0$, and we obviously have $\int_0^1 \sqrt{H(\delta)}\, d\delta < \infty$. It remains to verify conditions $[L2']$ and $[PE']$ of Nishiyama. With $\bar{\xi}_t = \sup_{y,\boldsymbol{\tau}} |\xi_t(y, \boldsymbol{\tau})| \leq \frac{1}{\sqrt{n}}\,|\,|Y_{t-j}| - 1\,|$ we have for every $\eta > 0$

$$\sum_{t=1}^n \bar{\xi}_t^2\, I(\,\bar{\xi}_t > \eta\,) \leq \frac{1}{n} \sum_{t=1}^n (|Y_{t-j}| - 1)^2\, I(\,|\,|Y_{t-j}| - 1\,| > \sqrt{n}\,\eta\,) = o_P(1).$$

and the latter holds, since the Y_t have finite second moments. This verifies Nishiyama's condition $[L2']$. In order to check $[PE']$ let $\mathcal{P}_{1,k}, \ldots, \mathcal{P}_{N+1,k}$ denote the partition constructed above at level 2^{-k} as constructed above. Nishiyama's condition $[PE']$ requires that

$$\sup_k \max_{1 \le j \le N(2^{-k})} \frac{\sqrt{\sum_{t=1}^n E_{t-1}(\mathcal{P}_{j,k})}}{2^{-k}} = O_P(1), \tag{6.23}$$

where $E_{t-1}(\mathcal{P}_{j,k}) = E\big(\,[\,\sup_{(y_1,\boldsymbol{\tau}_1),(y_2,\boldsymbol{\tau}_2)\in\mathcal{P}_{j,k}}(\,\xi_t(y_1,\boldsymbol{\tau}_1) - \xi_t(y_2,\boldsymbol{\tau}_2)\,)^2\,]\,|\mathcal{F}_{t-1}\big) \le 2/n\,\big(\,(\,|Y_{t-j}| - 1\,)^2\,(\,F_\varepsilon(y_{k+1}\,r(\mathbf{X}_t,\boldsymbol{\tau})) - F_\varepsilon(y_k\,r(\mathbf{X}_t,\boldsymbol{\tau}))\,)^2\,\big)$. By construction of the partitions we have on A_n that $E_{t-1}(\mathcal{P}_{j,k}) \le 2/n\,(\,|Y_{t-j}| - 1\,)^2\,2^{-2k}$, and hence (6.23) holds since $\frac{1}{n}\sum_{t=1}^n(\,|Y_{t-j}| - 1\,)^2 = O_P(1)$. By Nishiyama's result we now have that for all $\varepsilon, \delta > 0$ there exists a k such that

$$P\Big(\max_{1\le j\le N(2^{-k})}\ \sup_{(y_1,\boldsymbol{\tau}_1),\,(y_2,\boldsymbol{\tau}_2)\in\mathcal{P}_{j,k}} |\,Z_n(y_1,\boldsymbol{\tau}_1) - Z_n(y_2,\boldsymbol{\tau}_2)\,| > \delta\Big) < \varepsilon. \tag{6.24}$$

This is the second property asserted in(6.22). The first property in (6.22), i.e. the inequality, follows directly from the construction of the partition. This proves (6.21). Asymptotic normality of the finite dimensional distributions follows from the martingale convergence theorem, and calculation of the asserted covariance structure is a simple exercise.

The asserted convergence in distribution of T_2 is an easy consequence of the just verified first result of the theorem, and follows essentially by observing that if $F = G$ we have $\sigma_0(\cdot) = \nu_y$, and by applying a continuous mapping theorem. □

Proposition 6.1 *Suppose that* **(A1)** *and* **(A2)** *hold, and that* $N_B(\epsilon, \mathcal{C}, L_2(F)) < \infty$ *for all* $\epsilon > 0$.

(i.a) If $\mathcal{C}$ *is such that* $\mathbf{(C2)}_{F_n}$ *holds, and that* $M_F(\alpha) \in \mathcal{C}$ *for all* $0 \le \alpha \le 1$*, then*

$$\sup_{\alpha\in[0,1]} F(\widehat{M}_{\mathcal{C},F_n}(\alpha)\Delta M_F(\alpha)) = o_P(1), \qquad .$$

(i.b) The statement of (i.a) also holds with F *and* F_n *replaced by* G *and* G_n*, respectively.*

