

# A Residual Bootstrap for Conditional Value-at-Risk

Eric Beutner\*

Alexander Heinemann\*\*

Stephan Smeekes\*\*\*

December 15, 2024

## Abstract

A fixed-design residual bootstrap method is proposed for the two-step estimator of Francq and Zakoïan (2015) associated with the conditional Value-at-Risk. The bootstrap's consistency is proven for a general class of volatility models and intervals are constructed for the conditional Value-at-Risk. A simulation study reveals that the equal-tailed percentile bootstrap interval tends to fall short of its nominal value. In contrast, the reversed-tails bootstrap interval yields accurate coverage. We also compare the theoretically analyzed fixed-design bootstrap with the recursive-design bootstrap. It turns out that the fixed-design bootstrap performs equally well in terms of average coverage, yet leads on average to shorter intervals in smaller samples. An empirical application illustrates the interval estimation.

**Key words:** Residual bootstrap; Value-at-Risk; GARCH

**JEL codes:** C14; C15; C58

---

\*Department of Econometrics and Data Science, Vrije Universiteit Amsterdam, De Boelelaan 1105 1081 HV Amsterdam, Netherlands. E-mail address: e.a.beutner@vu.nl

\*\*Department of Quantitative Economics, Maastricht University, Tongersestraat 53, 6211 LM Maastricht, Netherlands. E-mail address: a.heinemann@alumni.maastrichtuniversity.nl (corresponding author)

\*\*\*Department of Quantitative Economics, Maastricht University, Tongersestraat 53, 6211 LM Maastricht, Netherlands. E-mail address: s.smeekes@maastrichtuniversity.nl

The authors thank Franz Palm, Hanno Reuvers, Jean-Michel Zakoïan and Christian Francq for useful comments and suggestions as well as Dewi Peerlings and Benoit Duvocelle for computational support. This research was financially supported by the Netherlands Organisation for Scientific Research (NWO).

# 1 Introduction

Risk management has tremendously developed in past decades becoming an increasing practice. With minimum capital requirements being enforced by current legislation (Basel III and Solvency II), financial institutions and insurance companies monitor risk by using conventional measures such as Value-at-Risk (VaR). Typically, the volatility dynamics are specified by a (semi-)parametric model leading to conditional risk measure versions. For GARCH-type models the conditional VaR reduces to the conditional volatility scaled by a quantile of the innovations' distribution. The latter is conventionally treated as additional parameter and forms together with the others the *risk parameter* (Francq and Zakoïan, 2015). The true parameters are generally unknown and need to be estimated to obtain an estimate for the conditional VaR. Clearly, this VaR evaluation is subject to estimation risk that needs to be quantified for appropriate risk management.

Whereas an estimator based on a single step is available after re-parameterization (Francq and Zakoïan, 2015), a widely used approach is the following two-step estimation procedure. First, the parameters of the stochastic volatility model are estimated. Arguably the most popular estimation method in a GARCH-type setting is the Gaussian quasi-maximum-likelihood (QML) method. Based on the model's residuals the quantile is estimated by its empirical counterpart in a second step. For realistic sample sizes (e.g. 500 or 1,000 daily observations) the estimators are subject to considerable estimation risk. In particular, the estimation uncertainty associated with the quantile estimator is substantial for extreme quantiles (e.g.  $\leq 5\%$ ).

To quantify the uncertainty around the point estimators, one traditionally relies on asymptotic theory while replacing the unknown quantities in the limiting distri-

bution by consistent estimates. An alternative approach – frequently employed in practice – is based on a bootstrap approximation. Regarding the estimators of the GARCH parameters, various bootstrap methods have been studied to approximate the estimators’ finite sample distribution including the subsample bootstrap (Hall and Yao, 2003), the block bootstrap (Corradi and Iglesias, 2008), the wild bootstrap (Shimizu, 2009) and the residual bootstrap. The residual bootstrap method is particularly popular and can be further divided into recursive (Pascual et al., 2006; Hidalgo and Zaffaroni, 2007; Jeong, 2017) and fixed (Shimizu, 2009; Cavaliere et al., 2018) design. Whereas in the former the bootstrap observations are generated recursively using the estimated volatility dynamics, the latter design keeps the dynamics of the bootstrap samples fixed at the value of the original series.

The estimation of the quantile and the conditional VaR have received only selected attention in the bootstrap literature and proposed bootstrap methods have been, to the best of our knowledge, exclusively investigated by means of simulation. Christoffersen and Gonçalves (2005) examine various quantile estimators and construct intervals for the conditional VaR using a recursive-design residual bootstrap method. In addition, Hartz et al. (2006) presume the innovation distribution to be standard normal such that the quantile parameter is known; they propose a resampling method based on a residual bootstrap and a bias-correction step to account for deviations from the normality assumption. In contrast, Spierdijk (2016) develops an  $m$ -out-of- $n$  without-replacement bootstrap to construct confidence intervals for ARMA-GARCH VaR.

