



# Adaptiveness of the empirical distribution of residuals in semi- parametric conditional location scale models

Christian Francq, Jean-Michel Zakoïan

## ► To cite this version:

Christian Francq, Jean-Michel Zakoïan. Adaptiveness of the empirical distribution of residuals in semi- parametric conditional location scale models. 2020. hal-02898909

**HAL Id: hal-02898909**

**<https://hal.science/hal-02898909>**

Preprint submitted on 14 Jul 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Adaptiveness of the empirical distribution of residuals in semi-parametric conditional location scale models

Christian Francq and Jean-Michel Zakoïan

*CREST and University of Lille, 5 Avenue Henri Le Chatelier, 91120 Palaiseau, France.*

E-mail: christian.francq@univ-lille3.fr, zakoian@ensae.fr

**Summary.** This paper addresses the problem of deriving the asymptotic distribution of the empirical distribution function  $\hat{F}_n$  of the residuals in a general class of time series models, including conditional mean and conditional heteroscedasticity, whose independent and identically distributed errors have unknown distribution  $F$ . We show that, for a large class of time series models (including the standard ARMA-GARCH), the asymptotic distribution of  $\sqrt{n}\{\hat{F}_n(\cdot) - F(\cdot)\}$  is impacted by the estimation but does not depend on the model parameters. It is thus neither asymptotically estimation free, as is the case for purely linear models, nor asymptotically model dependent, as is the case for some nonlinear models. The asymptotic stochastic equicontinuity is also established. We consider an application to the estimation of the conditional Value-at-Risk.

## 1. Introduction

For independent variables  $\eta_1, \dots, \eta_n$  with common distribution  $F$ , the celebrated Glivenko-Cantelli theorem states the uniform almost sure convergence of the empirical distribution,

$$\sup_{x \in \mathbb{R}} |F_n(x) - F(x)| \rightarrow 0, \quad a.s.$$

where  $F_n(x) = \frac{1}{n} \sum_{t=1}^n 1_{\{\eta_t \leq x\}}$  for all  $x$ . The usual Central Limit Theorem (CLT) shows that  $\sqrt{n}\{F_n(x) - F(x)\}$  is asymptotically  $\mathcal{N}(0, F(x)(1 - F(x)))$  distributed without any further assumption. Moreover, under the assumption that  $F$  is continuous, the sequence  $\{\sqrt{n}[F_n(\cdot) - F(\cdot)]\}$  is stochastically equicontinuous, in the sense that  $\sqrt{n}\{F_n(x_n) - F(x_n)\} \xrightarrow{\mathcal{L}} \mathcal{N}(0, F(x)(1 - F(x)))$  for any sequence  $(x_n)$  converging to  $x$  in probability, for any  $x$  (see Billingsley (1968, Section 22) and Andrews (1994)). The latter convergence in distribution has important statistical applications, for instance the derivation of the asymptotic distribution of the empirical quantiles (see below).

Time series observations are rarely independent and identically distributed (iid), but time series models often involve iid innovations. Modern time series analysis often relies on semi-parametric models in which the distribution  $F$  of the innovations  $\eta_t$  is not fully specified. Many such models are of the location-scale form  $y_t = m_t(\boldsymbol{\theta}_0) + \sigma_t(\boldsymbol{\theta}_0)\eta_t$ , where  $m_t$  and  $\sigma_t$  are functions of the past observations  $y_{t-i}, i > 0$ , and the sequence  $(\eta_t)$  is iid. Consistent estimation of the parameter  $\boldsymbol{\theta}_0$  can be achieved by several methods. In particular, the Quasi-Maximum Likelihood (QML) estimation method generally relies on a Gaussian criterion - written as if the distribution  $F$  was Gaussian - but provides consistent estimators while remaining agnostic concerning the true distribution.

Even if no distributional assumption is made during the estimation phase, using semi-parametric time series models may require information about the errors distribution  $F$ . For instance, in the location-scale model the *conditional* Value-at-Risk (VaR) at level  $\alpha \in (0, 1)$  is equal to  $m_t + \sigma_t \xi_\alpha$  where  $\xi_\alpha$  is the  $\alpha$ -quantile of  $F$ . Information about  $F$  can be obtained via the empirical distribution function of the residuals (e.d.f.r.)  $\hat{F}_n$ . Consistent estimation of the parameters produce residuals which are obviously not iid but converge to the innovations. One question of importance is whether the previous properties on  $F_n$  apply to  $\hat{F}_n$ , i.e. when innovations are replaced by residuals.

For many applications, the asymptotic distribution of the sequence  $\{\sqrt{n}[\hat{F}_n(\cdot) - F(\cdot)]\}$  is required. In the statistical literature, two kinds of results have been established. In some situations, the asymptotic distribution is asymptotically estimation free (AEF), in the sense that, asymptotically,  $\sqrt{n}\{\hat{F}_n(\cdot) - F(\cdot)\} \stackrel{d}{=} \sqrt{n}\{F_n(\cdot) - F(\cdot)\}$ . Such results go back to Boldin (1982), for the  $\text{AR}(p)$  process, and have been generalized to linear processes by Kreiss (1991) and Bai (1994). On the other hand, existing results for nonlinear processes show that the distribution of  $\sqrt{n}\{\hat{F}_n(\cdot) - F(\cdot)\}$  can be much more complicated. It can be asymptotically model dependent (AMD), in the sense

that the asymptotic variance depends on both the law of the innovations  $F$  and the parameter  $\theta_0$ . For different classes of conditionally heteroscedastic time series, the asymptotic distribution was investigated by Boldin (1998), and Lee and Taniguchi (2005) among others, and for the squared residuals by Horváth, Kokoszka and Teyssi re (2001) and Berkes and Horv th (2003). In all these works, the distribution of the statistics involving the residuals is AMD. To overcome the AMD, Koul and Ling (2006) proposed a modification of the Kolmogorov-Smirnov test based on residuals. See Berkes and Horv th (2002) for a review of the asymptotic behaviour of e.d.f.r.

The main contribution of this paper is to show that, for a large class of time series models (including the standard ARMA-GARCH), the asymptotic distribution of  $\sqrt{n}\{\widehat{F}_n(\cdot) - F(\cdot)\}$  is neither AEF nor AMD. In some cases, which we are able to characterize, the distribution is asymptotically model free (AMF), however it is impacted by the estimation, i.e. it is asymptotically estimation dependent (AED), though only through the errors distribution. The AMF property simplifies dramatically the use of such asymptotic results in practical applications based on potentially complex time series models. On the other hand, the AED property tells us that usual statistics available for observed iid processes cannot be directly employed.

We start by considering volatility modelling and QML estimation, which constitutes the most widely used framework for financial returns. The absence of a conditional mean and the choice of a specific estimation method allows us to derive finer results. However, from an extended perspective, we will also consider conditional location-scale models in which both first conditional moments are involved, and more general estimation methods. We will also investigate an application to the VaR estimation.

The rest of the paper is organized as follows. In the next section we introduce a general parametric volatility model and the main assumptions ensuring the consistency of the e.d.f.r. Next, we state our first main result, establishing the asymptotic distribution of the e.d.f.r. Section 3 considers the extension to conditional location-scale models. The application to VaR is developed in Section 4. In particular, we provide conditions for the strong consistency and asymptotic normality of the residuals quantiles. Section 5 concludes. Most proofs are postponed to Section 6 or to an Appendix.

## 2. Empirical distribution of the residuals for volatility models

In this section, we focus on conditional scale models of the multiplicative form

$$\epsilon_t = \sigma_t \eta_t, \quad \sigma_t = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0), \quad (1)$$

where  $(\eta_t)$  is a sequence of iid random variables,  $E(\eta_t^2) = 1$ ,  $\theta_0$  is a vector of unknown coefficients which belongs to a compact parameter set  $\Theta \subset \mathbb{R}^d$ , and  $\sigma$  is a positive function. Note that we do

not assume that  $E(\eta_t) = 0$ .

For instance, the standard GARCH( $p, q$ ) model writes

$$\epsilon_t = \sigma_t \eta_t, \quad \sigma_t^2 = \omega_0 + \sum_{i=1}^q \alpha_{0i} \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_{0j} \sigma_{t-j}^2, \quad (2)$$

where  $\boldsymbol{\theta}_0 = (\omega_0, \alpha_{01}, \dots, \alpha_{0p})'$  satisfies  $\omega_0 > 0, \alpha_{0i} \geq 0, \beta_{0j} \geq 0$ . Under strict stationarity, invertibility of the polynomial  $\mathcal{B}(z) = 1 - \sum_{j=1}^p \beta_{0j} z^j$  holds and the volatility  $\sigma_t^2$  can indeed be written as a linear function of the past values of  $\epsilon_t^2$ .

We start by considering the Gaussian QML estimation method for which explicit conditions for Consistency and Asymptotic Normality (CAN) can be displayed. More general estimators will be considered in Section 3. The setup can be described as follows. Given observations  $\epsilon_1, \dots, \epsilon_n$ , and using arbitrary initial values  $\tilde{\epsilon}_i$  for  $i \leq 0$ , we define for any  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ ,

$$\tilde{\sigma}_t(\boldsymbol{\theta}) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \boldsymbol{\theta}),$$

which will be used as a proxy of  $\sigma_t(\boldsymbol{\theta}) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \dots; \boldsymbol{\theta})$ .

A Gaussian QML estimator (QMLE) of  $\boldsymbol{\theta}_0$  is defined as

$$\hat{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t(\boldsymbol{\theta}), \quad \tilde{\ell}_t(\boldsymbol{\theta}) = \frac{\epsilon_t^2}{\tilde{\sigma}_t^2(\boldsymbol{\theta})} + \log \tilde{\sigma}_t^2(\boldsymbol{\theta}). \quad (3)$$

Let  $K > 0$  be a generic constant or random variable measurable with respect to  $\mathcal{F}_0$ , where  $\mathcal{F}_t$  denotes the  $\sigma$ -algebra generated by  $\{\eta_s, s \leq t\}$ . Let  $\rho \in (0, 1)$ . We shall assume the following.

**A1:**  $(\epsilon_t)$  is a strictly stationary, non-anticipative (i.e.  $\epsilon_t \in \mathcal{F}_t$ ) and ergodic solution of Model (1).

Moreover,  $E|\sigma_t|^r < \infty$  for some  $r > 0$ .

**A2:** For any real sequence  $(x_i)$ , the function  $\boldsymbol{\theta} \mapsto \sigma(x_1, x_2, \dots; \boldsymbol{\theta})$  is continuous differentiable.

Almost surely,  $\sigma_t(\boldsymbol{\theta}) \in (\underline{\omega}, \infty]$  for any  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$  and for some  $\underline{\omega} > 0$ . Moreover,  $\sigma_t(\boldsymbol{\theta}_0)/\sigma_t(\boldsymbol{\theta}) = 1$  a.s. iff  $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ .

**A3:**  $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} |\sigma_t(\boldsymbol{\theta}) - \tilde{\sigma}_t(\boldsymbol{\theta})| \leq K \rho^t$ .

**A4:** There exists a neighborhood  $V(\boldsymbol{\theta}_0)$  of  $\boldsymbol{\theta}_0$  such that  $E \left( \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})} \right)^r < \infty$  and  $E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\mathbf{D}_t(\boldsymbol{\theta})\|^r < \infty$ , where  $\mathbf{D}_t(\boldsymbol{\theta}) = \sigma_t^{-1}(\boldsymbol{\theta}) \partial \sigma_t(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ .

Assumptions **A1-A3** are a set of conditions ensuring the strong consistency of  $\hat{\boldsymbol{\theta}}_n$  (see Francq and Zakoian (2004, 2015)). Assumption **A4** is introduced to control the difference between the innovations and the residuals.<sup>1</sup>

<sup>1</sup>For the classical GARCH( $p, q$ ) model, under invertibility conditions on the lag polynomial  $\mathcal{B}_{\boldsymbol{\theta}}(z) = 1 - \sum_{j=1}^p \beta_j z^j$  and assuming that the law of  $\eta_t^2$  is nondegenerate, **A1-A4** reduce to the first part of **A1**.

Let the residuals  $\hat{\eta}_t = \epsilon_t / \tilde{\sigma}_t(\hat{\theta}_n)$ . We note that, at least for  $t$  large enough,  $\tilde{\sigma}_t(\hat{\theta}_n) \geq \underline{\omega}$  by **A2-A3**. We start by establishing a Glivenko-Cantelli result for the e.d.f.r.  $\hat{F}_n(x) = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t \leq x\}}$ .<sup>2</sup> The following assumption is simply denoted **A5** when it holds for all  $x \in \mathbb{R}$ .

**A5(x)**: For  $x \in \mathbb{R}$ , the cdf  $F$  of  $\eta$  is Lipschitz continuous in a neighborhood of  $x$ .

**THEOREM 1.** *Under **A1-A4**,*

- (a) *If **A5(x)** holds for  $x \in \mathbb{R}$ , we have  $|\hat{F}_n(x) - F(x)| \rightarrow 0$  a.s.*
- (b) *If **A5** holds we have  $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \rightarrow 0$  a.s.*

Notice that the previous uniform convergence was established by Stute (2001) in the case of ARCH( $q$ ) processes. It is known (see for instance Stute and Schumann (1980)) that in the case of a stationary ergodic sequence, the Glivenko-Cantelli theorem is valid without any assumption on  $F$ . The following example shows that **A5** is required for the consistency of the e.d.f.r.

**EXAMPLE 1 (GLIVENKO-CANTELLI IN FAILURE).** *Consider the scale model  $\epsilon_t = \sigma \eta_t, \sigma > 0$ , where  $\eta_t$  is distributed over  $\{-\sqrt{2}, 0, \sqrt{2}\}$  with  $P(\eta_t = -\sqrt{2}) = P(\eta_t = \sqrt{2}) = 1/4$ . Let  $\hat{\sigma}_n$  be a consistent estimator of  $\sigma$ . The residuals  $\hat{\eta}_t = \frac{\epsilon_t}{\hat{\sigma}_n}$  have the e.d.f.r.*

$$\hat{F}_n(x) = \left( \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\epsilon_t < 0} \right) \mathbb{1}_{-\sqrt{2} \frac{\sigma}{\hat{\sigma}_n} \leq x < 0} + \left( \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\epsilon_t \leq 0} \right) \mathbb{1}_{0 \leq x < \sqrt{2} \frac{\sigma}{\hat{\sigma}_n}} + \mathbb{1}_{x \geq \sqrt{2} \frac{\sigma}{\hat{\sigma}_n}}.$$

*It follows that  $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \geq |\hat{F}_n(-\sqrt{2}) - F(-\sqrt{2})| = 1/4$  whenever  $\sigma < \hat{\sigma}_n$ , which has a non vanishing probability for the QML. Indeed, we have  $\hat{\sigma}_n^2 = \frac{1}{n} \sum_{t=1}^n \epsilon_t^2$ , thus  $P(\sigma < \hat{\sigma}_n) = P(\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\epsilon_t^2=0} > 1/2) = P(X_n > n/2)$  where  $X_n \sim \mathcal{B}(n, 1/2)$ .*

We now derive the e.d.f.r. asymptotic distribution under the following additional assumptions.

**A6:**  $\theta_0$  belongs to the interior of  $\Theta$ .

**A7:** There exist no non-zero  $\mathbf{x} \in \mathbb{R}^d$  such that  $\mathbf{x}' \frac{\partial \sigma_t(\theta_0)}{\partial \theta} = 0$ , a.s.

**A8:** The function  $\theta \mapsto \sigma(x_1, x_2, \dots; \theta)$  has continuous second-order derivatives, and

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \sigma_t(\theta)}{\partial \theta} - \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} \right\| \leq K \rho^t.$$

**A9:** There exists a neighborhood  $V(\theta_0)$  of  $\theta_0$  such that

$$E \sup_{\theta \in V(\theta_0)} \left\{ \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|^4 + \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\|^2 + \left| \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \right|^4 + \left| \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \right|^4 \right\} < \infty.$$

Moreover,  $\kappa_4 := E|\eta_t|^4 < \infty$ .

<sup>2</sup>For the classical GARCH( $p, q$ ) model, Berkes and Horváth (2003) established a Glivenko-Cantelli theorem for the empirical cumulative distribution function (cdf) of the *squared* residuals.

**A10:** All the coordinates of  $\frac{\partial \sigma_t(\theta_0)}{\partial \theta}$  are (strictly) positive.

**A11:**  $\eta_1$  admits a density  $f$  which is continuous on  $\mathbb{R}$ .

**A12:** For any  $\theta \in \Theta$ , for any  $c > 0$ , and any sequence  $(x_i)$ , there exists  $\theta_c \in \Theta$  such that  $c\sigma(x_1, x_2, \dots; \theta) = \sigma(x_1, x_2, \dots; \theta_c)$ .

Assumption **A6** is required for the asymptotic normal distribution of  $\hat{\theta}_n$ . Assumptions **A7-A10** and **A12** are satisfied for the standard GARCH  $(p, q)$  model, under the already mentioned regularity assumptions. Assumption **A12** is a stability-by-scaling property, which seems a desirable assumption for any volatility model (examples are provided in Francq and Zakoian (2015)).

We now state our first main result showing that the law of  $\sqrt{n}(\hat{F}_n - F)$  is AED but AMF.

**THEOREM 2.** *Let **A1-A4**, **A6-A11** hold. Then, for any sequence  $(x_n)$  of random variables converging in probability to  $x \in \mathbb{R}$ ,*

$$\sqrt{n}(\hat{F}_n(x_n) - F(x_n)) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \{\mathbb{1}_{\eta_t < x} - F(x)\} - \frac{xf(x)}{2\sqrt{n}} \Omega' J^{-1} \sum_{t=1}^n (1 - \eta_t^2) D_t + o_P(1),$$

where  $\Omega = E(D_t)$ ,  $J = E(D_t D_t')$  with  $D_t = D_t(\theta_0)$ .

If in addition **A12** holds,

$$\sqrt{n}(\hat{F}_n(x_n) - F(x_n)) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, F(x)\{1 - F(x)\} + \frac{\{xf(x)\}^2}{4}(\kappa_4 - 1) + xf(x)\varrho(x)\right)$$

where  $\varrho(x) = E(\eta_0^2 \mathbb{1}_{\eta_0 < x}) - F(x)$ .

**REMARK 1.** *A noticeable outcome of this theorem is that, under **A12**, the asymptotic distribution of the e.d.f.r. only depends on the underlying distribution of the innovations. It is completely independent of the model from which these residuals are derived. In the sense of Robinson (1987), it means that the e.d.f.r. is adaptive to the unknown value of  $\theta_0$ , and even to the unknown form of the function  $\sigma$ . This does not mean that we retrieve the asymptotic distribution of  $F_n$  derived for iid data. The usual asymptotic variance,  $F(x)\{1 - F(x)\}$ , is only valid when  $xf(x) = 0$ . This is the case when  $x = 0$  because the innovations and residuals at a given date have the same sign. The effect of estimation also vanishes when  $|x|$  increases to  $\infty$  (under **A9**). Parameter estimation may reduce or increase the asymptotic variance (see Appendix).*

**REMARK 2.** *One practical interest of adaptiveness is that estimation of the asymptotic variance reduces to estimation of characteristics of the innovations distribution. In particular,  $f(x)$  can be estimated by a Kernel density estimator based on the residuals. Theorem 2.1 in Kulperger and Yu (2005) shows that, in the standard GARCH case, the use of residuals instead of innovations has no asymptotic impact on the estimation of the density.*

### 3. Including a conditional mean

In ARMA-GARCH models, the conditional variance specification is completed by a linear model for the conditional mean. In what follows, we consider a more general framework. We will not either restrict our investigations to the QML estimator. Suppose the model writes

$$\begin{cases} y_t = m_t + \epsilon_t, & \epsilon_t = \sigma_t \eta_t \\ m_t = m(y_{t-1}, y_{t-2}, \dots; \boldsymbol{\theta}_0), & \sigma_t = \sigma(y_{t-1}, y_{t-2}, \dots; \boldsymbol{\theta}_0) \end{cases} \quad (4)$$

under the same assumptions on  $(\eta_t)$ ,  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\Theta}$  as before. Model (4) includes the double-AR( $p$ ) of Ling (2007) in which  $\boldsymbol{\theta}_0 = (\phi_{01}, \dots, \phi_{0p}, \omega_0, \alpha_{01}, \dots, \alpha_{0p})' \in \mathbb{R}^p \times (0, \infty) \times [0, \infty)^p$  and

$$m(y_{t-1}, y_{t-2}, \dots; \boldsymbol{\theta}_0) = \sum_{i=1}^p \phi_{0i} y_{t-i}, \quad \sigma(y_{t-1}, y_{t-2}, \dots; \boldsymbol{\theta}_0) = \sqrt{\omega_0 + \sum_{i=1}^p \alpha_{0i} y_{t-i}^2},$$

but also more traditional ARMA-GARCH-type models, in which the volatility initially defined in terms of the  $\epsilon_{t-i}$ 's can be rewritten as a function of the  $y_{t-i}$ 's.

Let, for arbitrary initial values  $\tilde{y}_0, \tilde{y}_{-1}, \dots$ , for any  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ , and for  $1 \leq t \leq n$ ,

$$\tilde{m}_t(\boldsymbol{\theta}) = m(y_{t-1}, y_{t-2}, \dots, y_1, \tilde{y}_0, \tilde{y}_{-1}, \dots; \boldsymbol{\theta}), \quad \tilde{\sigma}_t(\boldsymbol{\theta}) = \sigma(y_{t-1}, y_{t-2}, \dots, y_1, \tilde{y}_0, \tilde{y}_{-1}, \dots; \boldsymbol{\theta}).$$

As in the previous section we denote without "tilde", the functions  $m_t(\boldsymbol{\theta})$  and  $\sigma_t(\boldsymbol{\theta})$  in which the initial values are replaced by variables at times anterior to 0. In the following assumptions,  $r > 0$  denotes a real number which can be chosen arbitrarily small.

**B1:**  $(y_t)$  is a strictly stationary, non-anticipative and ergodic solution of Model (4). Moreover,  $E|m_t|^r < \infty$  and  $E|\sigma_t|^r < \infty$ .