(ii) Suppose that $\mathcal{C}$ *is such that* $\mathbf{(C2)}_{G_{n,\widehat{\boldsymbol{\theta}}}}$ *holds, and that* $M_F(\alpha) \in \mathcal{C}$ *for all* $0 \le \alpha \le 1$*. In addition*

assume **(V2.a)** *and suppose that* $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1)$*, then under hypothesis (3.1)*

$$\sup_{\alpha \in [0,1]} F(\widehat{M}_{\mathcal{C},\widehat{\boldsymbol{\theta}}}(\alpha) \Delta M_F(\alpha)) = o_P(1).$$

PROOF. We first prove the assertion for $\widehat{M}_{\mathcal{C},F_n}(\alpha)$. Since by assumption (2.9) holds and $M_F(\alpha)$ is in $\mathcal{C}$, we know that for each $\alpha \in [0,1]$ there exists a unique $\lambda_\alpha \geq 0$ with $M_F(\alpha) = \{x : f(x) \geq \lambda_\alpha\}$. The proof of Proposition 3.5 in Polonik (1995) shows that if we let $H_\lambda(C) = F(C) - \lambda \,\mathrm{Leb}(C)$, then the assertion follows once we have shown that $\sup_\alpha |\, H_{\lambda_\alpha}(M_F(\alpha)) - H_{\lambda_\alpha}(\widehat{M}_{\mathcal{C},F_n}(\alpha))\,| = o_P(1)$, and this is what we show now. For each α let $\widehat{\alpha} = \min\{\, 0 \leq \eta \leq 1 : F_n(M_f(\eta)) \geq \alpha \,\}$ and let $\widehat{\beta}_\alpha = F_n(M_f(\widehat{\alpha}))$. Notice that by definition we have $\widehat{\alpha} = F(M_F(\widehat{\alpha}))$, and hence, since $\sup_\alpha |\,(F_n - F)(M_F(\alpha))\,| = o_P(1)$ we obtain $\sup_\alpha |\,\widehat{\beta}_\alpha - \widehat{\alpha}\,| = o_P(1)$. The fact that $\sup_\alpha |\,(F_n - F)(M_F(\alpha))\,| = o_P(1)$ follows from $\sup_\alpha |\,(F_n - F)(M_F(\alpha))\,| \leq \sup_{C \in \mathcal{C}} |(F_n - F)(C)| = o_P(1)$, and the latter follows from the assumed ergodicity of the underlying process and the finiteness of the bracketing number by using the standard bracketing trick (see e.g. Pollard, 1984, or van der Vaart and Wellner 1995). Further we have

$$\begin{aligned} 0 &\leq H_{\lambda_\alpha}(M_F(\alpha)) - H_{\lambda_\alpha}(\widehat{M}_{\mathcal{C},F_n}(\alpha)) \\ &= [\, H_{\lambda_\alpha}(M_F(\alpha)) - H_{\lambda_\alpha}(M_F(\widehat{\alpha}))\,] + [\, H_{\lambda_\alpha}(M_F(\widehat{\alpha})) - H_{\lambda_\alpha}(\widehat{M}_{\mathcal{C},F_n}(\alpha))\,] \\ &=: I + II. \end{aligned}$$

Now, by definition $\widehat{\beta}_\alpha = F_n(M_F(\widehat{\alpha})) \geq \alpha$, and hence $\mathrm{Leb}(\widehat{M}_{\mathcal{C},F_n}(\alpha)) \leq \mathrm{Leb}(M_F(\widehat{\alpha}))$. It follows that $II = F(M_F(\widehat{\alpha})) - F(\widehat{M}_{\mathcal{C},F_n}(\alpha)) - \lambda_\alpha\,(\,\mathrm{Leb}(M_F(\widehat{\alpha})) - \mathrm{Leb}(\widehat{M}_{\mathcal{C},F_n}(\alpha))\,) \leq F(M_F(\widehat{\alpha})) - F(\widehat{M}_{\mathcal{C},F_n}(\alpha))$
$= \widehat{\alpha} - F_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) + (F_n - F)(\widehat{M}_{\mathcal{C},F_n}(\alpha))$.

Consequently,

$$\begin{aligned} 0 \leq I + II &\leq I + (\widehat{\alpha} - \alpha) + (F_n(\widehat{M}_{\mathcal{C},F_n}(\alpha)) - \alpha) + \sup_{C \in \mathcal{C}} |(F_n - F)(C)| \\ &= I + (\widehat{\alpha} - \alpha) + o_P(1). \end{aligned}$$