This paper proposes a fixed-design residual bootstrap method to mimic the finite sample distribution of the two-step estimator and provides an algorithm for the construction of bootstrap intervals for the conditional VaR. The proposed bootstrap

method is proven to be consistent for a general class of volatility models. In particular, our framework does not only encompass GARCH but also several GARCH extensions such as the threshold GARCH (T-GARCH) of Zakoïan (1994) and the GJR-GARCH named after Glosten, Jagannathan and Runkle (1993). The bootstrap consistency is established under a set of mild assumptions, which relaxes moment conditions on the innovations imposed in the GARCH bootstrap literature. To the best of our knowledge this paper is the first to theoretically validate the residual bootstrap for the quantile and the conditional VaR.

The remainder of the paper is organized as follows. Section 2 specifies the model and the conditional VaR is derived. The two-step estimation procedure is described in Section 3 and asymptotic theory is provided under mild assumptions. In Section 4, a fixed-design residual bootstrap method is proposed and proven to be consistent. Further, bootstrap intervals are constructed for the conditional VaR and bootstrap extensions discussed. A simulation study is conducted in Section 5 and an empirical application illustrates the interval estimation based on the fixed-design residual bootstrap. Section 6 concludes and auxiliary results are gathered in the Appendix. Appendix A contains lemmas and their proofs while Appendix B is devoted to the related recursive-design residual bootstrap.

## 2 Model

We consider a conditional volatility model of the form

$$\epsilon_t = \sigma_t \eta_t \tag{2.1}$$

with  $t \in \mathbb{Z}$ , where  $\{\epsilon_t\}$  denotes the sequence of log-returns,  $\{\sigma_t\}$  is a volatility process and  $\{\eta_t\}$  is a sequence of independent and identically distributed (iid) variables satisfying  $\mathbb{E}[\eta_t^2] = 1$ . The volatility is presumed to be a measurable function of past observations

$$\sigma_t = \sigma_t(\theta_0) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) \quad (2.2)$$

with  $\sigma : \mathbb{R}^\infty \times \Theta \rightarrow (0, \infty)$  and  $\theta_0$  denotes the true parameter vector belonging to the parameter space  $\Theta \subset \mathbb{R}^r$ ,  $r \in \mathbb{N}$ . Subsequently, we consider two examples for the functional form of (2.2): the well-known GARCH model (Engle, 1982; Bollerslev, 1986) and the T-GARCH model of Zakoïan (1994). Whereas the first is frequently applied in practice, the second is motivated by our empirical application (see Section 5.2).

**Example 1.** Suppose  $\{\epsilon_t\}$  follows a GARCH(1, 1) process given by (2.1) and  $\sigma_t^2 = \omega_0 + \alpha_0 \epsilon_{t-1}^2 + \beta_0 \sigma_{t-1}^2$ , where  $\theta_0 = (\omega_0, \alpha_0, \beta_0)' \in (0, \infty) \times [0, \infty) \times [0, 1)$ . The recursive structure implies  $\sigma_t = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) = \sqrt{\sum_{k=1}^{\infty} \beta_0^{k-1} (\omega_0 + \alpha_0 \epsilon_{t-k}^2)}$ .

**Example 2.** Suppose  $\{\epsilon_t\}$  follows a T-GARCH(1, 1) process given by (2.1) and  $\sigma_t = \omega_0 + \alpha_0^+ \epsilon_{t-1}^+ + \alpha_0^- \epsilon_{t-1}^- + \beta_0 \sigma_{t-1}$  with parameters  $\theta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0)' \in (0, \infty) \times [0, \infty) \times [0, \infty) \times [0, 1)$  and  $\epsilon_t^+ = \max\{\epsilon_t, 0\}$  and  $\epsilon_t^- = \max\{-\epsilon_t, 0\}$ . The model's recursive structure yields  $\sigma_t = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \theta_0) = \sum_{k=1}^{\infty} \beta_0^{k-1} (\omega_0 + \alpha_0^+ \epsilon_{t-k}^+ + \alpha_0^- \epsilon_{t-k}^-)$ .