**B2:**  $\hat{\boldsymbol{\theta}}_n$  denotes any consistent estimator of  $\boldsymbol{\theta}_0$ . The functions  $\boldsymbol{\theta} \rightarrow m_t(\boldsymbol{\theta})$  and  $\boldsymbol{\theta} \rightarrow \sigma_t(\boldsymbol{\theta})$  are continuously differentiable. Almost surely,  $\sigma_t(\boldsymbol{\theta}) \in (\underline{\omega}, \infty]$  for any  $\boldsymbol{\theta} \in \boldsymbol{\Theta}$  and for some  $\underline{\omega} > 0$ .

**B3:**  $\sup_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} (|m_t(\boldsymbol{\theta}) - \tilde{m}_t(\boldsymbol{\theta})| + |\sigma_t(\boldsymbol{\theta}) - \tilde{\sigma}_t(\boldsymbol{\theta})|) \leq K_t \rho^t$  where  $K_t \in \mathcal{F}_{t-1}$  and  $\sup_t E(K_t^r) < \infty$ .

**B4:** For any neighborhood  $V(\boldsymbol{\theta}_0)$  of  $\boldsymbol{\theta}_0$ , we have

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\{ \left\| \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^r + |m_t(\boldsymbol{\theta}_0) - m_t(\boldsymbol{\theta})|^r + \|D_t(\boldsymbol{\theta})\|^r \right\} < \infty$$

where

$$D_t(\boldsymbol{\theta}) = \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

Note that the condition  $K \in \mathcal{F}_0$  in **A3** has been weakened in **B3**, in particular to be able to handle the ARMA-GARCH under standard conditions.

We start by extending the Glivenko-Cantelli-type Theorem 1. Let the residuals  $\hat{\eta}_t = \{y_t - \tilde{m}_t(\hat{\boldsymbol{\theta}}_n)\} / \tilde{\sigma}_t(\hat{\boldsymbol{\theta}}_n)$ .



THEOREM 3. Under **B1-B4** (instead of **A1-A4**), the conclusions of Theorem 1 hold.

Hence, extending the framework of the previous section does not alter the consistency results. We will now see that the same conclusion cannot be drawn concerning the asymptotic distribution of  $\sqrt{n}(\hat{F}_n - F)$ , which will no longer be model-free in general. Another difference, which constitutes a major difficulty, is that Assumption **A10** can no longer be made when a conditional mean is included.<sup>3</sup> We now assume that  $\hat{\theta}_n$  admits a Bahadur representation.

**B5:** The following expansion holds

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \Delta_{t-1} \mathbf{V}(\eta_t) + o_P(1),$$

where  $\mathbf{V}(\cdot)$  is a measurable function,  $\mathbf{V} : \mathbb{R} \mapsto \mathbb{R}^k$  for some positive integer  $k$ , and  $\Delta_{t-1}$  is a  $\mathcal{F}_{t-1}$ -measurable  $d \times k$  matrix. The variables  $\Delta_t$  and  $\mathbf{V}(\eta_t)$  belong to  $L^2$  with  $E\mathbf{V}(\eta_t) = 0$ ,  $\text{var}\{\mathbf{V}(\eta_t)\} = \Upsilon$  is nonsingular and  $E\Delta_t = \Lambda$  is full row rank.

Under this assumption, the CLT for stationary second-order martingale differences of Billingsley (1961) can be applied. It follows that the asymptotic law of  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  is a centered Gaussian with variance  $\Sigma := E(\Delta_t \Upsilon \Delta_t')$ . Assumption **B5** has to be verified on a case by case basis, given specific model and estimator.

REMARK 3. For the Gaussian QMLE  $\hat{\theta}_n$  of  $\theta_0$ ,

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta} Q_n(\theta), \quad Q_n(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t(\theta), \quad \tilde{\ell}_t(\theta) = \frac{\{y_t - \tilde{m}_t(\theta)\}^2}{\tilde{\sigma}_t^2(\theta)} + \log \tilde{\sigma}_t^2(\theta),$$

it can be shown that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = 2\mathbf{J}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \eta_t \frac{1}{\sigma_t} \frac{\partial m_t(\theta_0)}{\partial \theta} + (\eta_t^2 - 1) \frac{1}{\sigma_t} \frac{\partial \sigma_t(\theta_0)}{\partial \theta} \right\} + o_P(1), \quad (5)$$

where

$$\mathbf{J} = E \left( \frac{\partial \ell_t^2(\theta_0)}{\partial \theta \partial \theta'} \right) = 2\mathbf{J}_m + 4\mathbf{J}_\sigma, \quad \mathbf{J}_m = E \left( \frac{1}{\sigma_t^2} \frac{\partial m_t(\theta_0)}{\partial \theta} \frac{\partial m_t(\theta_0)}{\partial \theta'} \right)$$

and  $\mathbf{J}_\sigma = E D_t(\theta_0) D_t'(\theta_0)$ . The Bahadur expansion in **B5** is thus satisfied, if  $E(\eta_t) = 0$ , with  $k = 2$  and

$$\mathbf{V}(\eta_t) = (\eta_t, \eta_t^2 - 1)', \quad \Delta_{t-1} = 2\mathbf{J}^{-1} \left[ \frac{1}{\sigma_t} \frac{\partial m_t(\theta_0)}{\partial \theta} \quad \frac{1}{\sigma_t} \frac{\partial \sigma_t(\theta_0)}{\partial \theta} \right], \quad \Lambda = 2\mathbf{J}^{-1} [\Omega_m \quad \Omega_\sigma].$$

We also have  $\text{Var}_{as}\{\sqrt{n}(\hat{\theta}_n - \theta_0)\} = 4\mathbf{J}^{-1} \{\mathbf{J}_m + (\kappa_4 - 1)\mathbf{J}_\sigma + \mu_3(\mathbf{J}_{m\sigma} + \mathbf{J}_{m\sigma}')\} \mathbf{J}^{-1} = \Sigma$ , where  $\mathbf{J}_{m\sigma} = E \left( \frac{1}{\sigma_t^2} \frac{\partial m_t(\theta_0)}{\partial \theta} \frac{\partial \sigma_t(\theta_0)}{\partial \theta'} \right)$  and  $\mu_3 = E\eta_t^3$ .

<sup>3</sup>For instance in the model  $y_t = m_0 + \sigma_t \eta_t$  where  $\sigma_t^2 = \omega + \alpha(y_{t-1} - m)^2$ , the derivative of  $\sigma_t$  with respect to  $m$  is not positive, and even not of constant sign.

The assumptions of Section 2 have to be modified as follows.

**B6:** The function  $\boldsymbol{\theta} \rightarrow m_t(\boldsymbol{\theta})$  and  $\boldsymbol{\theta} \rightarrow \sigma_t(\boldsymbol{\theta})$  have continuous second-order derivatives, and

$$\sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{m}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| + \sup_{\boldsymbol{\theta} \in \Theta} \left\| \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} - \frac{\partial \tilde{\sigma}_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \leq K_t \rho^t,$$

where  $K_t$  is as in **B3**.

**B7:** There exists a neighborhood  $V(\boldsymbol{\theta}_0)$  of  $\boldsymbol{\theta}_0$  such that

$$E \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\{ \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|^4 + \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\|^2 \right\} < \infty.$$

Let the additional assumption

**B8:** For  $n$  large enough,  $n > n_0$  say, the conditional distribution of  $\eta_t$  given  $\hat{\boldsymbol{\theta}}_n$  and  $\mathcal{F}_{t-1}$  admits a density which is a.s. bounded, uniformly in  $n > n_0$  and  $t < n - t(n)$  where  $t(n) = o(\sqrt{n})$ .

Intuitively, the variables  $\hat{\boldsymbol{\theta}}_n$  and  $\eta_t$  should be asymptotically independent in most standard situations. For instance, consider the simple location model  $y_t = \theta_0 + \sigma \eta_t$ , with Gaussian innovations  $\eta_t$ . The conditional distribution of  $\eta_t$  given  $\hat{\boldsymbol{\theta}}_n$  and  $\mathcal{F}_{t-1}$ , where  $\hat{\boldsymbol{\theta}}_n$  is the sample mean, is Gaussian with variance  $1 - \{n - (t - 1)\}^{-1}$ . Thus **B8** is satisfied for  $n_0 > 1$  and  $t(n) = 1$ , with bound  $1/\sqrt{\pi}$ .

The next theorem uses the notion of discrete estimator, which has been introduced by Le Cam (1960) and used by many authors (e.g. Kreiss (1987)).

**THEOREM 4.** *Under **B1-B8** and **A9, A11**, if i)  $(x_n)$  is a non-random sequence converging to  $x \in \mathbb{R}$ ; or ii)  $x_n$  is a discrete  $\sqrt{n}$ -consistent estimator of  $x$ , then*

$$\sqrt{n} \left( \hat{F}_n(x_n) - F(x_n) \right) \xrightarrow{\mathcal{L}} \mathcal{N} \left( 0, F(x) \{1 - F(x)\} + \mathbf{H}'(x) \boldsymbol{\Sigma} \mathbf{H}(x) + 2 \mathbf{H}'(x) \boldsymbol{\Lambda} \boldsymbol{\varrho}(x) \right)$$

where  $\boldsymbol{\varrho}(x) = E(\mathbf{1}_{\eta_0 < x} \mathbf{V}(\eta_0))$  and  $\mathbf{H}(x) = f(x)(\boldsymbol{\Omega}_m + x \boldsymbol{\Omega}_\sigma)$ ,  $\boldsymbol{\Omega}_m = E\left(\frac{1}{\sigma_t} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}\right)$ ,  $\boldsymbol{\Omega}_\sigma = E \mathbf{D}_t(\boldsymbol{\theta}_0)$ .

Every sequence of  $\sqrt{n}$ -consistent estimators can be truncated to get a discrete sequence. The discretization can also be avoided, at the price of either reinforcing Assumption **B8**, or increasing the speed of convergence of the sequence  $(x_n)$  (see Appendix).

More explicit forms of the asymptotic variance can be derived for the QMLE under the following extension of Assumption **A12** which will be illustrated below.

**A12\*:** For any  $\boldsymbol{\theta} \in \Theta$ , for any  $c > 0$ , and any sequence  $(x_i)$ , there exists  $\boldsymbol{\theta}_c \in \Theta$  such that

$$c\sigma(x_1, x_2, \dots; \boldsymbol{\theta}) = \sigma(x_1, x_2, \dots; \boldsymbol{\theta}_c) \text{ and } m(x_1, x_2, \dots; \boldsymbol{\theta}) = m(x_1, x_2, \dots; \boldsymbol{\theta}_c).$$

The next result provides an explicit form for the asymptotic variance when the parameters of the conditional mean and variance are independent and are estimated by QML.

COROLLARY 1. Let  $\boldsymbol{\theta} = (\boldsymbol{\varphi}', \boldsymbol{\vartheta}')'$  and suppose that  $m_t = m_t(\boldsymbol{\varphi})$ ,  $\sigma_t = \sigma_t(\boldsymbol{\vartheta})$ . Under the assumptions of Theorem 4, under **A12\*** and using the QMLE of  $\boldsymbol{\theta}_0$ , we have

$$\begin{aligned} & \text{Var}_{as} \left\{ \sqrt{n} \left( \widehat{F}_n(x_n) - F(x_n) \right) \right\} \\ &= F(x) \{1 - F(x)\} + x f(x) \left\{ x f(x) \frac{\kappa_4 - 1}{4} + \varrho(x) \right\} + f(x) \{f(x) + 2E(\mathbb{1}_{\eta_0 < x} \eta_0)\} \\ & \times E \left( \frac{1}{\sigma_t} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}'} \right) \left\{ E \left( \frac{1}{\sigma_t^2} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}'} \right) \right\}^{-1} E \left( \frac{1}{\sigma_t} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} \right). \end{aligned}$$

REMARK 4. The latter variance is obviously AMF if  $E \left( \frac{1}{\sigma_t} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} \right) = \mathbf{0}$ . This occurs in particular for conditionally homoscedastic ARMA models without intercept.<sup>4</sup> However, this property is not general, as the following example shows. Let

$$X_t = \varphi_0 \mathbb{1}_{X_{t-1} > 0} + \sigma_0 \eta_t, \quad \varphi_0 \in \mathbb{R}, \quad \sigma_0 > 0,$$

where, for simplicity, the iid process  $(\eta_t)$  is endowed with a symmetric and continuous distribution  $F$ . Then  $(X_t)$  is strictly stationary<sup>5 6</sup> and straightforward calculation shows that  $P(X_t > 0) = \{3 - 2F(\varphi_0/\sigma_0)\}^{-1}$ . Thus

$$E \left( \frac{1}{\sigma_0} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}'} \right) \left\{ E \left( \frac{1}{\sigma_0^2} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}'} \right) \right\}^{-1} E \left( \frac{1}{\sigma_0} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} \right) = \frac{1}{3 - 2F(\varphi_0/\sigma_0)}.$$

For this model, the distribution of the e.d.f.r. is not AMF.

REMARK 5. When **A12\*** does not hold, even when the conditional mean and variance parameters are disentangled, the variance is AMD, as the following example shows. Consider the following DAR(1) model

$$X_t = \phi_0 X_{t-1} + \sqrt{1 + \alpha_0 X_{t-1}^2} \eta_t, \quad \alpha_0 > 0, \quad \phi_0 \in \mathbb{R}, \quad \boldsymbol{\theta} = (\phi, \alpha)'$$

under the strict stationarity condition  $E \log |\phi_0 + \sqrt{\alpha_0} \eta_t| < 0$  (see Ling, 2007). Then it can be shown that the asymptotic variance of  $\sqrt{n} \left( \widehat{F}_n(x_n) - F(x_n) \right)$  depends on the value of  $\boldsymbol{\theta}_0$  (see Appendix).

<sup>4</sup>Let  $\Phi(L)y_t = \Theta(L)\epsilon_t$  under standard assumptions on the lag polynomials  $\Phi(L) = 1 - \sum_{i=1}^p \phi_{0i} L^i$  and  $\Theta(L) = 1 - \sum_{j=1}^q \theta_{0j} L^j$ . Then  $m_t = \sum_{i=1}^p \phi_i y_{t-i} - \sum_{j=1}^q \theta_j \epsilon_{t-j}$ , from which we deduce  $\frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \phi_i} = y_{t-i} - \sum_{j=1}^q \theta_{0j} \frac{\partial \epsilon_{t-j}}{\partial \phi_i}$ , and  $\frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \theta_k} = - \sum_{j=1}^q \theta_{0j} \frac{\partial \epsilon_{t-j}}{\partial \theta_k} - \epsilon_{t-k}$ . We also have  $\frac{\partial \epsilon_t}{\partial \theta_k} = \Theta(L)^{-1} \epsilon_{t-k}$ , hence  $E \left( \frac{\partial \epsilon_t}{\partial \theta_k} \right) = 0$  and similarly  $E \left( \frac{\partial \epsilon_t}{\partial \phi_i} \right) = 0$ . We conclude that  $E \left( \frac{1}{\sigma_0} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} \right) = 0$ .

<sup>5</sup>This can be shown by noting that  $(X_t)$  is a Markov chain for which the ergodicity criterion of Feigin and Tweedie (1985, Theorem 1) applies. Indeed, for  $V(x) = |x| + 1$  and  $M > |\varphi_0| + \sigma_0 E|\eta_t|$ , we have  $E\{V(X_t)|X_{t-1}\} \leq |\varphi_0| + \sigma_0 E|\eta_t| + 1 \leq (1 - \delta)V(x)$  for  $|x| > M$  and  $\delta$  sufficiently small.

<sup>6</sup>This can also be shown by noting that  $Z_t := \mathbb{1}_{X_t > 0} = a_t Z_{t-1} + b_t(1 - Z_{t-1})$  where  $a_t = \mathbb{1}_{\varphi_0 + \sigma_0 \eta_t > 0}$ ,  $b_t = \mathbb{1}_{\sigma_0 \eta_t > 0}$ . Thus  $Z_t = b_t + \sum_{k=0}^{\infty} c_t \dots c_{t-k} b_{t-k-1}$  a.s. where  $c_t = a_t - b_t$ , where the existence of the infinite sum holds by absolute convergence of the partial sums.

We now consider in details the case of the ARMA( $P, Q$ )-GARCH( $p, q$ ) model estimated by Gaussian QML, which constitutes the model and estimation method most widely used in empirical works. Let

$$\Phi(L)X_t = \Theta(L)\epsilon_t, \quad \epsilon_t = \sigma_t\eta_t, \quad \sigma_t^2 = \omega + \sum_{i=1}^q \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^p \beta_j \sigma_{t-j}^2, \quad (6)$$

with  $\omega > 0, \alpha_i, \beta_j \geq 0$  and under standard assumptions on the lag polynomials  $\Phi(L) = 1 - \sum_{i=1}^P \phi_i L^i$  and  $\Psi(L) = 1 - \sum_{j=1}^Q \psi_j L^j$ . Note that this model does not satisfy the assumptions of Corollary 1 because the conditional variance function depends on the whole set of parameters. However, denoting by  $\varphi$  the ARMA parameters, Assumption **A12\*** is satisfied, with  $\theta_c = (\varphi', c^2\omega, c^2\alpha_1, \dots, c^2\alpha_q, \beta_1, \dots, \beta_p)'$ , which entails a considerable simplification (see Appendix).

**COROLLARY 2.** *For the ARMA-GARCH model (6) with symmetrically distributed innovations, under the assumptions ensuring the CAN of the QMLE, under **B8** and **A11**, and for the sequence  $(x_n)$  of Theorem 4, we have*

$$\begin{aligned} & \sqrt{n} \left( \hat{F}_n(x_n) - F(x_n) \right) \\ \xrightarrow{\mathcal{L}} & \mathcal{N} \left( 0, F(x)\{1 - F(x)\} + \frac{\{xf(x)\}^2}{4}(\kappa_4 - 1) + xf(x)E\{\mathbf{1}_{\eta_0 < x}(\eta_0^2 - 1)\} \right). \end{aligned}$$

In the symmetric ARMA-GARCH case, the asymptotic distribution of the empirical cdf of the residuals is thus adaptive to the parametric specification contrary to the examples of Remarks 4 and 5. In particular, in the pure ARMA case, estimating the (constant) innovations variance suffices to introduce estimation dependence in the asymptotic variance (compare to Bai (1994), Theorem 1).

We now exploit our stochastic equicontinuity results for estimating conditional quantiles.

#### 4. Estimation of the conditional VaR

The *conditional* VaR of the process  $(\epsilon_t)$  at risk level  $\alpha \in (0, 1)$ , denoted by  $\text{VaR}_t(\alpha)$ , is defined as the opposite of the  $\alpha$ -quantile of the conditional distribution of  $\epsilon_t$ . Assuming that this distribution is continuous,  $\text{VaR}_t(\alpha)$  solves

$$P_{t-1}[\epsilon_t < -\text{VaR}_t(\alpha)] = \alpha,$$

where  $P_{t-1}$  denotes the historical distribution conditional on  $\{\epsilon_u, u < t\}$ . When  $(\epsilon_t)$  is a non-anticipative solution of Model (1), the conditional VaR at level  $\alpha$  is then given by

$$\text{VaR}_t(\alpha) = -\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0)\xi_\alpha,$$

where  $\xi_\alpha = \inf\{x : F(x) \geq \alpha\}$  is the  $\alpha$ -quantile of the cdf  $F$  of  $\eta_t$ . A two-step estimator of the conditional VaR at level  $\alpha$  is thus  $\widehat{\text{VaR}}_t(\alpha) = -\tilde{\sigma}(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \hat{\theta}_n)\hat{\xi}_{n,\alpha}$ , where  $\hat{\xi}_{n,\alpha}$  is the

$\alpha$ -quantile of  $\hat{\eta}_1, \dots, \hat{\eta}_n$ , that is the  $[n\alpha]$ -th order statistics of the residuals, where  $[x]$  denotes the smallest integer larger than  $x$ . For standard GARCH, Asymmetric Power GARCH and DAR models, alternative estimators based on the quantile equivariance property and quantile regression have been recently studied by Zheng, Zhu, Li and Xiao (2018), Wang, Zhu, Li and Li (2019), and Zhu and Li (2019). For ARMA-GARCH models, estimators of  $\xi_\alpha$  based on extreme value theory were investigated by Hoga (2019).

We start by establishing the strong consistency of the empirical quantiles of the residuals. We make the following assumption.

**A13:** For  $\alpha \in (0, 1)$ , the cdf  $F$  of  $\eta$  satisfies:  $F(x) > \alpha$  whenever  $x > \xi_\alpha$ .

Assumption **A13** means that the quantile function  $F^-$  of  $\eta$  is right-continuous at  $\alpha$ .

**COROLLARY 3.** *For Model (1) under **A1-A5**, **A13**, for  $\alpha \in (0, 1)$  we have the strong convergence*

$$\hat{\xi}_{n,\alpha} \rightarrow \xi_\alpha \quad a.s.$$

Without Assumption **A13**, we have

$$[\liminf \hat{\xi}_{n,\alpha}, \limsup \hat{\xi}_{n,\alpha}] \subseteq [\xi_\alpha, \xi_\alpha^+] \quad a.s., \quad (7)$$

where  $\xi_\alpha^+ = \inf\{x : F(x) > \alpha\}$ .

When the residuals  $\hat{\eta}_t$  are replaced by innovations  $\eta_t$  (and  $\hat{\xi}_{n,\alpha}$  replaced by  $\xi_{n,\alpha}$ ) the inclusion in (7) is in fact an equality (see Appendix). In particular, the empirical quantile  $\xi_{n,\alpha}$  of the innovations, and in general the empirical quantile  $\hat{\xi}_{n,\alpha}$  of the residuals, do not converge when Assumption **A13** is not satisfied.

To establish the asymptotic distribution of  $\hat{\xi}_{n,\alpha}$ , we need the following assumption.

**A14:** the density  $f$  is strictly positive in a neighborhood of  $\xi_\alpha$ .

**COROLLARY 4.** *For Model (1), under **A1-A4**, **A6-A12** and **A14**, we have*

$$\sqrt{n}(\hat{\xi}_{n,\alpha} - \xi_\alpha) = -\frac{1}{f(\xi_\alpha)} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbf{1}_{\{\eta_t < \xi_\alpha\}} - \alpha) + \xi_\alpha \boldsymbol{\Omega}' \frac{\mathbf{J}^{-1}}{2\sqrt{n}} \sum_{t=1}^n (1 - \eta_t^2) \mathbf{D}_t + o_P(1).$$

Thus

$$\sqrt{n}(\hat{\xi}_{n,\alpha} - \xi_\alpha) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)} + \frac{\xi_\alpha \varrho(\xi_\alpha)}{f(\xi_\alpha)} + \frac{\kappa_4 - 1}{4} \xi_\alpha^2\right).$$

Direct proof of this result would be quite complex, but with the help of Theorem 2 it is simplistic.