In order to see that $\sup_\alpha |\widehat{\alpha} - \alpha| = o_P(1)$ we show that $\sup_\alpha |\widehat{\beta}_\alpha - \alpha| = o_P(1)$. (Recall that we already have seen above that $\sup_\alpha |\widehat{\beta}_\alpha - \widehat{\alpha}| = o_P(1)$.) Notice that for every $0 \leq \alpha \leq 1$ and $\epsilon > 0$

we have $o_P(1) = \sup_\alpha (F_n - F)(M_F(\alpha)| \geq |F_n(M_F(\min\{\alpha+\epsilon,1\})) - F(M_F(\min\{\alpha+\epsilon,1\}))| = |F_n(M_F(\min\{\alpha+\epsilon,1\})) - \min\{\alpha+\epsilon,1\}|$. This means that for every fixed $\epsilon > 0$ we have with probability tending to one that $F_n(M_F(\min\{\alpha+\epsilon,1\})) \geq \alpha$ uniformly in α. It follows that with probability tending to one we have (uniformly in α) that $\widehat{\beta}_\alpha \leq \min(\alpha+\epsilon,1)$. Since $\epsilon > 0$ was arbitrary and we also have $\alpha \leq \widehat{\beta}_\alpha$ (see above) we can conclude that $\sup_\alpha |\widehat{\beta}_\alpha - \alpha| = o_P(1)$.

It remains to show that $I = o_P(1)$ uniformly in α. Observe that by assumption the maps $\alpha \to \lambda_\alpha$, $\alpha \to \text{Leb}(M_F(\alpha))$ and $\alpha \to H_{\lambda_\alpha}(M_F(\alpha))$ are continuous. In particular we have $H_{\lambda_\alpha}(M_F(\alpha)) \to 1$ as $\alpha \to 1$. We have seen above that $\sup_\alpha |\widehat{\alpha} - \alpha| = o_P(1)$. Hence, for each $\eta > 0$ there exists an $\epsilon > 0$ such that for $\alpha > 1-\epsilon$ we have $1 \geq H_{\lambda_\alpha}(M_F(\alpha)) > 1-\eta$ as well as $1 \geq H_{\lambda_\alpha}(M_F(\widehat{\alpha})) > 1-2\eta$ with probability tending to one as $T \to \infty$. Hence for each η there exists an $\epsilon > 0$ such that $P[\sup_{\alpha > 1-\epsilon} |H_{\lambda_\alpha}(M_F(\alpha)) - H_{\lambda_\alpha}(M_F(\widehat{\alpha}))| > 2\eta] \to 0$. Further, for a given $\epsilon > 0$ there exist a $K > 0$ such that $\sup_{\alpha \leq 1-\epsilon} \text{Leb}(M_F(\alpha)) < K < \infty$. The function $\alpha \to \text{Leb}(M_F(\alpha))$ is (uniformly) continuous on $[0, 1-\epsilon]$. It follows that $I = |H_{\lambda_\alpha}(M_F(\alpha)) - H_{\lambda_\alpha}(M_F(\widehat{\alpha}))| \leq |F(M_F(\alpha)) - F(M_F(\widehat{\alpha}))| + \lambda_\alpha |\text{Leb}(M_F(\alpha)) - \text{Leb}(M_F(\widehat{\alpha}))| = |\widehat{\beta}_\alpha - \alpha| + \lambda_\alpha |\text{Leb}(M_F(\alpha)) - \text{Leb}(M_F(\widehat{\alpha}))| = o_P(1)$, because $0 \leq \lambda_\alpha \leq \sup_x f(x) < \infty$, uniformly in α. This concludes the proof of $(i.a)$. The proof for $(i.b)$ is mutatis mutandis the same.

In order to see (ii), the key point is to show that (under the present assumptions) we have

$$\sup_{C \in \mathcal{C}} |(G_{n,\widehat{\boldsymbol{\theta}}} - F)(C)| = o_P(1) \quad \text{as } n \to \infty. \tag{6.25}$$

Then, the above proof can be repeated mutatis mutandis by observing that under (3.1) we have $G_{\boldsymbol{\theta}_0} = F$. Property (6.25) can be seen as follows. We first argue that

$$\sup_{C \in \mathcal{C}} |(G_{n,\boldsymbol{\theta}_0} - F)(C)| = o_P(1).$$

Observe that under (3.1) we have $G_{n,\boldsymbol{\theta}_0}(C) = \frac{1}{n}\sum_{t=1}^n |\varepsilon_t|\, I(\mathbf{X}_t \in C)$. Since ε_t is independent from $\mathbf{X}_t$ we have $\text{E}(G_{n,\boldsymbol{\theta}_0}(C)) = F(C)$. Ergodic theorem implies that $G_{n,\boldsymbol{\theta}_0}(M_F(\alpha)) \to F(M_F(\alpha))$ for all α, and the usual bracketing trick implies that the convergence is uniform in α. Property (6.25) follows once we have shown that

$$\sup_{C\in\mathcal{C}} |(G_{n,\boldsymbol{\theta}_0} - G_{n,\widehat{\boldsymbol{\theta}}})(C)| = o_P(1).$$