Throughout the paper, for any cumulative distribution function (cdf), say  $G$ , we define the generalized inverse by  $G^{-1}(u) = \inf \{\tau \in \mathbb{R} : G(\tau) \geq u\}$  and write  $G(\cdot-)$  to denote its left limit. Generally, for an arbitrary real-valued random variable  $X$  (e.g. stock return) with cdf  $F_X$ , the VaR at level  $\alpha \in (0, 1)$ , is given by  $VaR_\alpha(X) = -F_X^{-1}(\alpha)$ . Let  $\mathcal{F}_n$  denote the  $\sigma$ -algebra generated by  $\{\epsilon_t, t \leq n\}$ . It fol-

lows that the conditional VaR of  $\epsilon_{n+1}$  given  $\mathcal{F}_n$  at level  $\alpha \in (0, 1)$  is  $VaR_\alpha(\epsilon_{n+1}|\mathcal{F}_n) = \sigma(\epsilon_n, \epsilon_{n-1}, \dots; \theta_0) VaR_\alpha(\eta_{n+1})$ . For given  $\alpha$ , the quantile of  $\eta_{n+1}$  is constant and can be treated as a parameter. Thus, denoting the cdf of the  $\eta_t$ 's by  $F$  and setting  $\xi_\alpha = F^{-1}(\alpha)$ , the conditional VaR of  $\epsilon_{n+1}$  given  $\mathcal{F}_n$  at level  $\alpha$  reduces to

$$VaR_\alpha(\epsilon_{n+1}|\mathcal{F}_n) = -\xi_\alpha \sigma_{n+1}(\theta_0). \quad (2.3)$$

Typically,  $\alpha$  is fixed at a sufficiently small level such that  $\xi_\alpha < 0$ . Except for special cases (e.g. normality of  $\eta_t$ ),  $\xi_\alpha$  is unknown and needs to be estimated just like  $\theta_0$ .

### 3 Estimation

We estimate the parameters  $\theta_0$  and  $\xi_\alpha$  following the two-step procedure of Francq and Zakoïan (2015, Section 4.2). In the first step, we estimate the conditional volatility parameter  $\theta_0$  by Gaussian QML. This approach is motivated as follows: if the innovations  $\{\eta_t\}$  were Gaussian, the variables  $\eta_t(\theta) = \epsilon_t/\sigma_t(\theta)$  would be iid  $N(0, 1)$  whenever  $\theta = \theta_0$ , where

$$\sigma_t(\theta) = \sigma(\epsilon_{t-1}, \dots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \dots; \theta). \quad (3.1)$$

The 'Q' in QML stands for 'quasi' and refers to the fact that  $F$  does not need to be the standard normal distribution function. Obviously, given a sample  $\epsilon_1, \dots, \epsilon_n$ , we generally cannot determine  $\sigma_t(\theta)$  completely. Replacing the unknown presample observations by arbitrary values, say  $\tilde{\epsilon}_t$ ,  $t \leq 0$ , we obtain

$$\tilde{\sigma}_t(\theta) = \sigma(\epsilon_{t-1}, \dots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \theta), \quad (3.2)$$

which serves as an approximation for (3.1). The QML estimator of  $\theta_0$  is defined by

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta} \tilde{L}_n(\theta) \quad (3.3)$$

with the criterion function specified by

$$\tilde{L}_n(\theta) = \frac{1}{n} \sum_{t=1}^n \tilde{\ell}_t(\theta) \quad \text{and} \quad \tilde{\ell}_t(\theta) = -\frac{1}{2} \left( \frac{\epsilon_t}{\tilde{\sigma}_t(\theta)} \right)^2 - \log \tilde{\sigma}_t(\theta).$$

In the second step, we estimate  $\xi_\alpha$  on the basis of the first-step residuals, i.e.  $\hat{\eta}_t = \epsilon_t / \tilde{\sigma}_t(\hat{\theta}_n)$ . The empirical  $\alpha$ -quantile of  $\hat{\eta}_1, \dots, \hat{\eta}_n$  is given by

$$\hat{\xi}_{n,\alpha} = \arg \min_{z \in \mathbb{R}} \frac{1}{n} \sum_{t=1}^n \rho_\alpha(\hat{\eta}_t - z), \quad (3.4)$$

where  $\rho_\alpha(u) = u(\alpha - \mathbb{1}_{\{u < 0\}})$  is the usual asymmetric absolute loss function (cf. Koenker and Xiao, 2006). Equivalently, we can write  $\hat{\xi}_{n,\alpha} = \hat{\mathbb{F}}_n^{-1}(\alpha)$  with  $\hat{\mathbb{F}}_n(x) = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t \leq x\}}$  being the empirical distribution function (edf) of the residuals.