**Proof.** Note that Assumptions **A5** and **A13** are satisfied under **A11** and **A14**. Thus, by Corollary 3,  $\widehat{\xi}_{n,\alpha} \rightarrow \xi_\alpha$  *a.s.* The empirical cdf being a step function with jumps of size  $1/n$ , we have  $\widehat{F}_n(\widehat{\xi}_{n,\alpha}) - \alpha \leq 1/n$ . It follows that, by Theorem 2,

$$\begin{aligned} \sqrt{n} \left( \alpha - F(\widehat{\xi}_{n,\alpha}) \right) &= \sqrt{n} \left( \widehat{F}_n(\widehat{\xi}_{n,\alpha}) - F(\widehat{\xi}_{n,\alpha}) \right) + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \{ \mathbb{1}_{\eta_t < \xi_\alpha} - F(\xi_\alpha) \} - \frac{\xi_\alpha f(\xi_\alpha)}{2\sqrt{n}} \boldsymbol{\Omega}' \mathbf{J}^{-1} \sum_{t=1}^n (1 - \eta_t^2) \mathbf{D}_t + o_P(1). \end{aligned}$$

By the delta method applied with the function  $F^{-1}$  we conclude.  $\square$

Again, we emphasize the fact that the distribution of the empirical quantile of the residuals is AMF, though not AEF (since the second and third terms in the asymptotic variance vanish when the residuals are replaced by the innovations). Asymptotic confidence intervals for the conditional VaR could be derived from the joint asymptotic distribution of  $\widehat{\boldsymbol{\theta}}_n$  and  $\widehat{\xi}_{n,\alpha}$  as in Theorem 4 of Francq and Zakoian (2015).

## 5. Concluding remarks

The main contribution of this article is to show that the asymptotic distribution of the residuals of conditional location-scale models can be model free. In the absence of a conditional mean, the effect of estimation on the asymptotic variance is dependent of the error distribution but does not depend on the parametric model. When the conditional mean is present in the model, we give explicit characterizations of situations where this unexpected adaptiveness property holds, leading to simple-to-implement tests on the error distribution.

We now point out several topics for future research. Multivariate extensions are far from trivial because the quantile function does not have a natural extension (see Hallin, Paindaveine and Šíman (2010) for a new multivariate concept of quantile). However, when the innovations have a spherical distribution, estimating the quantile of a linear combination reduces to estimating the cdf of any component of the innovation vector. It would be interesting to extend the results of our paper in this setting (in particular to handle applications involving the VaR of portfolios). Another area of extension concerns nonstationary time series. Ling (1998) showed that the inclusion of a unit root in autoregressive processes renders AMD the distribution of the e.d.f.r., unlike in the stationary case. In the case of an explosive GARCH component, the intercept cannot be consistently estimated (see Francq and Zakoian (2013b)) and the consistency of the cdf of the residuals is an open question.

## 6. Proofs

This section contains the proofs of Theorems 1, 2, 4. More detailed proofs can be found in the Appendix.

### 6.1. Proof of Theorem 1

Let  $\eta_t(\boldsymbol{\theta}) = \epsilon_t/\sigma_t(\boldsymbol{\theta})$  and  $\tilde{\eta}_t(\boldsymbol{\theta}) = \epsilon_t/\tilde{\sigma}_t(\boldsymbol{\theta})$ , so that  $\eta_t = \eta_t(\boldsymbol{\theta}_0)$  and  $\hat{\eta}_t = \tilde{\eta}_t(\hat{\boldsymbol{\theta}}_n)$ . By **A2-A3**,

$$\sup_{\boldsymbol{\theta} \in \Theta} |\eta_t(\boldsymbol{\theta}) - \tilde{\eta}_t(\boldsymbol{\theta})| \leq K \rho^t |\epsilon_t|.$$

Moreover, a Taylor expansion shows that

$$\eta_t(\hat{\boldsymbol{\theta}}_n) = \eta_t - \eta_t(\boldsymbol{\theta}_t) \mathbf{D}'_t(\boldsymbol{\theta}_t) (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \quad (8)$$

where  $\boldsymbol{\theta}_t$  is between  $\hat{\boldsymbol{\theta}}_n$  and  $\boldsymbol{\theta}_0$ . Under **A1-A3**,  $\hat{\boldsymbol{\theta}}_n$  tends to  $\boldsymbol{\theta}_0$  a.s. (see Francq and Zakoian (2013a), Theorem 1). Since  $\hat{\eta}_t = \tilde{\eta}_t(\hat{\boldsymbol{\theta}}_n)$ , it follows that

$$|\hat{\eta}_t - \eta_t| \leq |\eta_t| \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\mathbf{D}_t(\boldsymbol{\theta})\| \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| + K \rho^t |\epsilon_t|$$

for  $n$  large enough. Thus, we have

$$|\hat{\eta}_t - \eta_t| \leq K \left( \rho^t \sigma_t + \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \right) u_t |\eta_t|, \quad (9)$$

for  $n$  large enough, where  $u_t \in \mathcal{F}_{t-1}$  and, by **A1, A2, A4**,  $E(|u_t|^{r/2}) < \infty$ . Without loss of generality, assume that  $K = 1$  in (9). Let  $a_n = \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|$ . For all  $x \in \mathbb{R}$ ,  $\varepsilon > 0$ ,  $S > 0$  and  $M > 0$ , we then have

$$\begin{aligned} |1_{\{\hat{\eta}_t \leq x\}} - 1_{\{\eta_t \leq x\}}| &\leq 1_{\{x - (\rho^t \sigma_t + a_n) u_t |\eta_t| \leq \eta_t \leq x + (\rho^t \sigma_t + a_n) u_t |\eta_t|\}} \\ &\leq 1_{A_{t,\varepsilon,M,S}} + 1_{a_n > \varepsilon} + 1_{u_t |\eta_t| > M} + 1_{\sigma_t > S}, \end{aligned}$$

with the event  $A_{t,\varepsilon,M,S} = \{x - (\rho^t S + \varepsilon) M \leq \eta_t \leq x + (\rho^t S + \varepsilon) M\}$ . For  $t$  large enough such that  $\rho^t S \leq \varepsilon$ , we have  $A_{t,\varepsilon,M,S} \subset A_{t,2\varepsilon,M}$  with  $A_{t,\varepsilon,M} = \{x - \varepsilon M \leq \eta_t \leq x + \varepsilon M\}$ . Assumption **A5(x)** implies, for  $\varepsilon M$  small enough,  $E 1_{A_{t,\varepsilon,M}} = \int_{x-\varepsilon M}^{x+\varepsilon M} dF(y) \leq 2K_0 \varepsilon M$  where  $K_0$  is the Lipschitz constant. For all  $\kappa > 0$ , we thus have a small  $\varepsilon > 0$  and large  $M > 0$  and  $S > 0$  such that  $E \{1_{A_{t,2\varepsilon,M}} + 1_{u_t |\eta_t| > M} + 1_{\sigma_t > S}\} \leq \kappa$ . Recall that, under **A1-A3**,  $a_n$  tends to zero almost surely, and thus  $1_{a_n > \varepsilon} = 0$  for  $n$  large enough. Since  $\kappa$  can be chosen arbitrarily small, it follows that, almost surely, by the ergodic theorem we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n 1_{\{\hat{\eta}_t \leq x\}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n 1_{\{\eta_t \leq x\}} = P(\eta_t \leq x), \quad \forall x \in \mathbb{R}.$$

We have shown that  $|\hat{F}_n(x) - F(x)| \rightarrow 0$  a.s. and the uniform convergence follows from the fact that, for any  $\varepsilon > 0$ , any cdf has a finite number of jumps of size larger than  $\varepsilon$ .  $\square$

### 6.2. A Marcinkiewicz type law of large numbers

The following result, which will be used in the proof of Theorem 2 is of independent interest. Since we have not been able to find this result in the literature, we provide a proof, which is adapted from arguments kindly given to us by L. Horváth.

LEMMA 1. *Let  $(X_t, \mathcal{F}_t^X)$  a strictly stationary martingale difference sequence, where  $\mathcal{F}_t^X$  is a filtration, such that  $EX_t^2 < \infty$ . Then, for any  $s > 0$ ,  $\frac{1}{n^{1/2+s}} \sum_{t=1}^n X_t \rightarrow 0$  a.s.*

**Proof.** Let  $S_n = \sum_{t=1}^n X_t$ . We have, using the martingale difference property,

$$ES_n^2 = nEX_1^2. \quad (10)$$

In view of Theorem 15.1 in Burkholder (1973), we have the Rosenthal's type inequality

$$E \max_{1 \leq k \leq n} S_k^2 \leq Kn. \quad (11)$$

Letting  $\varsigma > 0$ , using (10) and the Markov inequality, we have

$$P\left(|S_{\lfloor k^\varsigma \rfloor}| \geq \varepsilon \lfloor k^\varsigma \rfloor^{s+1/2}\right) \leq \frac{ES_{\lfloor k^\varsigma \rfloor}^2}{\varepsilon^2 \lfloor k^\varsigma \rfloor^{2(s+1/2)}} \leq \frac{K \lfloor k^\varsigma \rfloor}{\varepsilon^2 \lfloor k^\varsigma \rfloor^{2(s+1/2)}} = \frac{K}{\varepsilon^2 \lfloor k^\varsigma \rfloor^{2s}},$$

which is summable provided  $\varsigma > 1/2s$ . It follows, by the Borel-Cantelli lemma, that

$$\frac{S_{\lfloor k^\varsigma \rfloor}}{(\lfloor k^\varsigma \rfloor)^{s+1/2}} \rightarrow 0, \quad a.s. \quad \text{as } k \rightarrow \infty \text{ when } \varsigma > 1/2s.$$

When  $n \in [k^\varsigma, (k+1)^\varsigma]$ , we write  $S_n = S_{\lfloor k^\varsigma \rfloor} + \sum_{i=\lfloor k^\varsigma \rfloor+1}^n X_i$ . Using (11) and noting that the length of the interval  $[k^\varsigma, (k+1)^\varsigma]$  is less than  $\varsigma(k+1)^{\varsigma-1}$  when  $\varsigma > 1$ , we have

$$P\left(\max_{\lfloor k^\varsigma \rfloor \leq j \leq \lfloor (k+1)^\varsigma \rfloor} \left| \sum_{i=\lfloor k^\varsigma \rfloor+1}^j X_i \right| \geq \varepsilon k^{\varsigma(s+1/2)}\right) \leq \frac{K}{\varepsilon^2 k^{2s\varsigma+1}},$$

which is always summable. The result follows by

$$\frac{|S_n|}{n^{s+1/2}} \leq \frac{|S_{\lfloor k^\varsigma \rfloor}|}{(\lfloor k^\varsigma \rfloor)^{s+1/2}} + \frac{\max_{\lfloor k^\varsigma \rfloor \leq j \leq \lfloor (k+1)^\varsigma \rfloor} \left| \sum_{i=\lfloor k^\varsigma \rfloor+1}^j X_i \right|}{k^{\varsigma(s+1/2)}}. \quad \square$$

### 6.3. Proof of Theorem 2

Recall that  $F_n(x) = \frac{1}{n} \sum_{t=1}^n \mathbf{1}_{\{\eta_t \leq x\}}$  and let

$$\begin{aligned} \widehat{e}_n(x) &= \sqrt{n}\{\widehat{F}_n(x) - F(x)\}, & e_n(x) &= \sqrt{n}\{F_n(x) - F(x)\}, & \tilde{\chi}_{t,n} &= \tilde{\sigma}_t(\widehat{\boldsymbol{\theta}}_n)/\sigma_t(\boldsymbol{\theta}_0), \\ \chi_{t,n} &= \sigma_t(\widehat{\boldsymbol{\theta}}_n)/\sigma_t(\boldsymbol{\theta}_0), & h_n(x) &= xf(x) \left\{ \frac{1}{n} \sum_{t=1}^n \mathbf{D}'_t(\boldsymbol{\theta}_0) \right\} \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0). \end{aligned}$$



We have  $\widehat{e}_n(x)$

$$= \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x \chi_{t,n}} - F(x \chi_{t,n})}_{\widehat{e}_{n,1}(x)} + \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n F(x \chi_{t,n}) - F(x)}_{\widehat{e}_{n,2}(x)} + \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x \tilde{\chi}_{t,n}} - \mathbb{1}_{\eta_t \leq x \chi_{t,n}}}_{\widehat{e}_{n,3}(x)}.$$

Let, for  $\mathbf{a}$  a vector of the same size as  $\boldsymbol{\theta}$  (sufficiently small so that  $\boldsymbol{\theta}_0 + \mathbf{a}/\sqrt{n} \in \boldsymbol{\Theta}$ ),

$$e_{n,1}(x, \mathbf{a}) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \{ \mathbb{1}_{\eta_t \leq x \gamma_{t,n}(\mathbf{a})} - F(x \gamma_{t,n}(\mathbf{a})) \}, \quad \gamma_{t,n}(\mathbf{a}) = \frac{\sigma_t(\boldsymbol{\theta}_0 + \frac{\mathbf{a}}{\sqrt{n}})}{\sigma_t(\boldsymbol{\theta}_0)}.$$

Write

$$e_{n,1}(x, \mathbf{a}) - e_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}), \quad (12)$$

where

$$z_{t,n}(x, \mathbf{a}) = \mathbb{1}_{\eta_t \leq x \gamma_{t,n}(\mathbf{a})} - F(x \gamma_{t,n}(\mathbf{a})) - \{ \mathbb{1}_{\eta_t \leq x} - F(x) \}.$$

We will establish a number of auxiliary lemmas.

LEMMA 2. *For any  $u > 0$  and sufficiently large  $n$ ,*

$$P \left( \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) \right| > u \right) \leq \frac{K}{nu^4} (x^2 \|\mathbf{a}\|^2 + 1).$$

**Proof.** By the Markov inequality

$$P \left( \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) \right| > u \right) \leq \frac{1}{n^2} \frac{1}{u^4} E \left( \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) \right)^4.$$

We note that, for fixed  $n, x$  and  $\mathbf{a}$ ,  $(z_{t,n}(x, \mathbf{a}), \mathcal{F}_t)_{1 \leq t \leq n}$  is a martingale difference sequence. Hence, by Rosenthal's inequality (see for instance Hall and Heyde (1980), Theorem 2.11)

$$E \left( \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) \right)^4 \leq K \left\{ E \left( \sum_{t=1}^n E(z_{t,n}^2(x, \mathbf{a}) | \mathcal{F}_{t-1}) \right)^2 + \sum_{t=1}^n E z_{t,n}^4(x, \mathbf{a}) \right\}.$$

Because  $|z_{t,n}(x, \mathbf{a})| \leq 2$ , it suffices to show that

$$E \left( \sum_{t=1}^n E(z_{t,n}^2(x, \mathbf{a}) | \mathcal{F}_{t-1}) \right)^2 \leq n K x^2 \|\mathbf{a}\|^2. \quad (13)$$

Noting that the second-order conditional moment of  $z_{t,n}(x, \mathbf{a})$  is the variance of a Bernoulli distribution, we have, using **A11**,

$$\sum_{t=1}^n E[z_{t,n}^2(x, \mathbf{a}) | \mathcal{F}_{t-1}] \leq \sum_{t=1}^n |F(x \gamma_{t,n}(\mathbf{a})) - F(x)| \leq \frac{K|x|}{\sqrt{n}} \sum_{t=1}^n \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_t^*)}{\partial \boldsymbol{\theta}} \right\| \|\mathbf{a}\|,$$

where  $\boldsymbol{\theta}_t^*$  is between  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\theta}_0 + \mathbf{a}/\sqrt{n}$ . It follows that

$$E \left\{ \sum_{t=1}^n E[z_{t,n}^2(x, \mathbf{a}) | \mathcal{F}_{t-1}] \right\}^2 \leq \frac{Kx^2}{n} \sum_{s,t=1}^n E \left\| \frac{1}{\sigma_s(\boldsymbol{\theta}_0)} \frac{\partial \sigma_s(\boldsymbol{\theta}_s^*)}{\partial \boldsymbol{\theta}} \right\| \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_t^*)}{\partial \boldsymbol{\theta}} \right\| \|\mathbf{a}\|^2,$$

and thus (13) holds.  $\square$

LEMMA 3. Let  $\mathcal{K}$  a compact subset of  $\mathbb{R}$ . We have  $\sup_{x \in \mathcal{K}} |n^{-1/2} \sum_{t=1}^n z_{t,n}(x, \mathbf{a})| = o_P(1)$ .

**Proof.** Fix  $\varepsilon > 0$  and let  $\mathcal{K} \subset \left[-\frac{N\varepsilon}{\sqrt{n}}, \frac{N\varepsilon}{\sqrt{n}}\right]$  with  $N = O(\sqrt{n})$ . Define  $x_j = \frac{j\varepsilon}{\sqrt{n}}$  for  $j = -N, -N+1, \dots, N-1, N$ . It follows that, by Lemma 2, for any  $u > 0$ , there exists  $K = K(u, \mathbf{a}, \varepsilon)$  such that

$$P \left( \max_{-N \leq j \leq N} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x_j, \mathbf{a}) \right| > u \right) \leq \frac{K}{\sqrt{n}}. \quad (14)$$

Noting that

$$\begin{aligned} \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) \right| &\leq \max_{-N \leq j \leq N-1} \sup_{x \in [x_j, x_{j+1}]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) \right| \\ &\leq \max_{-N \leq j \leq N-1} \left\{ \sup_{x \in [x_j, x_{j+1}]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) - \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x_j, \mathbf{a}) \right| \right. \\ &\quad \left. + \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x_j, \mathbf{a}) \right| \right\}, \end{aligned}$$

it remains to show that

$$\limsup_{n \rightarrow \infty} P \left\{ \max_{-N \leq j \leq N-1} \sup_{x \in [x_j, x_{j+1}]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) - \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x_j, \mathbf{a}) \right| > u \right\} = 0. \quad (15)$$

We have, for  $j = 0, \dots, N-1$ ,

$$\begin{aligned} &\sup_{x \in [x_j, x_{j+1}]} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x \gamma_{t,n}(\mathbf{a})} - F(x \gamma_{t,n}(\mathbf{a})) - \{ \mathbb{1}_{\eta_t \leq x_j \gamma_{t,n}(\mathbf{a})} - F(x_j \gamma_{t,n}(\mathbf{a})) \} \right| \\ &\leq \sup_{x \in [x_j, x_{j+1}]} \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x \gamma_{t,n}(\mathbf{a})} - \mathbb{1}_{\eta_t \leq x_j \gamma_{t,n}(\mathbf{a})} \\ &\quad + \sup_{x \in [x_j, x_{j+1}]} \frac{1}{\sqrt{n}} \sum_{t=1}^n F(x \gamma_{t,n}(\mathbf{a})) - F(x_j \gamma_{t,n}(\mathbf{a})) \\ &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x_{j+1} \gamma_{t,n}(\mathbf{a})} - \mathbb{1}_{\eta_t \leq x_j \gamma_{t,n}(\mathbf{a})} + \frac{1}{\sqrt{n}} \sum_{t=1}^n F(x_{j+1} \gamma_{t,n}(\mathbf{a})) - F(x_j \gamma_{t,n}(\mathbf{a})) \\ &\leq \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \{ \mathbb{1}_{\eta_t \leq x_{j+1} \gamma_{t,n}(\mathbf{a})} - F(x_{j+1} \gamma_{t,n}(\mathbf{a})) \} - \{ \mathbb{1}_{\eta_t \leq x_j \gamma_{t,n}(\mathbf{a})} - F(x_j \gamma_{t,n}(\mathbf{a})) \} \right| \\ &\quad + 2W_n(j, \mathbf{a}), \end{aligned}$$

where  $W_n(j, \mathbf{a}) = n^{-1/2} \sum_{t=1}^n \{F(x_{j+1}\gamma_{t,n}(\mathbf{a})) - F(x_j\gamma_{t,n}(\mathbf{a}))\}$ . Therefore,

$$\begin{aligned} & \sup_{x \in [x_j, x_{j+1}]} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x\gamma_{t,n}(\mathbf{a})} - F(x\gamma_{t,n}(\mathbf{a})) - \{\mathbb{1}_{\eta_t \leq x_j\gamma_{t,n}(\mathbf{a})} - F(x_j\gamma_{t,n}(\mathbf{a}))\} \right| \\ & \leq \frac{1}{\sqrt{n}} \left| \sum_{t=1}^n z_{t,n}(x_{j+1}, \mathbf{a}) \right| + \frac{1}{\sqrt{n}} \left| \sum_{t=1}^n z_{t,n}(x_j, \mathbf{a}) \right| + V_n(j) + 2W_n(j, \mathbf{a}), \end{aligned} \quad (16)$$

where  $V_n(j) = n^{-1/2} \left| \sum_{t=1}^n \{\mathbb{1}_{\eta_t \leq x_{j+1}} - F(x_{j+1})\} - \{\mathbb{1}_{\eta_t \leq x_j} - F(x_j)\} \right|$ . By Assumption **A11** and the mean-value theorem,  $F(x_{j+1}) - F(x_j) \leq M\varepsilon/\sqrt{n}$  where  $M = \sup_{x \in \mathbb{R}} f(x)$ . From (16) with  $\mathbf{a} = \mathbf{0}$ , it follows that

$$\begin{aligned} & \sup_{x \in [x_j, x_{j+1}]} \frac{1}{\sqrt{n}} \left| \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x} - F(x) - \{\mathbb{1}_{\eta_t \leq x_j} - F(x_j)\} \right| \\ & \leq V_n(j) + 2\sqrt{n}\{F(x_{j+1}) - F(x_j)\} \leq V_n(j) + 2M\varepsilon. \end{aligned} \quad (17)$$