We have

$$\begin{aligned}(G_{n,\boldsymbol{\theta}_0} - G_{n,\widehat{\boldsymbol{\theta}}})(C) &= \frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}\, n}\sum_{t=1}^{n}\Big(\frac{|Y_t|}{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)} - \frac{|Y_t|}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})}\Big) I(\mathbf{X}_t\in C) + \Big(\frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}} - \frac{1}{\widehat{\nu}_{\boldsymbol{\theta}_0}}\Big)\frac{1}{n}\sum_{t=1}^{n}|\varepsilon_t| \\ &=: \quad I_C \quad + \quad II\end{aligned}$$

As for I_C we have from **(V2.b)** that on the set $B_n = \{\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) < c\}$

$$|\, I_C\,| \le \frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}\, n}\sum_{t=1}^{n}\Big|\, 1 - \frac{\sigma_0(\mathbf{X}_t,\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_t,\widehat{\boldsymbol{\theta}})}\,\Big|\, |\varepsilon_t|\, I(\mathbf{X}_t\in C) = o_P(1)\,\frac{1}{\widehat{\nu}_{\widehat{\boldsymbol{\theta}}}\, n}\sum_{t=1}^{n}|\varepsilon_t|.$$

Again using **(V2.b)** it is easy to see that $\widehat{\nu}_{\widehat{\boldsymbol{\theta}}} - \widehat{\nu}_{\boldsymbol{\theta}_0} = o_P(1)$ and $\widehat{\nu}_{\boldsymbol{\theta}_0} = \frac{1}{n}\sum_{t=1}^{n}|\varepsilon_t| \to 1$ in probability. Since by assumption $\sqrt{n}(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_P(1)$ we can choose c so that $P(B_n)$ becomes arbitrarily small. Hence we can conclude that $\sup_{C\in\mathcal{C}} I_C = o_P(1)$. The fact that also $II = o_P(1)$ follows by using similar arguments. □

Proof of Theorem 3.3. First, a close inspection of the proof of Theorem 3.1 shows that under the present assumptions expansion (3.6) holds with the $o_P(1)$-term being uniform in the 'true' parameter $\boldsymbol{\theta}_0' \in B_{c/\sqrt{n}}(\boldsymbol{\theta}_0)$ for any $c > 0$. Since for $c \to \infty$ we have $P_{\boldsymbol{\theta}_0}\big(\widehat{\boldsymbol{\theta}} \in B_{c/\sqrt{n}}(\boldsymbol{\theta}_0)\big) \to 0$ as $n \to \infty$, expansion (3.6) also holds with $F_{\boldsymbol{\theta}_0}$-probability arbitrarily close to 1 in the bootstrap world (conditional on $\mathbf{X}_1,\ldots,\mathbf{X}_n$) where the true parameter is $\widehat{\boldsymbol{\theta}}$. For each $\boldsymbol{\theta}$ let $Y_{t,\boldsymbol{\theta}}$ denote realizations from our model under $\boldsymbol{\theta}$ and let $\mathbf{X}_{t,\boldsymbol{\theta}} = (|Y_{t-1,\boldsymbol{\theta}}|,\ldots,|Y_{t-p,\boldsymbol{\theta}}|)'$. The main term of expansion (3.6) under $\boldsymbol{\theta}$ can then be expressed as

$$\begin{aligned}&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}(\,|\varepsilon_t| - 1\,)\,[\,I(\mathbf{X}_{t,\boldsymbol{\theta}}\in M_{G_{\boldsymbol{\theta}}}(\alpha)) - \alpha\,] \;+\; (\,\sqrt{n}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\,)'\,\mathbf{b}_{\boldsymbol{\theta}}(M_{G_{\boldsymbol{\theta}}}(\alpha)) \\ =&\; W_{n,\boldsymbol{\theta}}(M_{G_{\boldsymbol{\theta}}}(\alpha)) \;+\; (\,\sqrt{n}\,(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta})\,)'\,\mathbf{b}_{\boldsymbol{\theta}}(M_{G_{\boldsymbol{\theta}}}(\alpha)) \\ =:&\; (I)_{\boldsymbol{\theta},\alpha} \;+\; (II)_{\boldsymbol{\theta},\alpha}, \qquad (6.26)\end{aligned}$$

where $W_{n,\boldsymbol{\theta}}(C)$ denotes the process defined in (3.4) using $\mathbf{X}_{t,\boldsymbol{\theta}}$. We will show that for any sequence