Having obtained estimators for  $\theta_0$  and  $\xi_\alpha$ , we turn to the estimation of the conditional VaR of the one-period ahead observation at level  $\alpha$ . Whereas the notation  $VaR_\alpha(\epsilon_{n+1} | \mathcal{F}_n)$  stresses the object's conditional nature, we henceforth proceed with the abbreviation  $VaR_{n,\alpha}$  for notational convenience. Employing (3.2) – (3.4) we can estimate  $VaR_{n,\alpha}$  by

$$\widehat{VaR}_{n,\alpha} = -\hat{\xi}_{n,\alpha} \tilde{\sigma}_{n+1}(\hat{\theta}_n). \quad (3.5)$$

Clearly, the estimator's large sample properties cannot be studied using traditional tools such as consistency since (3.5) does not permit a limit.

For the subsequent asymptotic analysis, we introduce the following assumptions.

**Assumption 1.** (Compactness)  $\Theta$  is a compact subset of  $\mathbb{R}^r$ .

**Assumption 2.** (Stationarity & Ergodicity)  $\{\epsilon_t\}$  is a strictly stationary and ergodic solution of (2.1) with (2.2).

**Assumption 3.** (Volatility process) The function  $\sigma : \mathbb{R}^\infty \times \Theta \rightarrow (0, \infty)$  is known and for any real sequence  $\{x_i\}$ , the function  $\theta \rightarrow \sigma(x_1, x_2, \dots; \theta)$  is continuous. Almost surely,  $\sigma_t(\theta) > \underline{\omega}$  for any  $\theta \in \Theta$  and some  $\underline{\omega} > 0$  and  $\mathbb{E}[\sigma_t^s(\theta_0)] < \infty$  for some  $s > 0$ . Moreover, for any  $\theta \in \Theta$ , we assume  $\sigma_t(\theta_0)/\sigma_t(\theta) = 1$  almost surely (a.s.) if and only if  $\theta = \theta_0$ .

**Assumption 4.** (Initial conditions) There exists a constant  $\rho \in (0, 1)$  and a random variable  $C_1$  measurable with respect to  $\mathcal{F}_0$  and  $\mathbb{E}[|C_1|^s] < \infty$  for some  $s > 0$  such that

$$(i) \sup_{\theta \in \Theta} |\sigma_t(\theta) - \tilde{\sigma}_t(\theta)| \leq C_1 \rho^t;$$

(ii)  $\theta \rightarrow \sigma(x_1, x_2, \dots; \theta)$  has continuous second-order derivatives satisfying

$$\sup_{\theta \in \Theta} \left\| \frac{\partial \sigma_t(\theta)}{\partial \theta} - \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} \right\| \leq C_1 \rho^t, \quad \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq C_1 \rho^t,$$

where  $\|\cdot\|$  denotes the Euclidean norm.

**Assumption 5.** (Innovation process) The innovations  $\{\eta_t\}$  satisfy

(i)  $\eta_t \stackrel{iid}{\sim} F$  with  $F$  being continuous,  $\mathbb{E}[\eta_t^2] = 1$  and  $\eta_t$  is independent of  $\{\epsilon_u : u < t\}$ ;

(ii)  $\eta_t$  admits a density  $f$  which is continuous and strictly positive around  $\xi_\alpha < 0$ ;

(iii)  $\mathbb{E}[\eta_t^4] < \infty$ .

**Assumption 6.** (Interior)  $\theta_0$  belongs to the interior of  $\Theta$  denoted by  $\mathring{\Theta}$ .



**Assumption 7.** (Non-degeneracy) There does not exist a non-zero  $\lambda \in \mathbb{R}^r$  such that  $\lambda' \frac{\partial \sigma_t(\theta_0)}{\partial \theta} = 0$  a.s.

**Assumption 8.** (Monotonicity) For any real sequence  $\{x_i\}$  and for any  $\theta_1, \theta_2 \in \Theta$  satisfying  $\theta_1 \leq \theta_2$  componentwise, we have  $\sigma(x_1, x_2, \dots; \theta_1) \leq \sigma(x_1, x_2, \dots; \theta_2)$ .

**Assumption 9.** (Moments) There exists a neighborhood  $\mathcal{V}(\theta_0)$  of  $\theta_0$  such that the following variables have finite expectation

$$(i) \sup_{\theta \in \mathcal{V}(\theta_0)} \left| \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \right|^a, \quad (ii) \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\|^b, \quad (iii) \sup_{\theta \in \mathcal{V}(\theta_0)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\|^c$$

for some  $a, b, c$  (to be specified).<sup>1</sup>

**Assumption 10.** (Scaling Stability) There exists a function  $g$  such that for any  $\theta \in \Theta$ , for any  $\lambda > 0$ , and any real sequence  $\{x_i\}$

$$\lambda \sigma(x_1, x_2, \dots; \theta) = \sigma(x_1, x_2, \dots; \theta_\lambda),$$

where  $\theta_\lambda = g(\theta, \lambda)$  and  $g$  is differentiable in  $\lambda$ .