Therefore, for  $j = 0, \dots, N-1$ , from (16)-(17),

$$\begin{aligned} & \sup_{x \in [x_j, x_{j+1}]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) - \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x_j, \mathbf{a}) \right| \\ & \leq \frac{1}{\sqrt{n}} \left| \sum_{t=1}^n z_{t,n}(x_{j+1}, \mathbf{a}) \right| + \frac{1}{\sqrt{n}} \left| \sum_{t=1}^n z_{t,n}(x_j, \mathbf{a}) \right| + 2W_n(j, \mathbf{a}) + 2V_n(j) + 2M\varepsilon, \end{aligned}$$

so

$$\begin{aligned} & \max_{0 \leq j \leq N-1} \sup_{x \in [x_j, x_{j+1}]} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) - \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x_j, \mathbf{a}) \right| \\ & \leq \max_{0 \leq j \leq N} \frac{2}{\sqrt{n}} \left| \sum_{t=1}^n z_{t,n}(x_j, \mathbf{a}) \right| + 2 \max_{0 \leq j \leq N-1} W_n(j, \mathbf{a}) + 2 \max_{0 \leq j \leq N-1} V_n(j) \\ & \quad + 2M\varepsilon. \end{aligned} \quad (18)$$

By the properties of the modulus of continuity of the empirical process (see Shorack and Wellner (1986), p. 542), under Assumption **A11** we have

$$\max_{0 \leq j \leq N-1} V_n(j) = o_P(1). \quad (19)$$

Now, using again the mean-value theorem,

$$\begin{aligned} \max_{0 \leq j \leq N-1} W_n(j, \mathbf{a}) & \leq \max_{0 \leq j \leq N-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n (x_{j+1} - x_j) \gamma_{t,n}(\mathbf{a}) M \\ & \leq \frac{M\varepsilon}{n} \sum_{t=1}^n \gamma_{t,n}(\mathbf{a}) \leq \frac{M\varepsilon}{n} \sum_{t=1}^n \left( 1 + \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_t^*)}{\partial \boldsymbol{\theta}} \right\| \frac{\|\mathbf{a}\|}{\sqrt{n}} \right), \end{aligned}$$

where  $\boldsymbol{\theta}_t^*$  is between  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\theta}_0 + \mathbf{a}/\sqrt{n}$ . Thus

$$\max_{0 \leq j \leq N-1} W_n(j, \mathbf{a}) \leq \frac{M\varepsilon}{n} \sum_{t=1}^n \left( 1 + \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left( \frac{\sigma_t(\boldsymbol{\theta})}{\sigma_t(\boldsymbol{\theta}_0)} \right) \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| \frac{\|\mathbf{a}\|}{\sqrt{n}} \right)$$

$$= \varepsilon \times O_P(1), \quad (20)$$

by Assumption **A9** and the ergodic theorem. Thus (14), (19) and (20) show that the right-hand side of (18) is an  $o_P(1)$ . It is clear that the same bound can be obtained when  $\max_{0 \leq j \leq N-1}$  is replaced by  $\max_{-N \leq j \leq -1}$ . Thus (15) is established.  $\square$

LEMMA 4. Let  $\mathcal{K}$  be a compact subset of  $\mathbb{R}$ . For any  $A > 0$  and  $\mathbf{A} = [-A, A]^d$ ,

$$\sup_{x \in \mathcal{K}} \sup_{\mathbf{a} \in \mathbf{A}} |e_{n,1}(x, \mathbf{a}) - e_n(x)| = o_P(1).$$

**Proof.** In view of (12) will show that

$$\sup_{\mathbf{a} \in \mathbf{A}} X_n(\mathbf{a}) = o_P(1), \quad \text{where} \quad X_n(\mathbf{a}) = \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) \right|. \quad (21)$$

Let  $\varepsilon > 0$  such that  $N := 2A/\varepsilon$  is an integer and define  $a(k) = -A + k\varepsilon$ , for  $1 \leq k \leq N$ . For any  $1 \leq k_1, k_2, \dots, k_d \leq N$  let  $\mathbf{k} = (k_1, \dots, k_d)$  and consider the grid of  $N^d$  points  $\mathbf{a}(\mathbf{k}) = (a(k_1), \dots, a(k_d))$ . Let also  $\mathbf{A}(\mathbf{k}) = \{(a_1, \dots, a_d) \in \mathbf{A} | a(k_i) - \varepsilon \leq a_i \leq a(k_i)\}$  and  $\mathbf{a}^*(\mathbf{k}) = (a(k_1) - \varepsilon, \dots, a(k_d) - \varepsilon)$ . We have, for  $j = 1, \dots, d$  and  $a_j \leq a(k_j)$

$$\begin{aligned} & F(x\gamma_{t,n}(a_1, \dots, a_{j-1}, a(k_j), a_{j+1}, \dots, a_d)) - F(x\gamma_{t,n}(a_1, \dots, a_{j-1}, a_j, a_{j+1}, \dots, a_d)) \\ &= f\{x\gamma_{t,n}(\mathbf{a}_{t,j}^*)\} \frac{x}{\sqrt{n}} (a(k_j) - a_j) \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_0 + \frac{\mathbf{a}_{t,j}^*}{\sqrt{n}})}{\partial \boldsymbol{\theta}'} \mathbf{e}_j, \end{aligned}$$

where  $\mathbf{e}_j$  is the  $j$ -th element of the canonical basis of  $\mathbb{R}^d$ , and  $\mathbf{a}_{t,j}^*$  is a point between the arguments of  $\gamma_{t,n}$  above. By **A11** and  $E|\eta_t| < \infty$ , we have  $\sup_x |x|f(x) < \infty$ . The latter difference is thus bounded, uniformly in  $x \in \mathbb{R}$  and  $a_j \in [a(k_j) - \varepsilon, a(k_j)]$ , by

$$K \frac{\varepsilon}{\sqrt{n}} \frac{1}{\sigma_t(\boldsymbol{\theta}_0 + \frac{\mathbf{a}_{t,j}^*}{\sqrt{n}})} \frac{\partial \sigma_t(\boldsymbol{\theta}_0 + \frac{\mathbf{a}_{t,j}^*}{\sqrt{n}})}{\partial \boldsymbol{\theta}'} \mathbf{e}_j.$$

Therefore, for  $n$  large enough,

$$\sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k})} \sup_{x \in \mathbb{R}} \sum_{t=1}^n |F(x\gamma_{t,n}(\mathbf{a})) - F(x\gamma_{t,n}(\mathbf{a}(\mathbf{k})))| \leq K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta})} \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|,$$

and thus, because  $\gamma_{t,n}(\cdot)$  is an increasing function of its arguments by **A10**,

$$\begin{aligned} & \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k})} \sup_{x \in \mathcal{K}} \left| \sum_{t=1}^n z_{t,n}(x, \mathbf{a}) - z_{t,n}(x, \mathbf{a}(\mathbf{k})) \right| \\ & \leq K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| + \sup_{x \in \mathcal{K}} \left| \sum_{t=1}^n \mathbf{1}_{\eta_t \leq x\gamma_{t,n}(\mathbf{a}(\mathbf{k}))} - \mathbf{1}_{\eta_t \leq x\gamma_{t,n}(\mathbf{a}^*(\mathbf{k}))} \right| \\ & \leq 2K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| + \sup_{x \in \mathcal{K}} \left| \sum_{t=1}^n z_{t,n}(x, \mathbf{a}(\mathbf{k})) \right| + \sup_{x \in \mathcal{K}} \left| \sum_{t=1}^n z_{t,n}(x, \mathbf{a}^*(\mathbf{k})) \right|. \end{aligned}$$

Note that

$$\begin{aligned}
& \sup_{\mathbf{a} \in \mathbf{A}} X_n(\mathbf{a}) \\
& \leq \max_{\mathbf{k} \in \{1, \dots, N\}^d} \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k})} \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n [z_{t,n}(x, \mathbf{a}) - z_{t,n}(x, \mathbf{a}(\mathbf{k}))] \right| + \max_{\mathbf{k} \in \{1, \dots, N\}^d} X_n(\mathbf{a}(\mathbf{k})) \\
& \leq \frac{2K\varepsilon}{n} \sum_{t=1}^n \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| + 2 \max_{\mathbf{k} \in \{1, \dots, N\}^d} X_n(\mathbf{a}(\mathbf{k})) + \max_{\mathbf{k} \in \{1, \dots, N\}^d} X_n(\mathbf{a}^*(\mathbf{k})).
\end{aligned}$$

By the ergodic theorem and **A9**, the first term in the r.h.s. is almost surely less than a constant times  $\varepsilon$  when  $n$  is large. The two other terms tend to zero in probability because  $X_n(\mathbf{a}) = o_P(1)$  by Lemma 3 and the maximas are over a finite number of points. Therefore (21) is established.  $\square$

LEMMA 5. *Let  $\mathcal{K}$  be a compact subset of  $\mathbb{R}$ . Then  $\sup_{x \in \mathcal{K}} |\widehat{e}_{n,2}(x) - h_n(x)| \rightarrow 0$  a.s.*

**Proof.** A Taylor expansion yields, for  $x_t^* = x\sigma_t(\boldsymbol{\theta}_t^*)/\sigma_t(\boldsymbol{\theta}_0)$  with  $\boldsymbol{\theta}_t^*$  between  $\widehat{\boldsymbol{\theta}}_n$  and  $\boldsymbol{\theta}_0$ ,

$$\begin{aligned}
& |\widehat{e}_{n,2}(x) - h_n(x)| \\
& = \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ F(x\chi_{t,n}) - F(x) - xf(x) \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right\} \right| \\
& \leq |x| \frac{1}{n} \sum_{t=1}^n \left| f(x_t^*) \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta}_t^*)}{\partial \boldsymbol{\theta}} - f(x) \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right| \left\| \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right\| \\
& \leq |x| \left( \frac{1}{n} \sum_{t=1}^n |f(x_t^*) - f(x)|^2 \right)^{1/2} \left( \frac{1}{n} \sum_{t=1}^n \left| \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta}_t^*)}{\partial \boldsymbol{\theta}} \right| \right)^{1/2} \left\| \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right\| \\
& \quad + |x| f(x) \frac{1}{n} \sum_{t=1}^n \left| \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta}_t^*)}{\partial \boldsymbol{\theta}} - \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right| \left\| \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right\| \tag{22}
\end{aligned}$$

by Cauchy-Schwarz inequality. We have, for any  $\Delta > 0$  and  $n$  large enough

$$\frac{1}{n} \sum_{t=1}^n \sup_{x \in \mathcal{K}} |f(x_t^*) - f(x)|^2 \leq \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\delta}: \|\boldsymbol{\delta}\| \leq \Delta} \sup_{x \in \mathcal{K}} \left| f\left(x \frac{\sigma_t(\boldsymbol{\theta}_0 + \boldsymbol{\delta})}{\sigma_t(\boldsymbol{\theta}_0)}\right) - f(x) \right|^2 + \mathbb{1}_{\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| > \Delta}.$$

The first term in the r.h.s. converges a.s. to

$$E \sup_{\boldsymbol{\delta}: \|\boldsymbol{\delta}\| \leq \Delta} \sup_{x \in \mathcal{K}} \left| f\left(x \frac{\sigma_t(\boldsymbol{\theta}_0 + \boldsymbol{\delta})}{\sigma_t(\boldsymbol{\theta}_0)}\right) - f(x) \right|^2$$

which can be made arbitrarily small for  $\Delta$  small enough by the dominated convergence theorem.

The last term in the r.h.s. is a.s. equal to zero for  $n$  large enough. It follows that the first term is the r.h.s. of the inequality (22) tends to zero almost surely uniformly in  $x \in \mathcal{K}$ . The second term is handled similarly.  $\square$

LEMMA 6. *For any  $s > 0$ , we have  $n^{1/2-s} \left\| \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right\| = o(1)$  a.s.*

**Proof.** This is a direct consequence of the Bahadur expansion

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = \frac{-\mathbf{J}^{-1}}{2\sqrt{n}} \sum_{t=1}^n (1 - \eta_t^2) \mathbf{D}_t + o_P(1) \quad (23)$$

and Lemma 1 with  $X_t$  being any component of  $(1 - \eta_t^2) \mathbf{D}_t$  (noting that  $EX_t^2 < \infty$  under **A9**).  $\square$

LEMMA 7. We have  $\sup_{x \in \mathbb{R}} |\hat{e}_{n,3}(x)| = o_P(1)$ .

**Proof.** In view of **A2-A3**, we have

$$\left| \eta_t(\hat{\boldsymbol{\theta}}_n) - \tilde{\eta}_t(\hat{\boldsymbol{\theta}}_n) \right| = \frac{|\tilde{\sigma}_t(\hat{\boldsymbol{\theta}}_n) - \sigma_t(\hat{\boldsymbol{\theta}}_n)|}{\tilde{\sigma}_t(\hat{\boldsymbol{\theta}}_n)\sigma_t(\hat{\boldsymbol{\theta}}_n)} \sigma_t |\eta_t| \leq K \rho^t \sigma_t |\eta_t|.$$

It follows that

$$\begin{aligned} \left| \mathbb{1}_{\eta_t \leq x \tilde{\chi}_{t,n}} - \mathbb{1}_{\eta_t \leq x \chi_{t,n}} \right| &= \left| \mathbb{1}_{\tilde{\eta}_t(\hat{\boldsymbol{\theta}}_n) \leq x} - \mathbb{1}_{\eta_t(\hat{\boldsymbol{\theta}}_n) \leq x} \right| \leq \mathbb{1}_{x - K \rho^t \sigma_t |\eta_t| \leq \eta_t(\hat{\boldsymbol{\theta}}_n) \leq x + K \rho^t \sigma_t |\eta_t|} \\ &\leq \mathbb{1}_{x - K \rho^{t/3} \leq \eta_t(\hat{\boldsymbol{\theta}}_n) \leq x + K \rho^{t/3}} + \mathbb{1}_{|\eta_t| \geq 1/\rho^{t/3}} + \mathbb{1}_{\sigma_t \geq 1/\rho^{t/3}}. \end{aligned}$$

Now, by Markov's inequality and **A1**

$$E \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\sigma_t \geq 1/\rho^{t/3}} = \frac{1}{\sqrt{n}} \sum_{t=1}^n P(\rho^{t/3} \sigma_t > 1) = O(n^{-1/2}).$$

Similarly it can be shown that  $n^{-1/2} \sum_{t=1}^n \mathbb{1}_{|\eta_t| \geq 1/\rho^{t/3}} = O_P(n^{-1/2})$ . It remains to show that

$$\sup_{x \in \mathbb{R}} \frac{1}{\sqrt{n}} \sum_{t=1}^n d_{t,n}(x) = o_P(1), \quad d_{t,n}(x) = \mathbb{1}_{\chi_{t,n}x - K \chi_{t,n} \rho^{t/3} \leq \eta_t \leq \chi_{t,n}x + K \chi_{t,n} \rho^{t/3}}.$$

Let  $(a_n)$  and  $(b_n)$  be two sequences of real numbers, and let  $s > 0$ . Consider the events  $A_n = \{\min_{1 \leq t < t' \leq n} |\eta_t - \eta_{t'}| \geq a_n\}$ . We have  $\sup_{x \in \mathbb{R}} n^{-1/2} \sum_{t=1}^n d_{t,n}(x) \leq \sum_{i=1}^5 c_{i,n}$ , with

$$\begin{aligned} c_{1,n} &= \sup_{x \in \mathbb{R}} \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n^{1/4} \rceil - 1} d_{t,n}(x) \leq \frac{1}{n^{1/4}}, \\ c_{2,n} &= \sup_{x \in \mathbb{R}} \frac{1}{\sqrt{n}} \sum_{t=\lceil n^{1/4} \rceil}^n \mathbb{1}_{\chi_{t,n}x - K b_n \rho_n \leq \eta_t \leq \chi_{t,n}x + K b_n \rho_n} \mathbb{1}_{A_n}, \quad \rho_n = \rho^{\frac{n^{1/4}}{3}}, \\ c_{3,n} &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\chi_{t,n} \geq b_n} \mathbb{1}_{\sqrt{n} \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \leq n^s}, \quad c_{4,n} = \sqrt{n} \mathbb{1}_{\sqrt{n} \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \geq n^s}, \quad c_{5,n} = \sqrt{n} \mathbb{1}_{A_n^c}. \end{aligned}$$

Note that  $c_{2,n} = 1/\sqrt{n}$  when  $a_n \geq 2K b_n \rho_n$ . Now, for some  $\boldsymbol{\theta}_t$  between  $\hat{\boldsymbol{\theta}}_n$  and  $\boldsymbol{\theta}_0$ , we have

$$\chi_{t,n} = 1 + \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

Using **A9** and assuming  $b_n > 1$ , it follows that

$$Ec_{3,n} = \frac{1}{\sqrt{n}} \sum_{t=1}^n P\left(\chi_{t,n} \geq b_n, \sqrt{n} \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \leq n^s\right)$$

$$\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n P \left( \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\| n^{s-1/2} \geq b_n - 1 \right) \leq K \frac{n^{2s}}{\sqrt{n}(b_n - 1)^2}.$$

By Lemma 6, we have  $c_{4,n} = 0$  a.s. for  $n$  large enough. Since  $f$  is bounded under **A11**, we have

$$P(|\eta_1 - \eta_2| < a_n) = \int \left( \int_{x-a_n}^{x+a_n} f(y) dy \right) f(x) dx \leq K a_n.$$

Therefore  $Ec_{5,n} = \sqrt{n}P(A_n^c) \leq \sqrt{n}K \binom{n}{2} a_n \leq K n^{2+1/2} a_n$ . By taking, for instance,  $a_n = n^{-3}$ ,  $b_n = \sqrt{n}$  and  $s < 1/2$ , it can be checked that  $c_{i,n} = o_P(1)$  for  $i = 1, \dots, 5$ . We conclude that

$$|\widehat{e}_{n,3}(x)| \leq \frac{1}{\sqrt{n}} \sum_{t=1}^n |\mathbb{1}_{\eta_t \leq x \tilde{\chi}_{t,n}} - \mathbb{1}_{\eta_t \leq x \chi_{t,n}}| = o_P(1). \quad \square$$

LEMMA 8. *Let  $\mathcal{K}$  be a compact subset of  $\mathbb{R}$ . We have  $\sup_{x \in \mathcal{K}} |\widehat{e}_n(x) - e_n(x) - h_n(x)| = o_P(1)$ .*

**Proof.** We have

$$|\widehat{e}_n(x) - e_n(x) - h_n(x)| \leq |\widehat{e}_{n,1}(x) - e_n(x)| + |\widehat{e}_{n,2}(x) - h_n(x)| + |\widehat{e}_{n,3}(x)|.$$

Note that  $\widehat{e}_{n,1}(x) = e_{n,1}(x, \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))$ . Thus, any  $u, A > 0$  and  $\mathbf{A} = [-A, A]^d$ ,

$$\begin{aligned} P \left( \sup_{x \in \mathcal{K}} |\widehat{e}_{n,1}(x) - e_n(x)| > u \right) &\leq P \left( \sup_{x \in \mathcal{K}} \sup_{\mathbf{a} \in \mathbf{A}} |e_{n,1}(x, \mathbf{a}) - e_n(x)| > u \right) \\ &\quad + P \left( \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \notin \mathbf{A} \right). \end{aligned}$$

By Lemma 4, the first term in the r.h.s. tends to zero, while the second term can be made arbitrarily small by letting  $A \rightarrow \infty$ . Thus  $\sup_{x \in \mathcal{K}} |\widehat{e}_{n,1}(x) - e_n(x)| = o_P(1)$ . The conclusion follows using Lemmas 5 and 7.  $\square$

The stochastic equicontinuity property stated in the following lemma follows from Billingsley (1968) (see Appendix).

LEMMA 9. *If  $x$  is a real number and  $(x_n)$  is a sequence of real random variables such that  $x_n \rightarrow x$  in probability, and if  $F$  is continuous, then  $e_n(x_n) - e_n(x) = o_P(1)$ .*

Now we complete the proof of Theorem 2 establishing the asymptotic distribution of  $\widehat{e}_n(x_n)$ . By **A11**,  $h_n(x_n) - h_n(x) = o_P(1)$ . By Lemmas 8-9 and (23), it then follows that

$$\begin{aligned} \widehat{e}_n(x_n) &= e_n(x_n) + h_n(x_n) + o_P(1) \\ &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \{\mathbb{1}_{\eta_t < x} - F(x)\} - \frac{x f(x)}{2\sqrt{n}} \boldsymbol{\Omega}' \mathbf{J}^{-1} \sum_{t=1}^n (1 - \eta_t^2) \mathbf{D}_t + o_P(1). \end{aligned}$$

We conclude by using  $\boldsymbol{\Omega}' \mathbf{J}^{-1} \boldsymbol{\Omega} = 1$  under **A12** (see Remark 3 in Francq and Zakoian, 2013a).  $\square$

#### 6.4. Proof of Theorem 4

By comparison with the proof of Theorem 2, the major difficulty comes from the non-monotonicity of the conditional mean and variance with respect to the parameters (**A10** can no longer be assumed). To solve this difficulty we will introduce a partition of the past, so that the mean and variance are monotonous conditional on this partition.