$\{\boldsymbol{\theta}_n\} \subset \mathcal{U}_{c/\sqrt{n}}(\boldsymbol{\theta}_0)$ there exist realizations of $Y_{t,\boldsymbol{\theta}_n}, t = 1, \ldots, n$ and $Y_{t,\boldsymbol{\theta}_0}, t = 1, \ldots, n$ such that we have both

$$\sup_{\alpha \in [0,1]} \big| (I)_{\boldsymbol{\theta}_n,\alpha} - (I)_{\boldsymbol{\theta}_0,\alpha} \big| = o_P(1) \qquad \text{and} \qquad \sup_{\alpha \in [0,1]} \big| (II)_{\boldsymbol{\theta}_n,\alpha} - (II)_{\boldsymbol{\theta}_0,\alpha} \big| = o_P(1).$$

This then implies the result. First we consider term (I). We have to show that

$$\sup_{\alpha \in [0,1]} \Big| W_{n,\boldsymbol{\theta}_n}(M_{G_{\boldsymbol{\theta}_n}}(\alpha)) - W_{n,\boldsymbol{\theta}_0}(M_{G_{\boldsymbol{\theta}_0}}(\alpha)) \Big| = o_P(1). \tag{6.27}$$

Since by assumption $W_{n,\boldsymbol{\theta}_n}$ and $W_{n,\boldsymbol{\theta}_0}$ are asymptotically equicontinuous, it follows that the process $Z_n(\alpha) := W_{n,\boldsymbol{\theta}_n}(M_{G_{\boldsymbol{\theta}_n}}(\alpha)) - W_{n,\boldsymbol{\theta}_0}(M_{G_{\boldsymbol{\theta}_0}}(\alpha)),\ \alpha \in [0,1]$ is asymptotically equicontinuous. It order to see (6.27) it thus suffice to show that the finite dimensional distributions of $Z_n(\alpha)$ converge to zero. To see that, observe that $Z_n(\alpha)$ is a sum of mean zero, uncorrelated random variables. Thus we have to show that $\mathrm{Var}(Z_n(\alpha)) \to 0$ as $n \to \infty$ for each $\alpha \in [0,1]$. We have

$$\begin{aligned}\mathrm{Var}(Z_n(\alpha) &= \mathrm{Var}\big[\,(\,|\varepsilon_t| - 1\,)\,[\,I(\mathbf{X}_{t,\boldsymbol{\theta}_n} \in M_{G_{\boldsymbol{\theta}_n}}(\alpha)) - I(\mathbf{X}_{t,\boldsymbol{\theta}_0} \in M_{G_{\boldsymbol{\theta}_0}}(\alpha))\,]\,\big]\\ &= \mathrm{Var}(|\varepsilon_t|)\,\mathrm{Var}\big[I(\mathbf{X}_{t,\boldsymbol{\theta}_n} \in M_{G_{\boldsymbol{\theta}_n}}(\alpha)) - I(\mathbf{X}_{t,\boldsymbol{\theta}_0} \in M_{G_{\boldsymbol{\theta}_0}}(\alpha))\big].\end{aligned}$$

and hence it is sufficient to show that

$$\lim_{n \to \infty} P\big(\mathbf{X}_{t,\boldsymbol{\theta}_n} \in M_{G_{\boldsymbol{\theta}_n}}(\alpha),\, \mathbf{X}_{t,\boldsymbol{\theta}_0} \in M_{G_{\boldsymbol{\theta}_0}}(\alpha)\,\big) = 0 \qquad \text{for all } \alpha \in [0,1], \tag{6.28}$$

where $Y_{t,\boldsymbol{\theta}_n}, t = 1, \ldots, n$ and $Y_{t,\boldsymbol{\theta}_0}, t = 1, \ldots, n$ are realizations with $\sup_{1 \le t \le n} \big|\,|Y_{t,\boldsymbol{\theta}_n}| - |Y_{t,\boldsymbol{\theta}_0}|\,\big| = o_P(1)$. Such realizations exist by assumption. In order to see (6.28) we will use our assumption that $g_{\boldsymbol{\theta}}$ are uniformly Lipschitz. Let L denote the corresponding Lipschitz constant. Since $M_{G_{\boldsymbol{\theta}}}(\alpha) = \{x : g_\theta(x) \ge \lambda_{\boldsymbol{\theta},\alpha}\}$ for some $\lambda_{\boldsymbol{\theta},\alpha} > 0$, we have for any $\mathbf{x} \in \mathcal{R}^d$ and $\eta > 0$ that

$$\big\{\,\mathbf{x} \in M_{G_{\boldsymbol{\theta}}}(\alpha),\, \mathbf{x} + \eta \in M_{G_{\boldsymbol{\theta}}}(\alpha)\,\big\} \subset \big\{\lambda_{\boldsymbol{\theta},\alpha} - L\,\eta \le g_{\boldsymbol{\theta}}(\mathbf{x}) \le \lambda_{\boldsymbol{\theta},\alpha} + L\,\eta\big\} = \big\{\,|g_{\boldsymbol{\theta}}(\mathbf{x}) - \lambda_{\boldsymbol{\theta},\alpha}| \le L\,\eta\,\big\}.$$