The previous set of assumptions is comparable to the conditions imposed by Francq and Zakoïan (2015). Assumption 3 calls for a correct specification of the volatility structure. If the researcher incorrectly specifies a volatility function  $\varsigma(\dots; \vartheta)$  instead, the estimator of the misspecified conditional volatility model  $\hat{\vartheta}_n$  will converge to a pseudo-true value, i.e.  $\vartheta_0 = \arg \min_{\vartheta} \mathbb{E} \left[ \frac{1}{2} \frac{\epsilon_t^2}{\varsigma_t^2(\vartheta)} + \log \varsigma_t(\vartheta) \right]$ . The corresponding edf of the residuals  $\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\epsilon_t / \varsigma_t(\hat{\vartheta}_n) \leq x\}}$  converges to  $\bar{F}(x) = \mathbb{E} \left[ F \left( x \frac{\varsigma_t(\vartheta_0)}{\varsigma_t(\theta_0)} \right) \right]$  in view of Lemma 1 while the  $\alpha$ -quantile estimator converges to  $\bar{F}^{-1}(\alpha)$ , which is generally different from  $F^{-1}(\alpha)$ . Thus, the correct specification of the volatility function is

---

<sup>1</sup>Note that the variables in (i)–(iii) are strictly stationary (Francq and Zakoïan, 2011, p. 181/406).

crucial and one can test for it using the recently developed test by Jiménez-Gamero et al. (2019); for further recent results on goodness-of-fit testing for GARCH models see Bardet et al. (2020). Regarding the innovation process we do not need to assume  $\mathbb{E}[\eta_t] = 0$  (cf. Francq and Zakoïan, 2004, Remark 2.5). The iid condition in Assumption 5(i) is vital for (2.3) to hold and is the basis of the residual bootstrap in Section 4.1. Under correct specification of the volatility process the iid assumption imposed on the innovations can be tested for by considering the errors  $\epsilon_t/\sigma(\epsilon_{t-1}, \epsilon_{t-2}, \dots; \hat{\theta}_n)$ ,  $t = 1, \dots, n$  and applying the test of Cho and White (2011). Whereas Cavaliere et al. (2018) assume the existence of the sixth moment of  $\eta_t$  for the fixed-design bootstrap in ARCH( $q$ ) models, we only require the fourth moment to be finite in Assumption 5(iii). In Assumption 8 the function  $\sigma(x_1, x_2, \dots; \theta)$  is presumed to be monotonically increasing in  $\theta$ , which is used to establish the strong consistency of the quantile estimator. While the monotonicity condition is a feature shared by various stochastic volatility models (cf. Berkes and Horváth, 2003, Lemma 4.1), it excludes the exponential GARCH (Nelson, 1991) and the log-GARCH (Geweke, 1986; Pantula, 1986). Further, we require higher order of moments in Assumption 9 for the bootstrap, which does not seem to be restrictive for the classical GARCH-type models (cf. Francq and Zakoïan, 2011, p. 165; Hamadeh and Zakoïan, 2011, p. 501). In particular, Assumption 9 is presumed to hold with  $a = \pm 12$ ,  $b = 12$  and  $c = 6$  for establishing the convergence of the bootstrap information matrix.

On the basis of the previous assumptions we extend the strong consistency result of Francq and Zakoïan (2015, Theorem 1) to the quantile estimator.

**Theorem 1.** *(Strong Consistency) Under Assumptions 1–3, 4(i) and 5(i) the estimator in (3.3) is strongly consistent, i.e.  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ . If in addition Assumptions 6 and 9(i) hold with  $a = -1$ , then the estimator in (3.4) satisfies  $\hat{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha$ .*

*Proof.* Francq and Zakoïan (2015, Theorem 1) establish  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ . The second claim follows from  $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \xrightarrow{a.s.} 0$  (Lemma 1 in Appendix A.1) and van der Vaart (2000, Theorem 21.2).  $\square$

To lighten notation, we henceforth write  $D_t(\theta) = \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta}$  and drop the argument when evaluated at the true parameter, i.e.  $D_t = D_t(\theta_0)$ . The next result provides the joint asymptotic distribution of  $\hat{\theta}_n$  and  $\hat{\xi}_{n,\alpha}$  and is due to Francq and Zakoïan (2015).