We recall some notations and introduce new ones. Let

$$\begin{aligned}\widehat{e}_n(x) &= \sqrt{n}\{\widehat{F}_n(x) - F(x)\}, & e_n(x) &= \sqrt{n}\{F_n(x) - F(x)\}, & \tilde{\chi}_{t,n} &= \frac{\tilde{\sigma}_t(\widehat{\boldsymbol{\theta}}_n)}{\sigma_t(\boldsymbol{\theta}_0)}, \\ \chi_{t,n} &= \frac{\sigma_t(\widehat{\boldsymbol{\theta}}_n)}{\sigma_t(\boldsymbol{\theta}_0)}, & \tilde{\lambda}_{t,n} &= \frac{\tilde{m}_t(\widehat{\boldsymbol{\theta}}_n) - m_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0)}, & \lambda_{t,n} &= \frac{m_t(\widehat{\boldsymbol{\theta}}_n) - m_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0)}, \\ \chi_{t,n}^* &= 1 + \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), & \lambda_{t,n}^* &= \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0), \\ H_n(x) &= f(x) \left( x \left\{ \frac{1}{n} \sum_{t=1}^n \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right\} + \left\{ \frac{1}{n} \sum_{t=1}^n \frac{1}{\sigma_t} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right\} \right) \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).\end{aligned}$$

We have

$$\begin{aligned}\widehat{e}_n(x) &= \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x\chi_{t,n}^* + \lambda_{t,n}^*} - F(x\chi_{t,n}^* + \lambda_{t,n}^*)}_{\widehat{e}_{n,4}(x)} + \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n F(x\chi_{t,n}^* + \lambda_{t,n}^*) - F(x)}_{\widehat{e}_{n,5}(x)} \\ &\quad + \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x\tilde{\chi}_{t,n} + \tilde{\lambda}_{t,n}} - \mathbb{1}_{\eta_t \leq x\chi_{t,n}^* + \lambda_{t,n}^*}}_{\widehat{e}_{n,6}(x)}.\end{aligned}\tag{24}$$

Let, for  $\mathbf{a}$  a vector of the same size as  $\boldsymbol{\theta}$ , sufficiently small so that  $\boldsymbol{\theta}_0 + \mathbf{a}/\sqrt{n} \in \boldsymbol{\Theta}$ ,

$$\begin{aligned}\lambda_{t,n}(\mathbf{a}) &= \frac{m_t(\boldsymbol{\theta}_0 + \frac{\mathbf{a}}{\sqrt{n}}) - m_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta}_0)}, & \gamma_{t,n}(\mathbf{a}) &= \frac{\sigma_t(\boldsymbol{\theta}_0 + \frac{\mathbf{a}}{\sqrt{n}})}{\sigma_t(\boldsymbol{\theta}_0)}, \\ e_{n,4}(x, \mathbf{a}) &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \mathbb{1}_{\eta_t \leq x\gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{a})} - F(x\gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{a})) \right\}, \\ \lambda_{t,n}^*(\mathbf{a}) &= \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \frac{\mathbf{a}}{\sqrt{n}}, & \gamma_{t,n}^*(\mathbf{a}) &= 1 + \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \frac{\mathbf{a}}{\sqrt{n}}.\end{aligned}$$

For  $\mathbf{a}, \mathbf{b} \in \mathbb{R}^d$  let

$$z_{t,n}^*(x, \mathbf{a}, \mathbf{b}) = \mathbb{1}_{\eta_t \leq x\gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b})} - F(x\gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b})) - \{\mathbb{1}_{\eta_t \leq x} - F(x)\}.$$

Write  $\underline{y}_{t-1} = (y_{t-1}, y_{t-2}, \dots)$ . Let  $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_{2^{2d}}\}$  the elements of  $\{-1, 1\}^{2d}$  taken in the lexicographical order (thus  $\mathbf{s}_1 = (-1, \dots, -1)$  and  $\mathbf{s}_{2^{2d}} = (1, \dots, 1)$ ). For any vector  $\mathbf{Z} \in \mathbb{R}^{2d}$ , let  $\mathbf{s}[\mathbf{Z}] = (s_1, \dots, s_{2d})$  denote the element of  $\mathcal{S}$  such that  $s_i = \text{sgn}(Z_i)$ ,  $i = 1, \dots, 2d$ . Define  $\{\mathcal{R}_i\}_{i \in I}$  a partition of  $\mathbb{R}^\infty$ , with  $I = \{1, \dots, 2^{2d}\}$  such that, for  $i \in I$ ,

$$\underline{y}_{t-1} \in \mathcal{R}_i \Leftrightarrow \mathbf{s} \left[ \left( \frac{\partial m_t}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0), \frac{\partial \sigma_t}{\partial \boldsymbol{\theta}'}(\boldsymbol{\theta}_0) \right)' \right] = \mathbf{s}_i := (\mathbf{s}_i^{(1)}, \mathbf{s}_i^{(2)})$$



where  $\mathbf{s}_i^{(j)} \in \{-1, 1\}^d, j = 1, 2$ .

LEMMA 10. *We have, for any  $u > 0$  and sufficiently large  $n$ , for any sequences  $(\mathbf{a}_i)_{i \in I}$  and  $(\mathbf{b}_i)_{i \in I}$*

$$P \left( \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^*(x, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| > u \right) \leq \frac{K}{nu^4} \left\{ (x \sup_{i \in I} \|\mathbf{a}_i\|)^2 + \sup_{i \in I} \|\mathbf{b}_i\|^2 + 1 \right\}.$$

**Proof.** We note that, for fixed  $n$  and  $x$ ,  $(\tilde{z}_t, \mathcal{F}_t)_{1 \leq t \leq n}$  is a martingale difference sequence, where  $\tilde{z}_t = \sum_{i \in I} z_{t,n}^*(x, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i}$ . Hence, by the Markov and Rosenthal inequalities

$$P \left( \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \tilde{z}_t \right| > u \right) \leq \frac{1}{n^2} \frac{1}{u^4} E \left( \sum_{t=1}^n \tilde{z}_t \right)^4 \leq \frac{K}{n^2 u^4} \left\{ E \left( \sum_{t=1}^n E(\tilde{z}_t^2 | \mathcal{F}_{t-1}) \right)^2 + \sum_{t=1}^n E \tilde{z}_t^4 \right\}.$$

We have

$$\begin{aligned} E(\tilde{z}_t^2 | \mathcal{F}_{t-1}) &= \sum_{i \in I} E[z_{t,n}^{*2}(x, \mathbf{a}_i, \mathbf{b}_i) | \mathcal{F}_{t-1}] \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \\ &\leq \sum_{i \in I} |F(x \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) - F(x)| \mathbb{1}_{y_{t-1} \in \mathcal{R}_i}. \end{aligned}$$

By **A11**, letting  $K = \sup_y f(y) < \infty$ ,

$$\begin{aligned} &|F(x \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) - F(x)| \\ &\leq |x| K |\gamma_{t,n}^*(\mathbf{a}_i) - 1| + K |\lambda_{t,n}^*(\mathbf{b}_i)| \\ &\leq \frac{K}{\sqrt{n}} \left\{ |x| \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| \|\mathbf{a}_i\| + \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| \|\mathbf{b}_i\| \right\}. \end{aligned}$$

The rest of the proof is similar to that of Lemma 2.  $\square$

LEMMA 11. *Let  $\mathcal{K}$  a compact subset of  $\mathbb{R}$ . We have, for any sequence  $(\mathbf{a}_i)_{i \in I}$  and  $(\mathbf{b}_i)_{i \in I}$  such that  $\sup_{i \in I} \|\mathbf{a}_i\| + \|\mathbf{b}_i\| < \infty$ ,*

$$\mathcal{X}_n((\mathbf{a}_i)_{i \in I}, (\mathbf{b}_i)_{i \in I}) := \sup_{x \in \mathcal{K}} \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^*(x, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| = o_P(1).$$

**Proof.** Using Lemma 10, the proof is similar to that of Lemma 3. See Appendix.  $\square$

LEMMA 12. *Let  $\mathcal{K}$  be a compact subset of  $\mathbb{R}$ . We have, for any  $A > 0$  and  $\mathbf{A} = [-A, A]^d$*

$$\sup_{x \in \mathcal{K}} \sup_{\mathbf{a} \in \mathbf{A}} |e_{n,4}(x, \mathbf{a}) - e_n(x)| = o_P(1).$$

**Proof.** Note that  $e_{n,4}(x, \mathbf{a}) - e_n(x) = n^{-1/2} \sum_{t=1}^n z_{t,n}^*(x, \mathbf{a}, \mathbf{a})$ . We will thus show that

$$\sup_{\mathbf{a}, \mathbf{b} \in \mathbf{A}} X_n^*(\mathbf{a}, \mathbf{b}) = o_P(1), \quad \text{where} \quad X_n^*(\mathbf{a}, \mathbf{b}) = \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^*(x, \mathbf{a}, \mathbf{b}) \right|. \quad (25)$$

Let  $\varepsilon > 0$  such that  $N := 2A/\varepsilon$  is an integer and define  $a(k) = -A + k\varepsilon$  for  $1 \leq k \leq N$ . For  $1 \leq k_1, k_2, \dots, k_d \leq N$ , let  $\mathbf{k} = (k_1, \dots, k_d)$  and consider the grid of  $N^d$  points  $\mathbf{a}(\mathbf{k}) = (a(k_1), \dots, a(k_d))$ . Let also  $\mathbf{A}(\mathbf{k}) = \{(a_1, \dots, a_d) \in \mathbf{A} | a(k_j) - \varepsilon \leq a_j \leq a(k_j), j \in \{1, \dots, d\}\}$ . Define, for  $i \geq 1$ , and  $\mathbf{s}_i = (\mathbf{s}_i^{(1)}, \mathbf{s}_i^{(2)}) = (s_{i1}^{(1)}, \dots, s_{id}^{(1)}, s_{i1}^{(2)}, \dots, s_{id}^{(2)})$

$$\bar{\mathbf{a}}_{\mathbf{s}_i^{(\ell)}}(\mathbf{k}) = (\bar{a}_{\mathbf{s}_{i1}^{(\ell)}}(k_1), \dots, \bar{a}_{\mathbf{s}_{id}^{(\ell)}}(k_d)), \quad \bar{a}_{\mathbf{s}_{ij}^{(\ell)}}(k_j) = a(k_j)\mathbb{1}_{\mathbf{s}_{ij}^{(\ell)}=1} + \{a(k_j) - \varepsilon\}\mathbb{1}_{\mathbf{s}_{ij}^{(\ell)}=-1},$$

$$\underline{\mathbf{a}}_{\mathbf{s}_i^{(\ell)}}(\mathbf{k}) = (\underline{a}_{\mathbf{s}_{i1}^{(\ell)}}(k_1), \dots, \underline{a}_{\mathbf{s}_{id}^{(\ell)}}(k_d)), \quad \underline{a}_{\mathbf{s}_{ij}^{(\ell)}}(k_j) = a(k_j)\mathbb{1}_{\mathbf{s}_{ij}^{(\ell)}=-1} + \{a(k_j) - \varepsilon\}\mathbb{1}_{\mathbf{s}_{ij}^{(\ell)}=1}.$$

We have, for  $j = 1, \dots, d$ ,

$$\begin{aligned} & F(x\gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(b_1, \dots, b_{j-1}, a(k_j), b_{j+1}, \dots, b_d)) - F(x\gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b})) \\ &= f\{x\gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b}_{t,j})\} \frac{a(k_j) - b_j}{\sqrt{n}} \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mathbf{e}_j, \end{aligned}$$

where  $\mathbf{e}_j$  is the  $j$ -th element of the canonical basis of  $\mathbb{R}^d$ , and  $\mathbf{b}_{t,j}$  is a point between the arguments of  $\lambda_{t,n}$  above. By **A11**, we have  $\sup_x f(x) < \infty$ . The absolute value of the latter difference is thus bounded, uniformly in  $x \in \mathbb{R}$ ,  $\mathbf{a} \in \mathbf{A}$  and  $\mathbf{b} \in \mathbf{A}(\mathbf{k})$ , by  $K \frac{\varepsilon}{\sqrt{n}} \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \left| \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mathbf{e}_j \right|$ .

Similarly, for  $j = 1, \dots, d$ ,

$$\begin{aligned} & \left| F(x\gamma_{t,n}^*(a_1, \dots, a_{j-1}, a(k_j), a_{j+1}, \dots, a_d) + \lambda_{t,n}^*(\mathbf{b})) - F(x\gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b})) \right| \\ &= \left| x f\{x\gamma_{t,n}^*(\mathbf{a}_{t,j}) + \lambda_{t,n}^*(\mathbf{b})\} \frac{a(k_j) - a_j}{\sqrt{n}} \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mathbf{e}_j \right| \leq K \frac{\varepsilon}{\sqrt{n}} \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \left| \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mathbf{e}_j \right|, \end{aligned}$$

uniformly in  $x \in \mathcal{K}$ ,  $\mathbf{b} \in \mathbf{A}$  and  $\mathbf{a} \in \mathbf{A}(\mathbf{k})$ . Therefore, for  $n$  large enough,

$$\begin{aligned} & \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k}), \mathbf{b} \in \mathbf{A}(\boldsymbol{\ell})} \sup_{x \in \mathcal{K}} \sum_{t=1}^n \left| F(x\gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b})) - F(x\gamma_{t,n}^*(\mathbf{a}(\mathbf{k})) + \lambda_{t,n}^*(\mathbf{b}(\boldsymbol{\ell}))) \right| \\ & \leq K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| + \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| := K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n Z_t. \end{aligned}$$

It can be shown (see Appendix) that

$$\begin{aligned} & \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k}), \mathbf{b} \in \mathbf{A}(\boldsymbol{\ell})} \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^*(x, \mathbf{a}, \mathbf{b}) - z_{t,n}^*(x, \mathbf{a}(\mathbf{k}), \mathbf{a}(\boldsymbol{\ell})) \right| \\ & \leq \mathcal{X}_n \left( (\bar{\mathbf{a}}_{\mathbf{s}_i^{(2)}}(\mathbf{k}))_{i \in I}, (\bar{\mathbf{a}}_{\mathbf{s}_i^{(1)}}(\boldsymbol{\ell}))_{i \in I} \right) + \mathcal{X}_n \left( (\underline{\mathbf{a}}_{\mathbf{s}_i^{(2)}}(\mathbf{k}))_{i \in I}, (\underline{\mathbf{a}}_{\mathbf{s}_i^{(1)}}(\boldsymbol{\ell}))_{i \in I} \right) \\ & \quad + \mathcal{X}_n \left( (\underline{\mathbf{a}}_{\mathbf{s}_i^{(2)}}(\mathbf{k}))_{i \in I}, (\bar{\mathbf{a}}_{\mathbf{s}_i^{(1)}}(\boldsymbol{\ell}))_{i \in I} \right) + \mathcal{X}_n \left( (\bar{\mathbf{a}}_{\mathbf{s}_i^{(2)}}(\mathbf{k}))_{i \in I}, (\underline{\mathbf{a}}_{\mathbf{s}_i^{(1)}}(\boldsymbol{\ell}))_{i \in I} \right) + 3K \frac{\varepsilon}{n} \sum_{t=1}^n Z_t. \end{aligned}$$

Note that

$$\sup_{\mathbf{a}, \mathbf{b} \in \mathbf{A}} X_n^*(\mathbf{a}, \mathbf{b})$$

$$\begin{aligned}
&\leq \max_{\mathbf{k}, \ell \in \{1, \dots, N\}^d} \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k}), \mathbf{b} \in \mathbf{B}(\ell)} \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n [z_{t,n}^*(x, \mathbf{a}, \mathbf{b}) - z_{t,n}^*(x, \mathbf{a}(\mathbf{k}), \mathbf{b}(\ell))] \right| \\
&\quad + \max_{\mathbf{k}, \ell \in \{1, \dots, N\}^d} X_n^*(\mathbf{a}(\mathbf{k}), \mathbf{a}(\ell)) \\
&\leq \frac{3K\varepsilon}{n} \sum_{t=1}^n Z_t + 5 \max_{(\mathbf{a}_i), (\mathbf{b}_i)} \mathcal{X}_n((\mathbf{a}_i)_{i \in I}, (\mathbf{b}_i)_{i \in I}),
\end{aligned}$$

where the last max is taken over all sequences of the form  $\bar{\mathbf{a}}_{s_i^{(\ell)}}(\mathbf{k})$  or  $\underline{\mathbf{a}}_{s_i^{(\ell)}}(\mathbf{k}), \ell \in \{1, 2\}$  and  $\mathbf{k}, \ell \in \{1, \dots, N\}^d$ . The end of the proof of (25) is similar to that of (21) using Lemma 11.  $\square$

LEMMA 13. *Let  $\mathcal{K}$  be a compact subset of  $\mathbb{R}$ . Then  $\sup_{x \in \mathcal{K}} |\hat{e}_{n,5}(x) - H_n(x)| \rightarrow 0$  a.s.*

**Proof.** A Taylor expansion yields, for  $x_t^*$  between  $x$  and  $x\chi_{t,n}^* + \lambda_{t,n}^*$ ,

$$F(x\chi_{t,n}^* + \lambda_{t,n}^*) - F(x) = f(x_t^*) \frac{1}{\sigma_t} \left( x \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} + \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right) (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0).$$

It follows that

$$\begin{aligned}
|\hat{e}_{n,5}(x) - H_n(x)| &= \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \{f(x_t^*) - f(x)\} \frac{1}{\sigma_t} \left( x \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} + \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right) (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \right| \\
&\leq \left( \frac{1}{n} \sum_{t=1}^n |f(x_t^*) - f(x)|^2 \right)^{1/2} \\
&\quad \times \left( \frac{1}{n} \sum_{t=1}^n |x|^2 \left\{ \left\| \frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\|^2 + \left\| \frac{1}{\sigma_t} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\|^2 \right\} \right)^{1/2} \|\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)\|
\end{aligned}$$

by Cauchy-Schwarz inequality. We have, for any  $\Delta > 0$  and  $n$  large enough

$$\begin{aligned}
&\frac{1}{n} \sum_{t=1}^n \sup_{x \in \mathcal{K}} |f(x_t^*) - f(x)|^2 \\
&\leq \frac{1}{n} \sum_{t=1}^n \sup_{\boldsymbol{\delta}_\theta: \|\boldsymbol{\delta}_\theta\| \leq \Delta} \sup_{x \in \mathcal{K}} \left| f \left( x + \frac{1}{\sigma_t} \left( x \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} + \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right) \boldsymbol{\delta}_\theta \right) - f(x) \right|^2 \\
&\quad + \mathbb{1}_{\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| > \Delta}.
\end{aligned}$$

The first term in the r.h.s. converges a.s. to

$$E \sup_{\boldsymbol{\delta}_\theta: \|\boldsymbol{\delta}_\theta\| \leq \Delta} \sup_{x \in \mathcal{K}} \left| f \left( x + \frac{1}{\sigma_t} \left( x \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} + \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \right) \boldsymbol{\delta}_\theta \right) - f(x) \right|^2$$

which can be made arbitrarily small for  $\Delta$  small enough by the dominated convergence theorem.

The last term in the r.h.s. is a.s. equal to zero for  $n$  large enough. The conclusion follows.  $\square$

Write

$$\hat{e}_{n,6}(x) = \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x\tilde{\chi}_{t,n} + \tilde{\lambda}_{t,n}} - \mathbb{1}_{\eta_t \leq x\chi_{t,n} + \lambda_{t,n}}}_{\hat{e}_{n,7}(x)} + \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\eta_t \leq x\chi_{t,n} + \lambda_{t,n}} - \mathbb{1}_{\eta_t \leq x\chi_{t,n}^* + \lambda_{t,n}^*}}_{\hat{e}_{n,8}(x)}.$$

LEMMA 14. We have  $|\widehat{e}_{n,6}(x_n)| = o_P(1)$ .

**Proof.** We prove in the Appendix that  $\sup_{x \in \mathbb{R}} |\widehat{e}_{n,7}(x)| = o_P(1)$ . From Taylor expansions we get

$$\begin{aligned} & |x\chi_{t,n} + \lambda_{t,n} - (x\chi_{t,n}^* + \lambda_{t,n}^*)| \\ & \leq \frac{2}{\sigma_t(\boldsymbol{\theta}_0)} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \left\{ |x| \left\| \frac{\partial^2 \sigma_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| + \left\| \frac{\partial^2 m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| \right\} \left\| \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right\|^2 := \delta_t(x) \left\| \widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right\|^2, \end{aligned}$$

where  $\boldsymbol{\theta}_t, \underline{\boldsymbol{\theta}}_t$  and  $\bar{\boldsymbol{\theta}}_t$  are between  $\boldsymbol{\theta}_0$  and  $\widehat{\boldsymbol{\theta}}_n$ . It follows that, for any  $\Delta_n > 0$ ,

$$|\widehat{e}_{n,8}(x_n)| \leq \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{1}_{x_n \chi_{t,n}^* + \lambda_{t,n}^* - \delta_t(x_n) \Delta_n^2 \leq \eta_t \leq x_n \chi_{t,n}^* + \lambda_{t,n}^* + \delta_t(x_n) \Delta_n^2} + \sqrt{n} \mathbf{1}_{\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| > \Delta_n}. \quad (26)$$

Let  $\Delta_n = n^{s-1/2}$ , for some  $s > 0$ . The last term in (26) is equal to  $c_{4,n} = 0$  for sufficiently large  $n$ . Now, first suppose that the sequence  $(x_n)$  is non random,  $x_n \rightarrow x$ . Note that the sequence  $(\delta_t(x_n))_t$  is stationary, and integrable by Assumptions **A9** and **B7**. Let  $f_{t,n}$  the conditional density of  $\eta_t$  introduced in Assumption **B8**, and let  $K$  its uniform upper bound. Then, by the law of iterated expectations, the expectation of the first term in (26) is bounded, for  $n > n_0$ , by

$$2\sqrt{n}\Delta_n^2 KE\{\delta_1(x_n)\} + \frac{t(n)}{\sqrt{n}} = o(1).$$

The conclusion follows in case i). Turning to case ii), using Lemma 4.4 in Kreiss (1987), we can deduce that the conclusion also holds.  $\square$

LEMMA 15. Let  $\mathcal{K}$  be a compact subset of  $\mathbb{R}$ . We have  $\sup_{x \in \mathcal{K}} |\widehat{e}_n(x) - e_n(x) - H_n(x)| = o_P(1)$ .