This implies the following. Let $\eta > 0$ be fixed, and let $A_n(\eta)$ denote the set where the $o_P(1)$-term is less than or equal to $\eta > 0$. Then we of course have $P(A_n^c(\eta)) = o(1)$ as $n \to \infty$ and we obtain

$$
\begin{aligned}
& P\big\{\mathbf{X}_{t,\boldsymbol{\theta}_0} \in M_{G_{\boldsymbol{\theta}_n}}(\alpha),\, \mathbf{X}_{t,\boldsymbol{\theta}_n} \in M_{G_{\boldsymbol{\theta}_n}}(\alpha)\big\} \\
& \qquad = P\big\{\mathbf{X}_{t,\boldsymbol{\theta}_0} \in M_{G_{\boldsymbol{\theta}_n}}(\alpha),\, \mathbf{X}_{t,\boldsymbol{\theta}_0} + o_P(1) \in M_{G_{\boldsymbol{\theta}_n}}(\alpha)\big\} \\
& \qquad \leq P\big\{(\mathbf{X}_{t,\boldsymbol{\theta}_0} \in M_{G_{\boldsymbol{\theta}_n}}(\alpha),\, \mathbf{X}_{t,\boldsymbol{\theta}_0} + o_P(1) \in M_{G_{\boldsymbol{\theta}_n}}(\alpha),\, A_n(\eta)\big\} + P(A_n^c(\eta)) \\
& \qquad \leq G_{\boldsymbol{\theta}_0}\big\{|g_{\boldsymbol{\theta}_n}(\mathbf{x}) - \lambda_{\boldsymbol{\theta}_n,\alpha}| \leq L\,\eta\big\} + o(1) \\
& \qquad = o(1) + o(1) \qquad \text{as } n \to \infty. \qquad (6.29)
\end{aligned}
$$

The last equality uses the fact that

$$
\begin{aligned}
& G_{\boldsymbol{\theta}_0}\big\{|g_{\boldsymbol{\theta}_n}(\mathbf{x}) - \lambda_{\boldsymbol{\theta}_n,\alpha}| \leq L\,\eta\big\} \\
\leq\ & G_{\boldsymbol{\theta}_0}\big\{|g_{\boldsymbol{\theta}_n}(\mathbf{x}) - \lambda_{\boldsymbol{\theta}_n,\alpha}| \leq L\,\eta,\ |g_{\boldsymbol{\theta}_n}(\mathbf{x}) - g_{\boldsymbol{\theta}_0}(\mathbf{x})| \leq L\eta\big\} + G_{\boldsymbol{\theta}_0}\big\{|g_{\boldsymbol{\theta}_n}(\mathbf{x}) - g_{\boldsymbol{\theta}_0}(\mathbf{x})| \leq L\eta\big\} \\
\leq\ & G_{\boldsymbol{\theta}_0}\big\{|g_{\boldsymbol{\theta}_0}(\mathbf{x}) - \lambda_{\boldsymbol{\theta}_n,\alpha}| \leq 2\,L\,\eta\big\} + G_{\boldsymbol{\theta}_0}\big\{|g_{\boldsymbol{\theta}_n}(\mathbf{x}) - g_{\boldsymbol{\theta}_0}(\mathbf{x})| \leq L\eta\big\} \\
\leq\ & \sup_{\lambda>0} G_{\boldsymbol{\theta}_0}\big\{|g_{\boldsymbol{\theta}_0}(\mathbf{x}) - \lambda| \leq 2\,L\,\eta\big\} + G_{\boldsymbol{\theta}_0}\big\{|g_{\boldsymbol{\theta}_n}(\mathbf{x}) - g_{\boldsymbol{\theta}_0}(\mathbf{x})| \leq L\eta\big\} = o(1).
\end{aligned}
$$