**Theorem 2.** (*Asymptotic Distribution*) Suppose Assumptions 1–7, 9 and 10 hold with  $a = b = 4$  and  $c = 2$ . Then, we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n - \theta_0) \\ \sqrt{n}(\xi_\alpha - \hat{\xi}_{n,\alpha}) \end{pmatrix} \xrightarrow{d} N(0, \Sigma_\alpha) \quad \text{with} \quad \Sigma_\alpha = \begin{pmatrix} \frac{\kappa-1}{4} J^{-1} & \lambda_\alpha J^{-1} \Omega \\ \lambda_\alpha \Omega' J^{-1} & \zeta_\alpha \end{pmatrix}, \quad (3.6)$$

where  $\kappa = \mathbb{E}[\eta_t^4]$ ,  $\Omega = \mathbb{E}[D_t]$ ,  $J = \mathbb{E}[D_t D_t']$ ,  $\lambda_\alpha = \xi_\alpha^{\frac{\kappa-1}{4}} + \frac{p_\alpha}{2f(\xi_\alpha)}$ ,  $\zeta_\alpha = \xi_\alpha^2 \frac{\kappa-1}{4} + \frac{\xi_\alpha p_\alpha}{f(\xi_\alpha)} + \frac{\alpha(1-\alpha)}{f^2(\xi_\alpha)}$  and  $p_\alpha = \mathbb{E}[\eta_t^2 \mathbb{1}_{\{\eta_t < \xi_\alpha\}}] - \alpha$ .

*Proof.* See Francq and Zakoïan (2015, Theorem 4) and note that Assumption 10 is needed to ensure  $\Omega' J^{-1} \Omega = 1$ .  $\square$

*Remark 1.* It is worth mentioning that the asymptotics in this theorem for  $\hat{\xi}_{n,\alpha}$  are for  $\alpha$  fixed while  $n$  goes to infinity. If, for instance,  $\alpha$  is very small for moderate  $n$  the distribution in the following theorem might not provide a good approximation. For such cases, approximations based on extreme value theory may provide better approximations to the unknown finite sample distribution.

In a GARCH( $p, q$ ) setting Gao and Song (2008) quantify the uncertainty around  $\hat{\theta}_n$  and  $\hat{\xi}_{n,\alpha}$  using (3.6) while replacing the unknown quantities in  $\Sigma_\alpha$  by consistent estimates. In this spirit  $\xi_\alpha$  can be substituted by  $\hat{\xi}_{n,\alpha}$  and  $\Omega, J, \kappa$  and  $p_\alpha$  can be

replaced by

$$\begin{aligned}\hat{\Omega}_n &= \frac{1}{n} \sum_{t=1}^n \hat{D}_t, & \hat{J}_n &= \frac{1}{n} \sum_{t=1}^n \hat{D}_t \hat{D}_t', \\ \hat{\kappa}_n &= \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^4, & \hat{p}_{n,\alpha} &= \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^2 \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}} - \alpha,\end{aligned}\tag{3.7}$$

respectively, with  $\hat{D}_t = \tilde{D}_t(\hat{\theta}_n)$  and  $\tilde{D}_t(\theta) = \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta}$ . The strong consistency of the estimators in (3.7) follow from Theorem 1 and Lemma 2 in Appendix A.1. Moreover, kernel smoothing is commonly employed to estimate the density  $f$ , i.e.

$$\hat{\mathbb{f}}_n^S(x) = \frac{1}{nh_n} \sum_{t=1}^n k\left(\frac{x - \hat{\eta}_t}{h_n}\right)\tag{3.8}$$

with kernel function  $k$  and bandwidth  $h_n > 0$ . Gao and Song (2008) consider Lipschitz-continuous kernels such as  $k(x) = \phi(x)$ , where  $\phi$  is the standard normal density function. An alternative estimator is based on the uniform kernel  $k(x) = \frac{1}{2} \mathbb{1}_{\{|x| \leq 1\}}$  yielding  $\hat{\mathbb{f}}_n^S(\hat{\xi}_{n,\alpha}) \xrightarrow{P} f(\xi_\alpha)$  whenever  $h_n \sim n^{-\varrho}$  for some  $\varrho \in (0, 1/2]$ . Based on (3.7) and (3.8), we obtain a consistent estimator for  $\Sigma_\alpha$  denoted by  $\hat{\Sigma}_{n,\alpha}$ .

Employing Theorem 2 we can study the asymptotic behavior of the conditional VaR estimator in (3.5). Since the conditional volatility varies over time, a limiting distribution cannot exist and therefore the concept of weak convergence is not applicable in this context. Beutner et al. (2019) advocate a *merging* concept that is discussed in the book of Dudley (2002, Section 11.7), i.e. two sequences of (random) probability measures  $\{P_n\}, \{Q_n\}$  *merge* (in probability) if and only if their bounded Lipschitz distance  $d_{BL}(P_n, Q_n)$  converges to zero (in probability).<sup>2</sup> Merging can be regarded as a generalization of weak convergence, where the latter corresponds to the

---

<sup>2</sup> Alternatively, merging can be defined in terms of the Prokhorov metric (Dudley, 2002, Theorem 11.7.1).