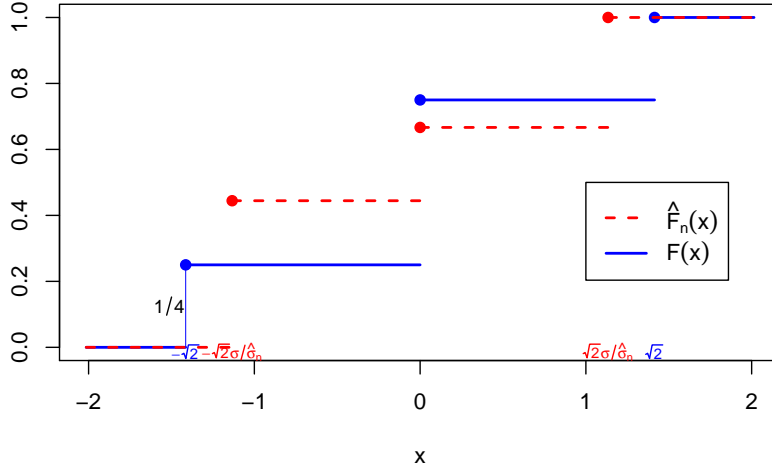
**Proof.** In view of (24),  $|\widehat{e}_n(x) - e_n(x) - H_n(x)| \leq |\widehat{e}_{n,4}(x) - e_n(x)| + |\widehat{e}_{n,5}(x) - H_n(x)| + |\widehat{e}_{n,6}(x)|$ . Note that  $\widehat{e}_{n,4}(x) = e_{n,4}(x, \sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0))$ . Thus, any  $u, A > 0$ ,  $\mathbf{A} = [-A, A]^d$ ,

$$\begin{aligned} P\left(\sup_{x \in \mathcal{K}} |\widehat{e}_{n,4}(x) - e_n(x)| > u\right) & \leq P\left(\sup_{x \in \mathcal{K}} \sup_{\mathbf{a} \in \mathbf{A}} |e_{n,4}(x, \mathbf{a}) - e_n(x)| > u\right) \\ & \quad + P\left(\sqrt{n}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \notin \mathbf{A}\right). \end{aligned}$$

The conclusion follows as in Lemma 8, using lemmas 12, 13 and 14.  $\square$

To complete the proof of Theorem 4, note that, by arguments used to prove Theorem 2, we have

$$\widehat{e}_n(x_n) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \{\mathbf{1}_{\eta_t < x} - F(x)\} + f(x)(\boldsymbol{\Omega}'_m + x\boldsymbol{\Omega}'_\sigma) \frac{1}{\sqrt{n}} \sum_{t=1}^n \boldsymbol{\Delta}_{t-1} \mathbf{V}(\eta_t) + o_P(1). \quad \square$$



**Fig. 1.** Failure of the Glivenko-Cantelli property for residuals when A5 does not hold (Example 1).

## Appendix

### A. Illustration of Example 1 and Remark 1

A graph illustrating the discrepancy between  $\hat{F}_n(x)$  and  $F(x)$  is displayed in Figure 1.

A graph illustrating the effect of the estimation on the asymptotic distribution of the e.d.f.r in displayed in Figure 2.

### B. Proof of Theorem 3

The proof requires the following changes in the proof of Theorem 1. Let  $\eta_t(\boldsymbol{\theta}) = (y_t - m_t(\boldsymbol{\theta}))/\sigma_t(\boldsymbol{\theta})$  and  $\tilde{\eta}_t(\boldsymbol{\theta}) = (y_t - \tilde{m}_t(\boldsymbol{\theta}))/\tilde{\sigma}_t(\boldsymbol{\theta})$ . We have  $\sup_{\boldsymbol{\theta} \in \Theta} |\eta_t(\boldsymbol{\theta}) - \tilde{\eta}_t(\boldsymbol{\theta})| \leq K_t \rho^t (1 + |y_t|)$ . Thus, by a Taylor expansion,

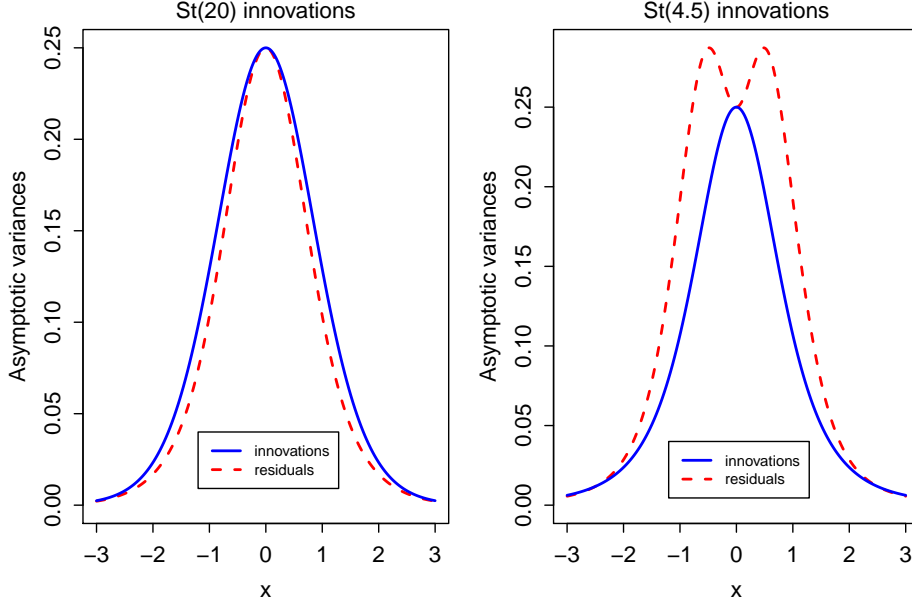
$$|\hat{\eta}_t - \eta_t| \leq \left\| \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right\| x_t + K_t \rho^t (1 + |y_t|) \quad (27)$$

for  $n$  large enough, where

$$x_t = \left\{ |\eta_t| \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{\sigma_t(\boldsymbol{\theta}_0)}{\sigma_t(\boldsymbol{\theta})} + \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{|m_t(\boldsymbol{\theta}_0) - m_t(\boldsymbol{\theta})|}{\sigma_t(\boldsymbol{\theta})} \right\} \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \|\mathbf{D}_t(\boldsymbol{\theta})\| + \sup_{\boldsymbol{\theta} \in V(\boldsymbol{\theta}_0)} \frac{1}{\sigma_t(\boldsymbol{\theta})} \left\| \frac{\partial m_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right\|. \quad (28)$$

In view of (27)-(28), we have

$$|\hat{\eta}_t - \eta_t| \leq K_t \rho^t (1 + |y_t|) + (u_t |\eta_t| + v_t) \left\| \hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 \right\|, \quad (29)$$



**Fig. 2.** Asymptotic variances of the empirical distribution for residuals (dotted red lines) and innovations (full blue line), for standardized Gaussian, GED (Generalized Error Distribution) and Student distributions.

for  $n$  large enough, where  $u_t, v_t \in \mathcal{F}_{t-1}$  and, by **B4**,  $E(|u_t|^{r/2}) < \infty$  and  $E(|v_t|^{r/2}) < \infty$ . The rest of the proof of the Glivenko-Cantelli property is similar to that of Theorem 1, using the consistency of  $\hat{\theta}_n$ .  $\square$

### C. Complements to Theorem 4

The next two results avoid the discreteness of the estimator in Theorem 4. Instead, we first assume faster convergence of the sequence  $(x_n)$ .

**PROPOSITION 1.** *Let  $x \in \mathbb{R}$  and let  $(x_n)$  a sequence of random variables such that  $\sqrt{n}(x_n - x) = o(1)$  a.s. Then, under **B1-B8** and **A9, A11**, the conclusions of Theorem 4 hold.*

**Proof.** The proof is the same as that of Theorem 4, except for Lemma 14. Let  $\varepsilon > 0$ ,  $\underline{x}_n = x - n^{-1/2}\varepsilon$ , and  $\bar{x}_n = x + n^{-1/2}\varepsilon$ . Then, by the law of iterated expectations, the expectation of the first term in (26) is bounded, for  $n$  large enough, by

$$\sqrt{n}(\bar{x}_n - \underline{x}_n)E(\chi_{1,n}^*) + 2\sqrt{n}\Delta_n^2 KE\{\delta_1(x_n)\} = 2\varepsilon + o(1).$$

Because  $\varepsilon$  can be chosen arbitrarily small, the conclusion follows.  $\square$

In our second result, discreteness is avoided at the price of reinforcing Assumption **B8**. For any sequence  $(x_n)$  of random variables, let the assumption

**B8** ( $x_n$ ): For  $n$  large enough,  $n > n_0$  say, the conditional distribution of  $\eta_t$  given  $(\hat{\theta}_n, x_n, \mathcal{F}_{t-1})$  admits a density which is a.s. bounded, uniformly in  $n > n_0$  and  $t < n - t(n)$  where  $t(n) = o(\sqrt{n})$ .

PROPOSITION 2. Let  $x \in \mathbb{R}$  and let  $(x_n)$  a sequence of random variables such that  $x_n \rightarrow x$  in probability. Then, under **B1-B7**, **B8**( $x_n$ ) and **A9**, **A11**, the conclusions of Theorem 4 hold.

**Proof.** Again, we just change the proof of Lemma 14. Noting that, for  $n > n_0$  and  $t < n - t(n)$

$$E \left( \mathbb{1}_{x_n \chi_{t,n}^* + \lambda_{t,n}^* - \delta_t(x_n) \Delta_n^2 \leq \eta_t \leq x_n \chi_{t,n}^* + \lambda_{t,n}^* + \delta_t(x_n) \Delta_n^2} \mid \hat{\theta}_n, x_n, \mathcal{F}_{t-1} \right) = 2K \delta_t(x_n) \Delta_n^2,$$

the conclusion follows.  $\square$

Noting that there is no guarantee that  $\gamma_{t,n}^*(\mathbf{a})$  be almost surely positive, we define

$$\gamma_{t,n}^{**}(\mathbf{a}) = |\gamma_{t,n}^*(\mathbf{a})|.$$

Before showing Lemma 11, we need to slightly modify Lemma 10. Let

$$z_{t,n}^{**}(x, \mathbf{a}, \mathbf{b}) = \mathbb{1}_{\eta_t \leq x \gamma_{t,n}^{**}(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b})} - F(x \gamma_{t,n}^{**}(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b})) - \{\mathbb{1}_{\eta_t \leq x} - F(x)\}.$$

LEMMA 16. We have, for any  $u > 0$  and sufficiently large  $n$ , for any sequences  $(\mathbf{a}_i)_{i \in I}$  and  $(\mathbf{b}_i)_{i \in I}$

$$P \left( \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^{**}(x, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| > u \right) \leq \frac{K}{nu^4} \left\{ (x \sup_{i \in I} \|\mathbf{a}_i\|)^2 + \sup_{i \in I} \|\mathbf{b}_i\|^2 + 1 \right\}.$$

**Proof.** Let  $\tilde{z}_t = \sum_{i \in I} z_{t,n}^{**}(x, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i}$ . Note that, for fixed  $n$  and  $x$ ,  $(\tilde{z}_t, \mathcal{F}_t)_{1 \leq t \leq n}$  is also a martingale difference sequence. Hence, we conclude by the arguments of the proof of Lemma 10 and

$$\begin{aligned} |F(x \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) - F(x)| &\leq |x|K|\gamma_{t,n}^{**}(\mathbf{a}_i) - 1| + K|\lambda_{t,n}^*(\mathbf{b}_i)| \\ &\leq |x|K|\gamma_{t,n}^*(\mathbf{a}_i) - 1| + K|\lambda_{t,n}^*(\mathbf{b}_i)|, \end{aligned}$$

using the elementary inequality  $||a| - |b|| \leq |a - b|$ .  $\square$

**Proof of Lemma 11.** With the notations of Lemma 3, Lemma 10 and Lemma 16 entail that for any  $u > 0$ ,

$$P \left( \max_{-N \leq j \leq N} \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^*(x_j, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| > u \right) \leq \frac{K}{\sqrt{n}}, \quad (30)$$

$$P \left( \max_{-N \leq j \leq N} \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^{**}(x_j, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| > u \right) \leq \frac{K}{\sqrt{n}}. \quad (31)$$

As argued in the proof of Lemma 3, using (30) it remains to show that

$$\max_{-N \leq j \leq N-1} \sup_{x \in [x_j, x_{j+1}]} \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n \{z_{t,n}^*(x, \mathbf{a}_i, \mathbf{b}_i) - z_{t,n}^*(x_j, \mathbf{a}_i, \mathbf{b}_i)\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| = o_P(1). \quad (32)$$

For all  $j$ , we have

$$\sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n \{z_{t,n}^*(x, \mathbf{a}_i, \mathbf{b}_i) - z_{t,n}^*(x_j, \mathbf{a}_i, \mathbf{b}_i)\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} = a_n(x) + b_n(x),$$

with

$$a_n(x) = \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \mathbb{1}_{\eta_t \leq x \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} - F(x \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) \right. \\ \left. - \mathbb{1}_{\eta_t \leq x_j \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} + F(x_j \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i}$$

and

$$b_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \mathbb{1}_{\eta_t \leq x} - F(x) - \mathbb{1}_{\eta_t \leq x_j} + F(x_j) \right\}.$$

By (17) and (19),  $\max_{-N \leq j \leq N-1} \sup_{x \in [x_j, x_{j+1}]} |b_n(x)| = o_P(1)$ . For  $j = 0, \dots, N-1$ ,

$$\begin{aligned} & \sup_{x \in [x_j, x_{j+1}]} |a_n(x)| \\ & \leq \frac{1}{\sqrt{n}} \sum_{i \in I} \sum_{t=1}^n \left\{ \mathbb{1}_{\eta_t \leq x_{j+1} \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} - \mathbb{1}_{\eta_t \leq x_j \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} \right\} \mathbb{1}_{\gamma_{t,n}^*(\mathbf{a}_i) \geq 0} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \\ & \quad + \frac{1}{\sqrt{n}} \sum_{i \in I} \sum_{t=1}^n \left\{ \mathbb{1}_{\eta_t \leq x_j \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} - \mathbb{1}_{\eta_t \leq x_{j+1} \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} \right\} \mathbb{1}_{\gamma_{t,n}^*(\mathbf{a}_i) < 0} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \\ & \quad + \frac{1}{\sqrt{n}} \sum_{i \in I} \sum_{t=1}^n \left\{ F(x_{j+1} \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) - F(x_j \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) \right\} \mathbb{1}_{\gamma_{t,n}^*(\mathbf{a}_i) \geq 0} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \\ & \quad + \frac{1}{\sqrt{n}} \sum_{i \in I} \sum_{t=1}^n \left\{ F(x_j \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) - F(x_{j+1} \gamma_{t,n}^*(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) \right\} \mathbb{1}_{\gamma_{t,n}^*(\mathbf{a}_i) < 0} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \\ & \leq \frac{1}{\sqrt{n}} \sum_{i \in I} \sum_{t=1}^n \left\{ \mathbb{1}_{\eta_t \leq x_{j+1} \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} - \mathbb{1}_{\eta_t \leq x_j \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \\ & \quad + \frac{1}{\sqrt{n}} \sum_{i \in I} \sum_{t=1}^n \left\{ \mathbb{1}_{\eta_t \leq -x_j \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} - \mathbb{1}_{\eta_t \leq -x_{j+1} \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)} \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \\ & \quad + \frac{1}{\sqrt{n}} \sum_{i \in I} \sum_{t=1}^n \left\{ F(x_{j+1} \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) - F(x_j \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \\ & \quad + \frac{1}{\sqrt{n}} \sum_{i \in I} \sum_{t=1}^n \left\{ F(-x_j \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) - F(-x_{j+1} \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \\ & \leq \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^{**}(x_{j+1}, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| + \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^{**}(x_j, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| \end{aligned}$$



$$+ \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^{**}(-x_{j+1}, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| + \left| \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^{**}(-x_j, \mathbf{a}_i, \mathbf{b}_i) \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \right| \\ + V_n^*(j) + 2W_n^*(j, \mathbf{a}_i, \mathbf{b}_i),$$

where

$$W_n^*(j, \mathbf{a}_i, \mathbf{b}_i) = \frac{1}{\sqrt{n}} \sum_{i \in I} \sum_{t=1}^n \{ F(x_{j+1} \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) - F(x_j \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) \\ + F(-x_j \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) - F(-x_{j+1} \gamma_{t,n}^{**}(\mathbf{a}_i) + \lambda_{t,n}^*(\mathbf{b}_i)) \} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i}, \\ V_n^*(j) = n^{-1/2} \left| \sum_{t=1}^n \{ \mathbb{1}_{\eta_t \leq x_{j+1}} - F(x_{j+1}) \} - \{ \mathbb{1}_{\eta_t \leq x_j} - F(x_j) \} \right. \\ \left. + \{ \mathbb{1}_{\eta_t \leq -x_j} - F(-x_j) \} - \{ \mathbb{1}_{\eta_t \leq -x_{j+1}} - F(-x_{j+1}) \} \right|.$$

Thus (32) is established using (31) and the arguments of the proof of Lemma 3.  $\square$

**Detailed proof of Lemma 12.** Note that  $e_{n,4}(x, \mathbf{a}) - e_n(x) = n^{-1/2} \sum_{t=1}^n z_{t,n}^*(x, \mathbf{a}, \mathbf{a})$ . We will thus show that

$$\sup_{\mathbf{a}, \mathbf{b} \in \mathbf{A}} X_n^*(\mathbf{a}, \mathbf{b}) = o_P(1), \quad \text{where} \quad X_n^*(\mathbf{a}, \mathbf{b}) = \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^*(x, \mathbf{a}, \mathbf{b}) \right|. \quad (33)$$

Let  $\varepsilon > 0$  such that  $N := 2A/\varepsilon$  is an integer and define  $a(k) = -A + k\varepsilon$  for  $1 \leq k \leq N$ . For  $1 \leq k_1, k_2, \dots, k_d \leq N$ , let  $\mathbf{k} = (k_1, \dots, k_d)$  and consider the grid of  $N^d$  points  $\mathbf{a}(\mathbf{k}) = (a(k_1), \dots, a(k_d))$ . Let also  $\mathbf{A}(\mathbf{k}) = \{(a_1, \dots, a_d) \in \mathbf{A} | a(k_j) - \varepsilon \leq a_j \leq a(k_j), j \in \{1, \dots, d\}\}$ . Define, for  $i \geq 1$ , and  $\mathbf{s}_i = (\mathbf{s}_i^{(1)}, \mathbf{s}_i^{(2)}) = (s_{i1}^{(1)}, \dots, s_{id}^{(1)}, s_{i1}^{(2)}, \dots, s_{id}^{(2)})$

$$\bar{\mathbf{a}}_{\mathbf{s}_i^{(\ell)}}(\mathbf{k}) = (\bar{a}_{s_{i1}^{(\ell)}}(k_1), \dots, \bar{a}_{s_{id}^{(\ell)}}(k_d)), \quad \bar{a}_{s_{ij}^{(\ell)}}(k_j) = a(k_j) \mathbb{1}_{s_{ij}^{(\ell)}=1} + \{a(k_j) - \varepsilon\} \mathbb{1}_{s_{ij}^{(\ell)}=-1},$$

$$\underline{\mathbf{a}}_{\mathbf{s}_i^{(\ell)}}(\mathbf{k}) = (\underline{a}_{s_{i1}^{(\ell)}}(k_1), \dots, \underline{a}_{s_{id}^{(\ell)}}(k_d)), \quad \underline{a}_{s_{ij}^{(\ell)}}(k_j) = a(k_j) \mathbb{1}_{s_{ij}^{(\ell)}=-1} + \{a(k_j) - \varepsilon\} \mathbb{1}_{s_{ij}^{(\ell)}=1}.$$

We have, for  $j = 1, \dots, d$ ,

$$F(x \gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(b_1, \dots, b_{j-1}, a(k_j), b_{j+1}, \dots, b_d)) - F(x \gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b})) \\ = f\{x \gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b}_{t,j})\} \frac{a(k_j) - b_j}{\sqrt{n}} \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mathbf{e}_j,$$

where  $\mathbf{e}_j$  is the  $j$ -th element of the canonical basis of  $\mathbb{R}^d$ , and  $\mathbf{b}_{t,j}$  is a point between the arguments of  $\lambda_{t,n}$  above. By **A11**, we have  $\sup_x f(x) < \infty$ . The absolute value of the latter difference is thus bounded, uniformly in  $x \in \mathbb{R}$ ,  $\mathbf{a} \in \mathbf{A}$  and  $\mathbf{b} \in \mathbf{A}(\mathbf{k})$ , by  $K \frac{\varepsilon}{\sqrt{n}} \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \left| \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mathbf{e}_j \right|$ .