The first term in the last line is $o(1)$ because of **(A2)**, and the fact that second term is $o(1)$ follows from an application of the dominated convergence theorem. This completes the proof of (6.27). Now we consider term (II). To simplify notation we present the proof for the case $q = 1$. Using representation (3.10) we write

$$
(II)_{\boldsymbol{\theta}_n,\alpha} - (II)_{\boldsymbol{\theta}_0,\alpha} = \big[b_{\boldsymbol{\theta}_n}(M_{G_{\boldsymbol{\theta}_n}}(\alpha)) - b_{\boldsymbol{\theta}_0}(M_{G_{\boldsymbol{\theta}_0}}(\alpha))\big]\frac{1}{\sqrt{n}}\sum_{t=1}^{n} h(\varepsilon_t)\psi_{\boldsymbol{\theta}_n}(\mathbf{X}_{t,\boldsymbol{\theta}_n}) \qquad (6.30)
$$

$$
+\, b_{\boldsymbol{\theta}_0}(M_{G_{\boldsymbol{\theta}_0}}(\alpha))\,\frac{1}{\sqrt{n}}\sum_{t=1}^{n} h(\varepsilon_t)\big(\psi_{\boldsymbol{\theta}_n}(\mathbf{X}_{t,\boldsymbol{\theta}_n}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_0})\big) + o_P(1). \qquad (6.31)
$$

First we consider the sum in (6.31). Write $\psi_{\boldsymbol{\theta}_n}(\mathbf{X}_{t,\boldsymbol{\theta}_n}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_0}) = \big[\psi_{\boldsymbol{\theta}_n}(\mathbf{X}_{t,\boldsymbol{\theta}_n}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_n})\big] - \big[\psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_n}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_0})\big]$. For the latter we have to consider differences of the form $\psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_0} + o_P(1)) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_0})$. Since the assumed Lipschitz continuity of $\boldsymbol{\psi}_{\boldsymbol{\theta}_0}$ implies that $\sup_{\mathbf{x}}|\psi_{\boldsymbol{\theta}_0}(\mathbf{x}+\eta) - \psi_{\boldsymbol{\theta}_0}(\mathbf{x}+\eta)| \leq L\,\eta$ for some $L > 0$, we obtain that $|\psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_n}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_0})| = o_P(1)$. An application of the dominated convergence theorem gives

$$\mathrm{Var}\Big[\frac{1}{\sqrt{n}}\sum_{t=1}^{n} h(\varepsilon_t)\big(\psi_{\boldsymbol{\theta}_n}(\mathbf{X}_{t,\boldsymbol{\theta}_0}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_0})\big)\Big] = \sigma_h^2 \,\mathrm{Var}[\psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_n}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_0})] = o(1). \tag{6.32}$$

As for the term involving differences of the form $\psi_{\boldsymbol{\theta}_n}(\mathbf{X}_{t,\boldsymbol{\theta}_n}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_n})$ we have

$$\begin{aligned} &\frac{1}{\sqrt{n}}\sum_{t=1}^{n} h(\varepsilon_t)\big[\psi_{\boldsymbol{\theta}_n}(\mathbf{X}_{t,\boldsymbol{\theta}_n}) - \psi_{\boldsymbol{\theta}_0}(\mathbf{X}_{t,\boldsymbol{\theta}_n})\big] \\ \leq\;& \sqrt{n}|\boldsymbol{\theta}_n - \boldsymbol{\theta}_0|\,\frac{1}{n}\sum_{t=1}^{n}\big|h(\,\varepsilon_t) - \mathrm{E}(h(\varepsilon_t)\,\big|\,K(\mathbf{X}_{t,\boldsymbol{\theta}_0}) = \; o(1)\cdot O_P(1) = o_P(1). \end{aligned} \tag{6.33}$$

Combining (6.32) and (6.33) shows that the sum in (6.31) is $o_P(1)$. It remains to consider the term on the right hand side of (6.30). We have already shown above that the sum in (6.30) behaves like the sum with $\mathbf{X}_{t,\boldsymbol{\theta}_n}$ replaced by $\mathbf{X}_{t,\boldsymbol{\theta}_0}$, and thus this sum is $O_P(1)$. The proof is completed once we have shown that

$$\sup_{\alpha\in[0,1]}\big\|\mathbf{b}_{\boldsymbol{\theta}_n}(M_{G_{\boldsymbol{\theta}_n}}(\alpha)) - \mathbf{b}_{\boldsymbol{\theta}_0}(M_{G_{\boldsymbol{\theta}_0}}(\alpha))\big\| = o(1) \quad \text{as } n\to\infty.$$