case  $Q_n = Q$  for all  $n$  with  $Q$  denoting the limiting distribution. While the Portman-teau lemma states several equivalent definitions of weak convergence of probability measures, it must be noted that this equivalence breaks down in the context of merging (D'Aristotile et al., 1988, Ex. 1.1). The bounded Lipschitz distance appears to be an appropriate and practical metric to study the asymptotic behavior of the VaR estimator. Presuming two independent samples, one for parameter estimation and one for conditioning, the delta method suggests<sup>3</sup> that the VaR estimator, centered at  $VaR_{n,\alpha}$  and inflated by  $\sqrt{n}$ , and

$$N \left( 0, \begin{pmatrix} -\xi_\alpha \frac{\partial \sigma_{n+1}(\theta_0)}{\partial \theta} \\ \sigma_{n+1} \end{pmatrix}' \Sigma_\alpha \begin{pmatrix} -\xi_\alpha \frac{\partial \sigma_{n+1}(\theta_0)}{\partial \theta} \\ \sigma_{n+1} \end{pmatrix} \right) \quad (3.9)$$

given  $\mathcal{F}_n$  merge in probability. Equation (3.9) highlights once more the relevance of the merging concept since its conditional variance still depends on  $n$  and does not converge as  $n \rightarrow \infty$ . Together with Theorem 1 and  $\hat{\Sigma}_{n,\alpha} \xrightarrow{p} \Sigma_\alpha$ , it yields a  $100(1-\gamma)\%$  confidence interval for  $VaR_{n,\alpha}$  with bounds (cf. Francq and Zakoïan, 2015, Eq. (23))

$$\widehat{VaR}_{n,\alpha} \pm \frac{\Phi^{-1}(\gamma/2)}{\sqrt{n}} \left\{ \begin{pmatrix} -\hat{\xi}_{n,\alpha} \frac{\partial \tilde{\sigma}_{n+1}(\hat{\theta}_n)}{\partial \theta} \\ \tilde{\sigma}_{n+1}(\hat{\theta}_n) \end{pmatrix}' \hat{\Sigma}_{n,\alpha} \begin{pmatrix} -\hat{\xi}_{n,\alpha} \frac{\partial \tilde{\sigma}_{n+1}(\hat{\theta}_n)}{\partial \theta} \\ \tilde{\sigma}_{n+1}(\hat{\theta}_n) \end{pmatrix} \right\}^{1/2}, \quad (3.10)$$

where  $\Phi$  is the standard normal cdf. However, with the exception of perhaps some experimental settings, researchers rarely have a replicate, independent of the original series, at hand. Beutner et al. (2019) provide an asymptotic justification for the interval on the basis of a single sample using a simple sample-split approach coupled

---

<sup>3</sup>Since the delta method follows from the continuous mapping theorem, which in turn relies on the Portmanteau lemma, it is not directly applicable in this merging context. In reference to Francq and Zakoïan (2015, page 162), we therefore use the verb *suggest*. In the case at hand the delta method hints at the correct approximate distribution (3.9). The sentence's claim can be formally shown by the definition of the bounded Lipschitz metric in the spirit of the proof of Corollary 1.

with a weak dependence condition (e.g. strong mixing). Although the interval in (3.10) is well-justified, it may perform poorly since the density estimation appears rather sensitive regarding the choice of bandwidth (see Gao and Song, 2008, Section 4). Bootstrap methods offer an alternative way to quantify the uncertainty around the estimators.

## 4 Bootstrap

Bootstrap approximations frequently provide better insight into the actual distribution than the asymptotic approximation, yet they require a careful set-up. Hall and Yao (2003) show that conventional bootstrap methods are inconsistent in a GARCH model lacking finite fourth moment in the case of the squared innovations' distribution not being in the domain of attraction of the normal distribution. They consider a subsample bootstrap instead and study its asymptotic properties. In correspondence, an  $m$ -out-of- $n$  without-replacement bootstrap is proposed by Spierdijk (2016) to construct confidence intervals for ARMA-GARCH VaR.

Pascual et al. (2006) present a residual bootstrap in a GARCH(1, 1) setting and assess its finite sample properties by means of simulation. Their bootstrap scheme follows a recursive design in which the bootstrap observations are generated iteratively using the estimated volatility dynamics. Building upon their results, Christoffersen and Gonçalves (2005) construct bootstrap confidence intervals for (conditional) VaR and Expected Shortfall and compare them to competitive methods within the GARCH(1, 1) model. Theoretical results on the recursive-design residual bootstrap are provided by Hidalgo and Zaffaroni (2007) and Jeong (2017) for the ARCH( $\infty$ ) and GARCH( $p, q$ ) model, respectively.