Similarly, for  $j = 1, \dots, d$ ,

$$|F(x \gamma_{t,n}^*(a_1, \dots, a_{j-1}, a(k_j), a_{j+1}, \dots, a_d) + \lambda_{t,n}^*(\mathbf{b})) - F(x \gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b}))|$$

$$= \left| x f \left\{ x \gamma_{t,n}^*(\mathbf{a}_{t,j}) + \lambda_{t,n}^*(\mathbf{b}) \right\} \frac{a(k_j) - a_j}{\sqrt{n}} \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mathbf{e}_j \right| \leq K \frac{\varepsilon}{\sqrt{n}} \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \left| \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}'} \mathbf{e}_j \right|,$$

uniformly in  $x \in \mathcal{K}$ ,  $\mathbf{b} \in \mathbf{A}$  and  $\mathbf{a} \in \mathbf{A}(\mathbf{k})$ . Therefore, for  $n$  large enough,

$$\begin{aligned} & \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k}), \mathbf{b} \in \mathbf{A}(\ell)} \sup_{x \in \mathcal{K}} \sum_{t=1}^n \left| F \left( x \gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b}) \right) - F \left( x \gamma_{t,n}^*(\mathbf{a}(\mathbf{k})) + \lambda_{t,n}^*(\mathbf{b}(\ell)) \right) \right| \\ & \leq K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| + \left\| \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial m_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \right\| := K \frac{\varepsilon}{\sqrt{n}} \sum_{t=1}^n Z_t. \end{aligned}$$

For  $\mathbf{a} \in \mathbf{A}(\mathbf{k})$  and  $\mathbf{b} \in \mathbf{A}(\ell)$  we have, for  $y_{t-1} \in \mathcal{R}_i$  and  $x \geq 0$ ,

$$x \gamma_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(1)}}(\ell)) \leq x \gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b}) \leq x \gamma_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(1)}}(\ell)),$$

while, for  $x \leq 0$ ,

$$x \gamma_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(1)}}(\ell)) \leq x \gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b}) \leq x \gamma_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(1)}}(\ell)).$$

It follows that,

$$\begin{aligned} & \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k}), \mathbf{b} \in \mathbf{A}(\ell)} \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n z_{t,n}^*(x, \mathbf{a}, \mathbf{b}) - z_{t,n}^*(x, \mathbf{a}(\mathbf{k}), \mathbf{a}(\ell)) \right| \\ & \leq \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k}), \mathbf{b} \in \mathbf{A}(\ell)} \sup_{x \in \mathcal{K}} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left| \mathbb{1}_{\eta_t \leq x \gamma_{t,n}^*(\mathbf{a}(\mathbf{k})) + \lambda_{t,n}^*(\mathbf{a}(\ell))} - \mathbb{1}_{\eta_t \leq x \gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b})} \right| \\ & \quad + \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k}), \mathbf{b} \in \mathbf{A}(\ell)} \sup_{x \in \mathcal{K}} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left| F \left( x \gamma_{t,n}^*(\mathbf{a}(\mathbf{k})) + \lambda_{t,n}^*(\mathbf{a}(\ell)) \right) - F \left( x \gamma_{t,n}^*(\mathbf{a}) + \lambda_{t,n}^*(\mathbf{b}) \right) \right| \\ & \leq \sup_{x \in \mathcal{K}} \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \mathbb{1}_{\eta_t \leq x \gamma_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(1)}}(\ell))} - \mathbb{1}_{\eta_t \leq x \gamma_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(1)}}(\ell))} \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \mathbb{1}_{x \geq 0} \\ & \quad + \sup_{x \in \mathcal{K}} \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ \mathbb{1}_{\eta_t \leq x \gamma_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(1)}}(\ell))} - \mathbb{1}_{\eta_t \leq x \gamma_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(1)}}(\ell))} \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \mathbb{1}_{x < 0} \\ & \quad + K \frac{\varepsilon}{n} \sum_{t=1}^n Z_t \\ & = \sup_{x \in \mathcal{K}} \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ z^*(x, \bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k}), \bar{\mathbf{a}}_{s_i^{(1)}}(\ell)) - z^*(x, \underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k}), \underline{\mathbf{a}}_{s_i^{(1)}}(\ell)) \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \mathbb{1}_{x \geq 0} \\ & \quad + \sup_{x \in \mathcal{K}} \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ F(x \gamma_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(1)}}(\ell))) - F(x \gamma_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(1)}}(\ell))) \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \mathbb{1}_{x \geq 0} \\ & \quad + \sup_{x \in \mathcal{K}} \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ z^*(x, \underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k}), \bar{\mathbf{a}}_{s_i^{(1)}}(\ell)) - z^*(x, \bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k}), \underline{\mathbf{a}}_{s_i^{(1)}}(\ell)) \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \mathbb{1}_{x < 0} \\ & \quad + \sup_{x \in \mathcal{K}} \sum_{i \in I} \frac{1}{\sqrt{n}} \sum_{t=1}^n \left\{ F(x \gamma_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(1)}}(\ell))) - F(x \gamma_{t,n}^*(\bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})) + \lambda_{t,n}^*(\underline{\mathbf{a}}_{s_i^{(1)}}(\ell))) \right\} \mathbb{1}_{y_{t-1} \in \mathcal{R}_i} \mathbb{1}_{x < 0} \\ & \quad + K \frac{\varepsilon}{n} \sum_{t=1}^n Z_t \end{aligned}$$

$$\begin{aligned}
&\leq \mathcal{X}_n((\bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})), (\bar{\mathbf{a}}_{s_i^{(1)}}(\ell))) + \mathcal{X}_n((\underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})), (\underline{\mathbf{a}}_{s_i^{(1)}}(\ell))) \\
&\quad + \mathcal{X}_n((\underline{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})), (\bar{\mathbf{a}}_{s_i^{(1)}}(\ell))) + \mathcal{X}_n((\bar{\mathbf{a}}_{s_i^{(2)}}(\mathbf{k})), (\underline{\mathbf{a}}_{s_i^{(1)}}(\ell))) + 3K \frac{\varepsilon}{n} \sum_{t=1}^n Z_t.
\end{aligned}$$

Note that

$$\begin{aligned}
\sup_{\mathbf{a}, \mathbf{b} \in \mathbf{A}} X_n^*(\mathbf{a}, \mathbf{b}) &\leq \max_{\mathbf{k}, \ell \in \{1, \dots, N\}^d} \sup_{\mathbf{a} \in \mathbf{A}(\mathbf{k}), \mathbf{b} \in \mathbf{B}(\ell)} \sup_{x \in \mathcal{K}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n [z_{t,n}^*(x, \mathbf{a}, \mathbf{b}) - z_{t,n}^*(x, \mathbf{a}(\mathbf{k}), \mathbf{b}(\ell))] \right| \\
&\quad + \max_{\mathbf{k}, \ell \in \{1, \dots, N\}^d} X_n^*(\mathbf{a}(\mathbf{k}), \mathbf{a}(\ell)) \\
&\leq \frac{3K\varepsilon}{n} \sum_{t=1}^n Z_t + 5 \max_{(\mathbf{a}_i), (\mathbf{b}_i)} \mathcal{X}_n((\mathbf{a}_i), (\mathbf{b}_i)),
\end{aligned}$$

where the last max is taken over all sequences of the form  $\bar{\mathbf{a}}_{s_i^{(\ell)}}(\mathbf{k})$  or  $\underline{\mathbf{a}}_{s_i^{(\ell)}}(\mathbf{k}), \ell \in \{1, 2\}$  and  $\mathbf{k}, \ell \in \{1, \dots, N\}^d$ . The end of the proof of (33) is similar to that of (21) using Lemma 11.  $\square$

**Proof that  $\sup_{x \in \mathbb{R}} |\hat{e}_{n,7}(x)| = o_P(1)$  in Lemma 14.**

By **B2-B3**,  $|\eta_t(\hat{\boldsymbol{\theta}}_n) - \tilde{\eta}_t(\hat{\boldsymbol{\theta}}_n)| \leq K_t \rho^t$ . It follows that

$$\begin{aligned}
\left| \mathbb{1}_{\eta_t \leq x \tilde{\chi}_{t,n} + \tilde{\lambda}_{t,n}} - \mathbb{1}_{\eta_t \leq x \chi_{t,n} + \lambda_{t,n}} \right| &= \left| \mathbb{1}_{\tilde{\eta}_t(\hat{\boldsymbol{\theta}}_n) \leq x} - \mathbb{1}_{\eta_t(\hat{\boldsymbol{\theta}}_n) \leq x} \right| \leq \mathbb{1}_{x - K_t \rho^t \leq \eta_t(\hat{\boldsymbol{\theta}}_n) \leq x + K_t \rho^t} \\
&\leq \mathbb{1}_{x - \rho^{t/2} \leq \eta_t(\hat{\boldsymbol{\theta}}_n) \leq x + \rho^{t/2}} + \mathbb{1}_{K_t \rho^{t/2} \geq 1}.
\end{aligned}$$

We have, by **B3** and Markov inequality

$$E \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{K_t \rho^{t/2} \geq 1} = \frac{1}{\sqrt{n}} \sum_{t=1}^n P(K_t \rho^{t/2} \geq 1) \leq \frac{1}{\sqrt{n}} \sum_{t=1}^n E(K_t^r \rho^{tr/2}) = O(n^{-1/2}).$$

Letting

$$d_{t,n}(x) = \mathbb{1}_{x - \rho^{t/2} \leq \eta_t(\hat{\boldsymbol{\theta}}_n) \leq x + \rho^{t/2}} = \mathbb{1}_{\chi_{t,n} x - \chi_{t,n} \rho^{t/2} + \lambda_{t,n} \leq \eta_t \leq \chi_{t,n} x + \chi_{t,n} \rho^{t/2} + \lambda_{t,n}},$$

we have

$$\sup_{x \in \mathbb{R}} \frac{1}{\sqrt{n}} \sum_{t=1}^n d_{t,n}(x) \leq \sum_{i=1}^5 c_{i,n},$$

with

$$\begin{aligned}
c_{1,n} &= \sup_{x \in \mathbb{R}} \frac{1}{\sqrt{n}} \sum_{t=1}^{\lceil n^{1/4} \rceil - 1} d_{t,n}(x) \leq \frac{1}{n^{1/4}}, \\
c_{2,n} &= \sup_{x \in \mathbb{R}} \frac{1}{\sqrt{n}} \sum_{t=\lceil n^{1/4} \rceil}^n \mathbb{1}_{\chi_{t,n} x - b_n \rho_n + \lambda_{t,n} \leq \eta_t \leq \chi_{t,n} x + b_n \rho_n + \lambda_{t,n}} \mathbb{1}_{A_n}, \quad \rho_n = \rho^{\frac{n^{1/4}}{2}}, \\
c_{3,n} &= \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\chi_{t,n} \geq b_n} \mathbb{1}_{\sqrt{n} \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \leq n^s}, \quad c_{4,n} = \sqrt{n} \mathbb{1}_{\sqrt{n} \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \geq n^s}, \quad c_{5,n} = \sqrt{n} \mathbb{1}_{A_n^c},
\end{aligned}$$

where  $(a_n)$  and  $(b_n)$  are sequences of real numbers,  $A_n = \{\min_{1 \leq t < t' \leq n} |\eta_t - \eta_{t'}| \geq a_n\}$  and  $s > 0$ .

Note that  $c_{2,n} = 1/\sqrt{n}$  when  $a_n \geq 2b_n \rho_n$ . Now, for some  $\boldsymbol{\theta}_t$  between  $\hat{\boldsymbol{\theta}}_n$  and  $\boldsymbol{\theta}_0$ , we have

$$\chi_{t,n} = 1 + \frac{1}{\sigma_t(\boldsymbol{\theta}_0)} \frac{\partial \sigma_t(\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}'} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0). \quad (34)$$

Using **B7** and choosing for instance  $b_n = \sqrt{n} + 1$ , we obtain

$$\begin{aligned} Ec_{3,n} &= \frac{1}{\sqrt{n}} \sum_{t=1}^n P \left( \chi_{t,n} \geq b_n, \sqrt{n} \|\hat{\theta}_n - \theta_0\| \leq n^s \right) \\ &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n P \left( \sup_{\theta \in V(\theta_0)} \left\| \frac{1}{\sigma_t(\theta_0)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| n^{s-1/2} \geq \sqrt{n} \right) = o(1). \end{aligned}$$

By Lemma 1 and **B5**, we have almost surely  $c_{4,n} = 0$  for  $n$  large enough. In the proof of Lemma 7, we have seen that  $Ec_{5,n} = o(1)$  when, for instance,  $a_n = n^{-3}$ . Hence the lemma is established.  $\square$

**Detailed proof of Lemma 14.** From (34) and a similar expansion for  $\lambda_{t,n}$

$$\begin{aligned} &|x\chi_{t,n} + \lambda_{t,n} - (x\chi_{t,n}^* + \lambda_{t,n}^*)| \\ &= \left| \frac{1}{\sigma_t(\theta_0)} \left\{ x \left( \frac{\partial \sigma_t(\theta_t)}{\partial \theta'} - \frac{\partial \sigma_t(\theta_0)}{\partial \theta'} \right) + \left( \frac{\partial m_t(\theta_t)}{\partial \theta'} - \frac{\partial m_t(\theta_0)}{\partial \theta'} \right) \right\} (\hat{\theta}_n - \theta_0) \right| \\ &\leq \frac{1}{\sigma_t(\theta_0)} \left\{ |x| \left\| \frac{\partial \sigma_t(\theta_t)}{\partial \theta} - \frac{\partial \sigma_t(\theta_0)}{\partial \theta} \right\| + \left\| \frac{\partial m_t(\theta_t)}{\partial \theta} - \frac{\partial m_t(\theta_0)}{\partial \theta} \right\| \right\} \|\hat{\theta}_n - \theta_0\| \\ &\leq \frac{1}{\sigma_t(\theta_0)} \left\{ |x| \left\| \frac{\partial^2 \sigma_t(\bar{\theta}_t)}{\partial \theta \partial \theta'} (\theta_t - \theta_0) \right\| + \left\| \frac{\partial^2 m_t(\bar{\theta}_t)}{\partial \theta \partial \theta'} (\theta_t - \theta_0) \right\| \right\} \|\hat{\theta}_n - \theta_0\| \\ &\leq \frac{2}{\sigma_t(\theta_0)} \sup_{\theta \in V(\theta_0)} \left\{ |x| \left\| \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| + \left\| \frac{\partial^2 m_t(\theta)}{\partial \theta \partial \theta'} \right\| \right\} \|\hat{\theta}_n - \theta_0\|^2 := \delta_t(x) \|\hat{\theta}_n - \theta_0\|^2, \end{aligned}$$

where  $\theta_t, \bar{\theta}_t$  and  $\bar{\theta}_t$  are between  $\theta_0$  and  $\hat{\theta}_n$ . It follows that, for any  $\Delta_n > 0$ ,

$$\begin{aligned} |\hat{e}_{n,8}(x_n)| &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n |\mathbb{1}_{\eta_t \leq x_n \chi_{t,n} + \lambda_{t,n}} - \mathbb{1}_{\eta_t \leq x_n \chi_{t,n}^* + \lambda_{t,n}^*}| \\ &\leq \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{x_n \chi_{t,n}^* + \lambda_{t,n}^* - \delta_t(x_n) \Delta_n^2 \leq \eta_t \leq x_n \chi_{t,n}^* + \lambda_{t,n}^* + \delta_t(x_n) \Delta_n^2} + \sqrt{n} \mathbb{1}_{\|\hat{\theta}_n - \theta_0\| > \Delta_n}. \end{aligned} \quad (35)$$

Let  $\Delta_n = n^{s-1/2}$ , for some  $s > 0$ . The last term in (35) is equal to  $c_{4,n} = 0$  for sufficiently large  $n$ . Now, first suppose that the sequence  $(x_n)$  is non random,  $x_n \rightarrow x$ . Note that the sequence  $(\delta_t(x_n))_t$  is stationary, and integrable by Assumptions **A9** and **B7**. Let  $f_{t,n}$  the conditional density of  $\eta_t$  introduced in Assumption **B8**, and let  $K$  its uniform upper bound. Then, by the law of iterated expectations, the expectation of the first term in (35) is bounded, for  $n > n_0$ , by

$$2\sqrt{n}\Delta_n^2 K E\{\delta_1(x_n)\} + \frac{t(n)}{\sqrt{n}} = o(1).$$

The conclusion follows in case i). Turning to case ii), using Lemma 4.4 in Kreiss (1987), we can deduce that the conclusion also holds.  $\square$

## D. QMLE of the location scale model

### D.1. Consequence of Assumption A12\* for the QMLE

Consider the function  $g : c \mapsto g(c) = Q_n(\hat{\theta}_{n,c}) = \frac{1}{n} \sum_{t=1}^n \frac{\{y_t - \tilde{m}_t(\hat{\theta}_n)\}^2}{c^2 \tilde{\sigma}_t^2(\hat{\theta}_n)} + \log\{c^2 \tilde{\sigma}_t^2(\hat{\theta}_n)\}$ . Because the function  $g$  reaches its minimum for  $c = 1$ , we must have  $\hat{\mu}_2 = 1$  a.s.

Now we have

$$\hat{\mu}_2 := \frac{1}{n} \sum_{t=1}^n \tilde{\eta}_t^2(\hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^n \eta_t^2(\hat{\theta}_n) + o_P(1) = \frac{1}{n} \sum_{t=1}^n \eta_t^2 + \frac{\partial \mu_{2,n}(\theta_0)}{\partial \theta'} (\hat{\theta}_n - \theta_0) + o_P(1),$$

where

$$\mu_{2,n}(\theta) = \frac{1}{n} \sum_{t=1}^n \eta_t^2(\theta), \quad \frac{\partial \mu_{2,n}(\theta_0)}{\partial \theta'} = \frac{-2}{n} \sum_{t=1}^n \eta_t \left\{ \frac{1}{\sigma_t} \frac{\partial m_t(\theta_0)}{\partial \theta'} + \eta_t \frac{1}{\sigma_t} \frac{\partial \sigma_t(\theta_0)}{\partial \theta'} \right\} = -2\Omega'_\sigma + o_P(1).$$

Thus

$$\sqrt{n}(\hat{\mu}_2 - 1) = \frac{1}{\sqrt{n}} \sum_{t=1}^n (\eta_t^2 - 1) - 2\Omega'_\sigma \sqrt{n}(\hat{\theta}_n - \theta_0) + o_P(1). \quad (36)$$

It follows from (36) and (5) that

$$\begin{aligned} \sqrt{n}(\hat{\mu}_2 - 1) &= -4\Omega'_\sigma J^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \eta_t \frac{1}{\sigma_t} \frac{\partial m_t(\theta_0)}{\partial \theta} \\ &\quad + \frac{1}{\sqrt{n}} \sum_{t=1}^n (\eta_t^2 - 1) \left( 1 - 4\Omega'_\sigma J^{-1} \frac{1}{\sigma_t} \frac{\partial \sigma_t(\theta_0)}{\partial \theta} \right) + o_P(1). \end{aligned} \quad (37)$$

Since, in (37), the asymptotic distribution of the l.h.s. is degenerate, we deduce that

$$\begin{aligned} 0 &= 16\Omega'_\sigma J^{-1} \{J_m + (\kappa_4 - 1)J_\sigma + \mu_3(J_{m\sigma} + J_{\sigma m})\} J^{-1} \Omega_\sigma + (\kappa_4 - 1)(1 - 8\Omega'_\sigma J^{-1} \Omega_\sigma) \\ &\quad - 4\mu_3(\Omega'_\sigma J^{-1} \Omega_m + \Omega'_m J^{-1} \Omega_\sigma). \end{aligned} \quad (38)$$

We retrieve the fact that in the absence of location  $m$ ,  $J_\sigma = J/4$  and thus  $\Omega'_\sigma J_\sigma^{-1} \Omega_\sigma = 1$ . In the Gaussian case we also get  $\Omega'_\sigma J_\sigma^{-1} \Omega_\sigma = 1$  (with or without a location).

Finally, (38) writes

$$4\Omega'_\sigma \Sigma \Omega_\sigma + (\kappa_4 - 1)(1 - 8\Omega'_\sigma J^{-1} \Omega_\sigma) - 4\mu_3(\Omega'_\sigma J^{-1} \Omega_m + \Omega'_m J^{-1} \Omega_\sigma) = 0.$$

### D.2. Proof of Corollary 1

Let  $\theta = (\varphi', \vartheta')'$ . We have

$$\begin{aligned} \Omega_m &= \left( E \left( \frac{1}{\sigma_t} \frac{\partial m_t(\varphi_0)}{\partial \varphi'} \right), 0 \right)', \quad \Omega_\sigma = \left( 0, E \left( \frac{1}{\sigma_t} \frac{\partial \sigma_t(\vartheta_0)}{\partial \vartheta'} \right) \right)', \quad J_{m\sigma} = J'_{\sigma m} = 0, \\ J_m &= \begin{pmatrix} K & 0 \\ 0 & 0 \end{pmatrix}, \quad J_\sigma = \begin{pmatrix} 0 & 0 \\ 0 & L \end{pmatrix}, \quad J = \begin{pmatrix} 2K & 0 \\ 0 & 4L \end{pmatrix}, \quad \Sigma = \begin{pmatrix} K^{-1} & 0 \\ 0 & \frac{\kappa_4 - 1}{4} L^{-1} \end{pmatrix}, \end{aligned}$$

where  $\mathbf{K} = E\left(\frac{1}{\sigma_t^2} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}'}\right)$ ,  $\mathbf{L} = E\left(\frac{1}{\sigma_t^2} \frac{\partial \sigma_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \frac{\partial \sigma_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'}\right)$ . Moreover, (38) entails

$$E\left(\frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'}\right) \left\{ E\left(\frac{1}{\sigma_t^2} \frac{\partial \sigma_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}} \frac{\partial \sigma_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}'}\right) \right\}^{-1} E\left(\frac{1}{\sigma_t} \frac{\partial \sigma_t(\boldsymbol{\vartheta}_0)}{\partial \boldsymbol{\vartheta}}\right) = 1.$$

The conclusion follows.  $\square$

### D.3. Proof of Remark 5

We have

$$\boldsymbol{\Omega}_m = \begin{pmatrix} E\left(\frac{X_{t-1}}{\sigma_t}\right) \\ 0 \end{pmatrix}, \quad \boldsymbol{\Omega}_\sigma = \begin{pmatrix} 0 \\ E\left(\frac{X_{t-1}^2}{2\sigma_t^2}\right) \end{pmatrix}, \quad \mathbf{J}_m = \begin{pmatrix} E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right) & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{J}_\sigma = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{4}E\left(\frac{X_{t-1}^4}{\sigma_t^4}\right) \end{pmatrix}.$$

Hence

$$\mathbf{J} = \begin{pmatrix} 2E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right) & 0 \\ 0 & E\left(\frac{X_{t-1}^4}{\sigma_t^4}\right) \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \left\{E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right)\right\}^{-1} & 0 \\ 0 & (\kappa_4 - 1) \left\{E\left(\frac{X_{t-1}^4}{\sigma_t^4}\right)\right\}^{-1} \end{pmatrix},$$

$$\mathbf{H}(x) = f(x) \begin{pmatrix} E\left(\frac{X_{t-1}}{\sigma_t}\right) \\ xE\left(\frac{X_{t-1}^2}{2\sigma_t^2}\right) \end{pmatrix}, \quad \boldsymbol{\Lambda} = \begin{pmatrix} \frac{E\left(\frac{X_{t-1}}{\sigma_t}\right)}{E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right)} & 0 \\ 0 & \frac{E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right)}{E\left(\frac{X_{t-1}^4}{\sigma_t^4}\right)} \end{pmatrix},$$

$$\mathbf{H}(x)' \boldsymbol{\Sigma} \mathbf{H}(x) = f(x)^2 \frac{\left\{E\left(\frac{X_{t-1}}{\sigma_t}\right)\right\}^2}{E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right)} + \{xf(x)\}^2 \frac{\kappa_4 - 1}{4} \frac{\left\{E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right)\right\}^2}{E\left(\frac{X_{t-1}^4}{\sigma_t^4}\right)},$$

$$2\mathbf{H}(x)' \boldsymbol{\Lambda} \boldsymbol{\varrho}(x) = 2f(x) \frac{\left\{E\left(\frac{X_{t-1}}{\sigma_t}\right)\right\}^2}{E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right)} E\{\mathbb{1}_{\eta_0 < x} \eta_0\} + xf(x) \frac{\left\{E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right)\right\}^2}{E\left(\frac{X_{t-1}^4}{\sigma_t^4}\right)} E\{\mathbb{1}_{\eta_0 < x} (\eta_0^2 - 1)\}.$$