We have

$$\begin{aligned} &\mathbf{b}_{\boldsymbol{\theta}_n}(M_{G_{\boldsymbol{\theta}_n}}(\alpha)) - \mathbf{b}_{\boldsymbol{\theta}_0}(M_{G_{\boldsymbol{\theta}_0}}(\alpha)) \\ &= \mathrm{E}\big[\frac{\dot\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_n},\boldsymbol{\theta}_n)}{\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_n},\boldsymbol{\theta}_n)}(I(\mathbf{X}_{t,\boldsymbol{\theta}_n}\in M_{G_{\boldsymbol{\theta}_n}}(\alpha)-\alpha) - \frac{\dot\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_0},\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_0},\boldsymbol{\theta}_0)}(I(\mathbf{X}_{t,\boldsymbol{\theta}_0}\in M_{G_{\boldsymbol{\theta}_0}}(\alpha)-\alpha)\big] \\ &= \mathrm{E}\big[\frac{\dot\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_n},\boldsymbol{\theta}_n)}{\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_n},\boldsymbol{\theta}_n)}(I(\mathbf{X}_{t,\boldsymbol{\theta}_n}\in M_{G_{\boldsymbol{\theta}_n}}(\alpha) - I(\mathbf{X}_{t,\boldsymbol{\theta}_0}\in M_{G_{\boldsymbol{\theta}_0}}(\alpha))\big] \qquad (6.34) \\ &\qquad + \mathrm{E}\big[\big(\frac{\dot\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_n},\boldsymbol{\theta}_n)}{\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_n},\boldsymbol{\theta}_n)} - \frac{\dot\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_0},\boldsymbol{\theta}_0)}{\sigma_0(\mathbf{X}_{t,\boldsymbol{\theta}_0},\boldsymbol{\theta}_0)}\big)(I(\mathbf{X}_{t,\boldsymbol{\theta}_0}\in M_{G_{\boldsymbol{\theta}_0}}(\alpha)-\alpha)\big] \qquad (6.35) \end{aligned}$$

An application of Cauchy-Schwarz inequality together with (6.29) shows that (6.34) is $o_P(1)$ uniformly in α. Our assumptions also assure that we can apply dominated convergence theorem to see that also (6.35) is $o_P(1)$ uniformly in α. This completes the proof.

References

Billingsley, P. (1999). *Convergence of Probability Measures* (2nd edit). Wiley, New York.

Chen, M. and An, H.Z. (1997). A Kolmogorov-Smirnov type test for conditional heteroskedasticity in time series. *Statistics and Probability Letters*, **33**, 321-331.

Doukhan, P. Massart, P., and Rio, E. (1994): A functional central limit theorem for strongly mixing processes. *Ann. Inst. Henri Poincaré, Probab. et Statist.* **30**, 63-82.

Doukhan, P. Massart, P., and Rio, E. (1995): Invariance principles for absolute regular empirical processes. *Ann. Inst. Henri Poincaré, Probab. et Statist.* **31**, 393-427.

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

Härdle, W. and Tsybakov, A. (1997). Local polynomial estimators of the volatility function in nonparametric autoregression. *J. of Econometrics*, **81**, 223-242.

Hall, P. and Wilson, S.R. (1991). Two guide lines of bootstrap hypothesis testing. *Biometrics*, **47**, 757-762.

Hinkley, D. V. (1988). Bootstrap methods (with discussion). *Journal of the Royal Statistical Society*, **B**, **50**, 321-337.

Khmaladze, E.V. (1981). Martingale approach in the theory of goodness-of-fit tests. *Theory. Probab. Appl.*, **26**, 240-257.

Khmaladze, E.V. (1988). An innovation approach to goodness of fit in $\mathbf{R}^m$. *The Annals of Statistics*, **16**, 1503-1516.

Koul, H. L., Stute, W. (1999). Nonparametric model checks for time series. *The Annals of Statistics*, **27**, 204-236.

Laïb, N. (2003). Nonparametric tests for conditional variance functions in time series. *Australian & New Zealand Journal of Statistics*, **45**, 461-475.

Morgan, J.P. (1996). *RiskMetrics Technical Document*. Fourth edition, New York.

Müller, A. and Stoyan, D. (2002). Comparison methods for stochastic models and risks. Wiley, New York.

Nishiyama, Y. (1996): A central limit theorem for ℓ^∞-valued martingale difference arrays and its application. Preprint 971, Dept. Mathematics, Utrecht Univ.

Nishiymam, Y. (2000): Weak convergence of some classes of martingales with jumps. *The Annals of Probability*, **28**, 685-712.

Pollard, D. (1984). *Convergence of stochastic processes.* Springer, New York.

Polonik, W. (1995). Measuring mass concentration and estimating density contour clusters - an excess mass approach. *The Annals of Statistics*, **23**, 855-881.

Polonik, W. (1999). Concentration and goodness-of-fit in higher dimensions: (Asymptotically) distribution-free methods. *The Annals of Statistics*, **27**, 1210-1229.

Stute, W. (1997). Nonparametric model checks for regression. *The Annals of Statistics*, **25**, 613-641.

van der Vaart, A. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer, New York.