Finally, for this model,

$$\begin{aligned} \text{Var}_{as} \left\{ \sqrt{n} \left( \widehat{F}_n(x_n) - F(x_n) \right) \right\} &= F(x) \{1 - F(x)\} + f(x) \frac{\left\{E\left(\frac{X_{t-1}}{\sigma_t}\right)\right\}^2}{E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right)} \{f(x) + 2E\{\mathbb{1}_{\eta_0 < x} \eta_0\}\} \\ &\quad + xf(x) \frac{\left\{E\left(\frac{X_{t-1}^2}{\sigma_t^2}\right)\right\}^2}{E\left(\frac{X_{t-1}^4}{\sigma_t^4}\right)} \left\{ xf(x) \frac{\kappa_4 - 1}{4} + E\{\mathbb{1}_{\eta_0 < x} (\eta_0^2 - 1)\} \right\}. \end{aligned}$$

which is clearly model-dependent. Note that if the distribution of  $\eta_t$  is symmetric we have  $E\left(\frac{X_{t-1}}{\sigma_t}\right) = 0$  and thus the second term in the right-hand side vanishes. If, in addition,  $\alpha_0 \rightarrow 0$ , the asymptotic variance converges to

$$F(x) \{1 - F(x)\} + xf(x) \frac{1 + \phi_0^2}{6\phi_0^2} \left\{ xf(x) \frac{\kappa_4 - 1}{4} + E\{\mathbb{1}_{\eta_0 < x} (\eta_0^2 - 1)\} \right\}.$$

#### D.4. Proof of Corollary 2

Let  $\boldsymbol{\theta} = (\boldsymbol{\varphi}', \boldsymbol{\vartheta}')'$ , where  $\boldsymbol{\varphi} = (\phi_1, \dots, \phi_P, \psi_1, \dots, \psi_Q)'$ ,  $\boldsymbol{\vartheta} = (\omega, \alpha_1, \dots, \alpha_q, \beta_1, \dots, \beta_p)'$ , and  $m_t = m_t(\boldsymbol{\varphi})$ ,  $\sigma_t = \sigma_t(\boldsymbol{\varphi}, \boldsymbol{\vartheta})$ . We know that the asymptotic variance  $\boldsymbol{\Sigma}$  of the QMLE is block-diagonal (see Francq and Zakoian (2019), Remark 7.4). Moreover, we have  $\frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \phi_i} = \Theta(L)^{-1} X_{t-i}$  and  $\frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \psi_j} = -\Theta(L)^{-1} \epsilon_{t-j}$ . Hence  $E\left(\frac{1}{\sigma_t} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}}\right) = 0$  using the fact that  $\sigma_t$  is an even function of the  $\eta_{t-i}, i > 0$ . Since  $\frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\vartheta}} = 0$  we thus have  $\boldsymbol{\Omega}_m = \mathbf{0}$  and it can similarly be shown that  $\boldsymbol{\Omega}_\sigma$  is of the form  $\boldsymbol{\Omega}_\sigma = \frac{1}{2}(\mathbf{0}', \boldsymbol{\omega}'_\sigma)'$ , where  $\boldsymbol{\omega}_\sigma$  is a vector of dimension  $p + q + 1$ . We also have

$$\mathbf{J}_m = \begin{pmatrix} E\left(\frac{1}{\sigma_t^2} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}} \frac{\partial m_t(\boldsymbol{\varphi}_0)}{\partial \boldsymbol{\varphi}'}\right) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}, \quad \mathbf{J}_\sigma = \begin{pmatrix} E\left(\frac{1}{\sigma_t^2} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}} \frac{\partial \sigma_t(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}'}\right) & \mathbf{0} \\ \mathbf{0} & \frac{1}{4}\mathbf{K} \end{pmatrix}$$

where  $\mathbf{K} = E\left(\frac{1}{\sigma_t^4} \frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\vartheta}} \frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\vartheta}'}\right)$ . The block-diagonality of  $\mathbf{J}_\sigma$  follows from the fact that  $\frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\vartheta}}$  is an even function of the  $\eta_{t-i}, i > 0$ , while  $\frac{1}{\sigma_t^2} \frac{\partial \sigma_t^2(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\varphi}}$  is an odd function. Hence, for some matrices  $\mathbf{M}$  and  $\mathbf{N}$  which we do not need to make explicit,

$$\mathbf{J} = \begin{pmatrix} \mathbf{M} & \mathbf{0} \\ \mathbf{0} & \mathbf{K} \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{N} & \mathbf{0} \\ \mathbf{0} & (\kappa_4 - 1)\mathbf{K}^{-1} \end{pmatrix}.$$

Since Assumption **A12\*** is satisfied, (38) holds and can be written  $\boldsymbol{\omega}'_\sigma \mathbf{K}^{-1} \boldsymbol{\omega}_\sigma = 1$ . We also have  $\mathbf{H}(x) = \frac{xf(x)}{2}(\mathbf{0}', \boldsymbol{\omega}'_\sigma)'$ , hence

$$\mathbf{H}(x)' \boldsymbol{\Sigma} \mathbf{H}(x) = \{xf(x)\}^2 \frac{\kappa_4 - 1}{4}, \quad \mathbf{H}(x)' \boldsymbol{\Lambda} \boldsymbol{\varrho}(x) = \frac{xf(x)}{2} E\{\mathbb{1}_{\eta_0 < x}(\eta_0^2 - 1)\}.$$

The conclusion follows.  $\square$

#### E. Proof of the stochastic equicontinuity property in Lemma 9

Noting that  $U_t := F(\eta_t)$  is uniformly distributed on  $[0, 1]$ , we have

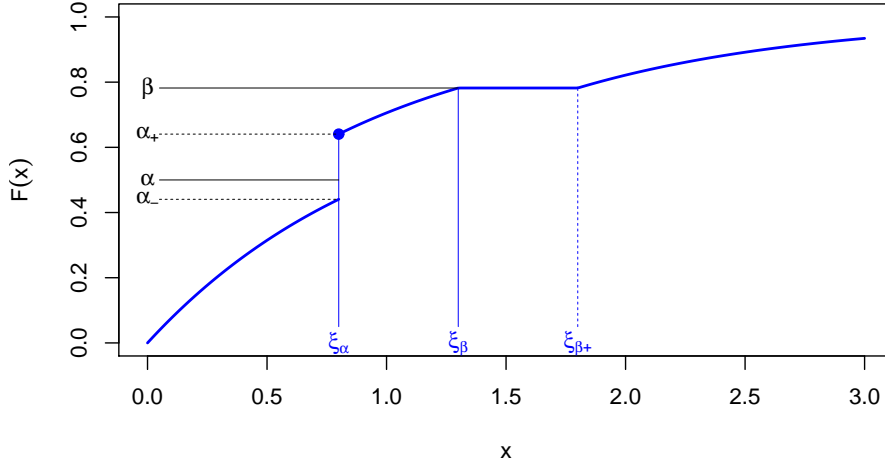
$$e_n(x) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{U_t \leq F(x)} - F(x) = Y_n(F(x)), \quad Y_n(u) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{U_t \leq u} - u.$$

Billingsley (1968) studied the modulus of continuity of  $\{Y_n(u), u \in [0, 1]\}$  and showed in (22.13) that, for each  $\varepsilon > 0$  and  $\eta > 0$ , there exists  $\tau \in (0, 1]$  such that

$$P\left(\sup_{|u-v|<\tau} |Y_n(u) - Y_n(v)| \geq \varepsilon\right) \leq \eta \quad (39)$$

for large  $n$ . For any  $\varepsilon > 0$  and  $\delta > 0$ , we have

$$P(|e_n(x_n) - e_n(x)| \geq \varepsilon) \leq P\left(\sup_{|y-x| \leq \delta} |e_n(y) - e_n(x)| \geq \varepsilon\right) + P(|x_n - x| \geq \delta).$$



**Fig. 3.** Example of cdf.

The last probability tends to zero because  $x_n \rightarrow x$  in probability. Now, note that

$$P\left(\sup_{|y-x|\leq\delta} |e_n(y) - e_n(x)| \geq \varepsilon\right) = P\left(\sup_{|u-v|\leq F(\delta)} |Y_n(u) - Y_n(v)| \geq \varepsilon\right) \leq \eta$$

when  $n$  is large enough to satisfy (39) with  $\tau > F(\delta)$ . Since  $\eta$  can be taken arbitrarily small, the result follows.  $\square$

## F. Complements for Section 4

### F.1. Proof and complements to Corollary 3

The different items introduced in Corollary 3 and Lemma 17 are illustrated in Figure 3.

**Proof of Corollary 3.** The first convergence being a particular case of the second part of the theorem – since under **A13**,  $\xi_\alpha = \xi_\alpha^+$  – it will be sufficient to prove the latter. Let  $\epsilon > 0$ . By definition of  $\xi_\alpha$  and  $\xi_\alpha^+$  we have,

$$F(\xi_\alpha - \epsilon) < \alpha - \delta \quad \text{and} \quad F(\xi_\alpha^+ + \epsilon) > \alpha + \delta, \quad (40)$$

for some  $\delta > 0$ . By Theorem 1, assume  $n$  large enough so that  $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| < \delta$  a.s. It follows from (40) that

$$\xi_\alpha - \epsilon \leq \hat{\xi}_{n,\alpha} \leq \xi_\alpha^+ + \epsilon. \quad (41)$$

Indeed, if  $\xi_\alpha - \epsilon > \hat{\xi}_{n,\alpha}$  then  $F(\xi_\alpha - \epsilon) \geq F(\hat{\xi}_{n,\alpha}) = F(\hat{\xi}_{n,\alpha}) - \hat{F}_n(\hat{\xi}_{n,\alpha}) + \hat{F}_n(\hat{\xi}_{n,\alpha}) \geq \alpha - \delta$  which contradicts the first inequality in (40). Moreover, if  $\xi_\alpha^+ + \epsilon < \hat{\xi}_{n,\alpha}$  then  $F(\xi_\alpha^+ + \epsilon) \leq \hat{F}_n(\xi_\alpha^+ + \epsilon) + \delta \leq$



$\alpha + \delta$  which contradicts the second inequality in (40). The strong convergence of  $\widehat{\xi}_{n,\alpha}$  to the set  $[\xi_\alpha, \xi_\alpha^+]$  follows from (41). The conclusion follows.  $\square$

The asymptotic behaviour of the empirical quantile when the residuals are replaced by innovations is given for the sake of comparison in the following lemma. Having been unable to find a reference for this result, we provide a proof for completeness. Let  $\xi_{n,\alpha}$  the empirical  $\alpha$ -quantile of the innovations  $\eta_1, \dots, \eta_n$ .

LEMMA 17. *For the iid sequence  $(\eta_t)$  and  $\alpha \in (0, 1)$  we have:*

(a) *If  $\xi_\alpha^+ \neq \xi_\alpha$ , then  $\liminf \xi_{n,\alpha} = \xi_\alpha$ ,  $\limsup \xi_{n,\alpha} = \xi_\alpha^+$ , a.s.*

*Moreover,  $P(\xi_{n,\alpha} \leq \xi_\alpha) = 1 - P(\xi_{n,\alpha} \geq \xi_\alpha^+) \rightarrow 1/2$  as  $n \rightarrow \infty$ .*

(b) *If  $\xi_\alpha^+ = \xi_\alpha$  (Assumption **A13**), then  $\xi_{n,\alpha} \rightarrow \xi_\alpha$  a.s. Moreover, letting  $\alpha_- = P(\eta_t < \xi_\alpha)$  and  $\alpha_+ = P(\eta_t \leq \xi_\alpha)$ :*

(i) *If  $\alpha > \alpha_-$ :  $P(\xi_{n,\alpha} < \xi_\alpha) \rightarrow 0$  as  $n \rightarrow \infty$ .*

(ii) *If  $\alpha < \alpha_+$ :  $P(\xi_{n,\alpha} \leq \xi_\alpha) \rightarrow 1$  as  $n \rightarrow \infty$ .*

(iii) *If  $\alpha = \alpha_-$ :  $P(\xi_{n,\alpha} < \xi_\alpha) \rightarrow 1/2$  as  $n \rightarrow \infty$ .*

(iv) *If  $\alpha = \alpha_+$ :  $P(\xi_{n,\alpha} \leq \xi_\alpha) \rightarrow 1/2$  as  $n \rightarrow \infty$ .*

**Proof.** The argument used in the proof of Corollary 3 shows that  $\xi_\alpha - \epsilon \leq \xi_{n,\alpha} \leq \xi_\alpha^+ + \epsilon$ , noting that the usual Glivenko-Cantelli lemma does not require assumption **A5**. Because there are no  $\eta_t$ 's in the interval  $(\xi_\alpha, \xi_\alpha^+)$ , it follows that, for sufficiently large  $n$ ,

$$\xi_\alpha - \epsilon \leq \xi_{n,\alpha} \leq \xi_\alpha \quad \text{or} \quad \xi_\alpha^+ \leq \xi_{n,\alpha} \leq \xi_\alpha^+ + \epsilon. \quad (42)$$

If  $\xi_\alpha^+ \neq \xi_\alpha$ , the last inequalities entail that  $\{\liminf \xi_{n,\alpha}, \limsup \xi_{n,\alpha}\} \subset \{\xi_\alpha, \xi_\alpha^+\}$ , a.s. Note that  $\xi_{n,\alpha} \geq z$  iff  $\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{\eta_t < z} < \alpha$ . Moreover  $P(\eta_t < \xi_\alpha^+) = P(\eta_t \leq \xi_\alpha) = F(\xi_\alpha) \geq \alpha$ . Because  $\xi_\alpha^+ > \xi_\alpha$ , we actually have  $F(\xi_\alpha) = \alpha$ . Therefore

$$P(\xi_{n,\alpha} \geq \xi_\alpha^+) = P\left(\frac{1}{n} \sum_{t=1}^n \mathbf{1}_{\eta_t < \xi_\alpha^+} < \alpha\right) = P(Z_n < 0),$$

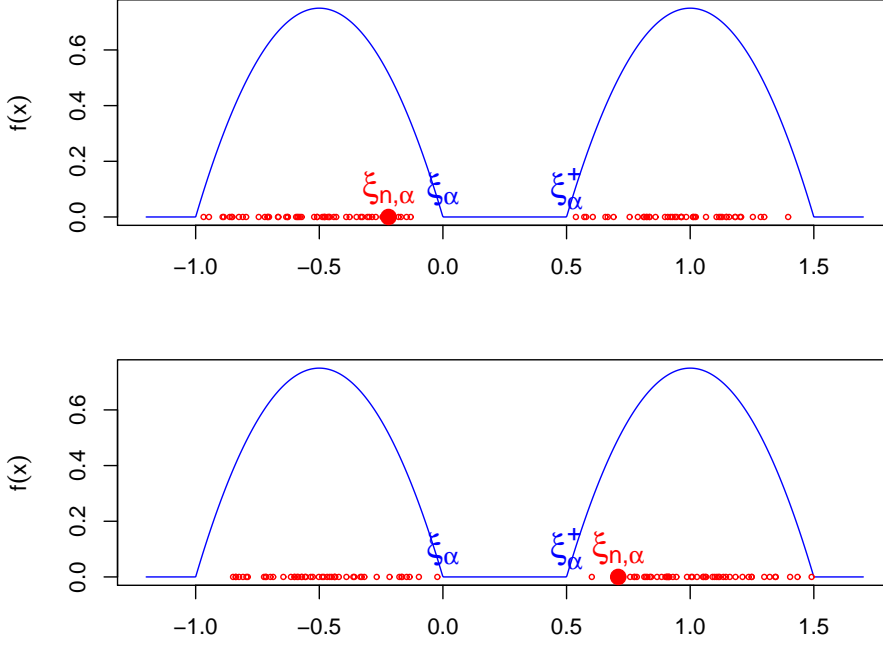
where

$$Z_n = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\mathbf{1}_{\eta_t < \xi_\alpha^+} - \alpha}{\sqrt{\alpha(1-\alpha)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1).$$

The conclusion follows in case 1.

Now, if  $\xi_\alpha^+ = \xi_\alpha$ , the strong consistency of  $\xi_{n,\alpha}$  follows from (42). To show the results in (a)-(d), the previous arguments can be adapted: for instance we have

$$P(\xi_{n,\alpha} \geq \xi_\alpha) = P\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\mathbf{1}_{\eta_t < \xi_\alpha} - \alpha_-}{\sqrt{\alpha_-(1-\alpha_-)}} < \sqrt{n} \frac{\alpha - \alpha_-}{\sqrt{\alpha_-(1-\alpha_-)}}\right),$$



**Fig. 4.** For  $\alpha = 1/2$ , computations of  $\xi_{n,\alpha}$  for two simulations of length 101 of a density which does not satisfy **A13**.

from which (a) and (c) can be deduced. The arguments for (b) and (d) are similar, using the characterization:  $\xi_{n,\alpha} \leq z$  iff  $\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\eta_t \leq z} \geq \alpha$ .  $\square$

The first part of this lemma shows that, for iid observations, (7) is replaced by

$$[\liminf \xi_{n,\alpha}, \limsup \xi_{n,\alpha}] = [\xi_\alpha, \xi_\alpha^+] \quad a.s.$$

In particular, the empirical quantile does not converge when Assumption **A13** is not satisfied, as illustrated in Figure 4. The second part shows, in particular, that if  $\alpha_- < \alpha < \alpha_+$  then  $P(\xi_{n,\alpha} = \xi_\alpha)$  tends to 1. On the other hand, when  $\alpha_- = \alpha_+$  (no mass at  $\xi_\alpha$ ) then  $P(\xi_{n,\alpha} = \xi_\alpha)$  tends to 0.

When **A13** does not hold, not only  $\xi_{n,\alpha}$  but also the sample quantile  $\hat{\xi}_{n,\alpha}$  of the residuals may not converge to  $\xi_\alpha$ . For instance, in the constant scale model,  $\epsilon_t = \sigma \eta_t$ , we have  $\hat{\eta}_t = \frac{\sigma}{\hat{\sigma}} \eta_t$  where the ratio converges to 1 and thus both  $\xi_{n,\alpha}$  and  $\hat{\xi}_{n,\alpha} = \frac{\sigma}{\hat{\sigma}} \xi_{n,\alpha}$  do not converge.

## References

- Andrews, D. W.** (1994) Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43–72.
- Bai, J.** (1994) Weak convergence of the sequential empirical processes of residuals in ARMA Models. *The Annals of Statistics* 22, 2051–2061.
- Berkes, I. and L. Horváth** (2002). Empirical processes of residuals. In *Empirical process techniques for dependent data* (pp. 195–209). Birkhäuser, Boston, MA.
- Berkes, I. and L. Horváth** (2003) Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Processes and their Applications* 105, 271–298.
- Billingsley, P.** (1961) The Lindeberg-Lévy theorem for martingales. *Proceedings of the American Mathematical Society* 12, 788–792.
- Billingsley, P.** (1968) *Convergence of probability measures*. Wiley, New York.
- Boldin, M. V.** (1982) Estimation of the distribution of noise in an autoregressive scheme. *Theory of Probability and Applications* 27 866–871.
- Boldin, M. V.** (1998) On residual empirical distribution functions in ARCH models with applications to testing and estimation. *Mitt. Math. Sem. Giessen* 235, 49–66.
- Burkholder, D.L.** (1973) Distribution function inequalities for martingales. *The Annals of Probability* 1, 19–42.
- Feigin, P.D. and R.L. Tweedie** (1985) Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis* 6, 1–14.
- Francq, C. and J.-M. Zakoïan** (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10, 605–637.
- Francq, C. and J.-M. Zakoïan** (2013a) Optimal predictions of powers of conditionally heteroscedastic processes. *Journal of the Royal Statistical Society: Series B* 75, 345–367.
- Francq, C. and J.-M. Zakoïan** (2013b) Inference in nonstationary asymmetric GARCH models. *The Annals of Statistics* 41, 1970–1998.
- Francq, C. and J.-M. Zakoïan** (2015) Risk parameter estimation in volatility models. *Journal of Econometrics* 184, 158–173.

- Francq, C. and J.-M. Zakořan** (2019) *GARCH Models: Structure, Statistical Inference and Financial Applications*. 2nd Edition. John Wiley.
- Hall, P., and C.C. Heyde** (1980) *Martingale limit theory and its application*. Academic press, New York.
- Hallin, M., Paindaveine, D. and M. Šiman** (2010) Multivariate quantiles and multiple-output regression quantiles: From  $L_1$  optimization to halfspace depth. *The Annals of Statistics* 38, 635–669.
- Hoga, Y.** (2019) Confidence intervals for conditional tail risk measures in ARMA-GARCH models. *Journal of Business and Economic Statistics* 37, 613–624.
- Horváth, L., Kokoszka, P. and G. Teyssière** (2001) Empirical process of the squared residuals of an ARCH sequence. *The Annals of Statistics* 29, 445–469.
- Koul, H. L., and S. Ling** (2006) Fitting an error distribution in some heteroscedastic time series models. *The Annals of Statistics* 34, 994–1012.
- Kreiss, J. P.** (1987). On adaptive estimation in stationary ARMA processes. *The Annals of Statistics* 15, 112–133.
- Kreiss, J. P.** (1991). Estimation of the distribution function of noise in stationary processes. *Metrika* 38, 28–297.
- Kulperger, R., and H. Yu** (2005) High moment partial sum processes of residuals in GARCH models and their applications. *The Annals of Statistics* 33, 2395–2422.
- Lee, S. and M. Taniguchi** (2005) Asymptotic theory for ARCH-SM models: LAN and residual empirical processes. *Statistica Sinica* 15, 215–234.
- Ling, S.** (1998) Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models. *The Annals of Statistics* 26, 741–754.
- Ling, S.** (2007) A double AR( $p$ ) model: structure and estimation. *Statistica Sinica* 17, 161–175.
- Robinson, P.M.** Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55, 875–891.
- Shorack, G.R. and J.A. Wellner** (1986) *Empirical processes with applications to statistics*. Wiley, New-York.

- Stute, W.** (2001) Residual analysis for ARCH( $p$ )-time series. *Test* 10, 393–403.
- Stute, W. and G. Schumann** (1980) A general Glivenko-Cantelli theorem for stationary sequences of random observations. *Scandinavian Journal of Statistics* 7, 102–104.
- Wang, G., Zhu, K., Li, G. and W. K. Li** (2019) Hybrid quantile estimation for asymmetric power GARCH models. arXiv:1911.09343.
- Zheng, Y., Zhu, Q., Li, G., and Z. Xiao** (2018) Hybrid quantile regression estimation for time series models with conditional heteroscedasticity. *Journal of the Royal Statistical Society: Series B* 80, 975–993.
- Zhu, Q. and G. Li** (2019) Quantile double autoregression. Available at SSRN 3335753.