In contrast, Shimizu (2009) considers fixed-design variants of the wild and the residual bootstrap in which the ARMA-GARCH dynamics of the bootstrap samples are kept fixed at the values of the original series. The bootstrap estimators are based on a single Newton-Raphson iteration simplifying the proofs of first-order asymptotic validity. Shimizu’s approach for the residual bootstrap is also employed in a multivariate GARCH setting by Francq et al. (2016). Recently, Cavaliere et al. (2018) study the fixed-design residual bootstrap in the context of ARCH( $q$ ) models and propose a bootstrap Wald statistic based on a QML bootstrap estimator. While their theory has been developed independently to ours, their simulation study indicates that the fixed-design bootstrap performs as well as the recursive-design bootstrap.

## 4.1 Fixed-design Residual Bootstrap

We propose a fixed-design residual bootstrap procedure, described in Algorithm 1, to approximate the distribution of the estimators in (3.3) – (3.5).

**Algorithm 1.** (*Fixed-design residual bootstrap*)

1. For  $t = 1, \dots, n$ , generate  $\eta_t^* \stackrel{iid}{\sim} \hat{\mathbb{F}}_n$  and the bootstrap observation  $\epsilon_t^* = \tilde{\sigma}_t(\hat{\theta}_n)\eta_t^*$ .
2. Calculate the bootstrap estimator

$$\hat{\theta}_n^* = \arg \max_{\theta \in \Theta} L_n^*(\theta) \tag{4.1}$$

with the bootstrap criterion function given by

$$L_n^*(\theta) = \frac{1}{n} \sum_{t=1}^n \ell_t^*(\theta) \quad \text{and} \quad \ell_t^*(\theta) = -\frac{1}{2} \left( \frac{\epsilon_t^*}{\tilde{\sigma}_t(\theta)} \right)^2 - \log \tilde{\sigma}_t(\theta).$$

3. For  $t = 1, \dots, n$  compute the bootstrap residual  $\hat{\eta}_t^* = \epsilon_t^* / \tilde{\sigma}_t(\hat{\theta}_n^*)$  and obtain

$$\hat{\xi}_{n,\alpha}^* = \arg \min_{z \in \mathbb{R}} \frac{1}{n} \sum_{t=1}^n \rho_\alpha(\hat{\eta}_t^* - z). \quad (4.2)$$

4. Obtain the bootstrap estimator of the conditional VaR

$$\widehat{VaR}_{n,\alpha}^* = -\hat{\xi}_{n,\alpha}^* \tilde{\sigma}_{n+1}(\hat{\theta}_n^*). \quad (4.3)$$

*Remark 2.* In contrast to the literature, the bootstrap errors are drawn with replacement from the residuals rather than the standardized residuals. In fact, re-centering would be inappropriate in the case of  $\mathbb{E}[\eta_t] \neq 0$ . In addition, re-scaling of the residuals is typically redundant as  $\frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^2 = 1$  is implied by  $\hat{\theta}_n \in \mathring{\Theta}$  under Assumption 10; see Francq and Zakoïan, 2011, p. 182/406 and note that the solution requires  $\hat{\theta}_n$  belonging to the interior (Francq and Zakoïan, Oct. 2018, personal communication).

*Remark 3.* The term ‘fixed-design’ refers to the fact that the bootstrap observations are generated using  $\tilde{\sigma}_t(\hat{\theta}_n) = \sigma(\epsilon_{t-1}, \dots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \hat{\theta}_n)$ . In contrast, a recursive-design scheme replicates the model’s dynamic structure, i.e.  $\epsilon_t^* = \sigma_t^* \eta_t^*$  with  $\sigma_t^* = \sigma(\epsilon_{t-1}^*, \dots, \epsilon_1^*, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \hat{\theta}_n)$  and  $\eta_t^* \stackrel{iid}{\sim} \hat{\mathbb{F}}_n$ , which is computationally more demanding. We refer to Appendix B for a complete description. See also Cavaliere et al. (2018) for more theoretical insights on the difference in the design in an ARCH( $q$ ).

*Remark 4.* Whereas (4.1) involves a nonlinear optimization, Shimizu (2009) proposes a Newton-Raphson type bootstrap estimator instead. The Newton-Raphson bootstrap estimator corresponding to (4.1) is given by

$$\hat{\theta}_n^{*NR} = \hat{\theta}_n + \hat{J}_n^{-1} \frac{1}{2n} \sum_{t=1}^n \hat{D}_t(\eta_t^{*2} - 1),$$



which can considerably speed up computations.

In the following subsection we show the asymptotic validity of the fixed-design bootstrap procedure described in Algorithm 1.

## 4.2 Bootstrap Consistency

Subsequently, we employ the usual notation for bootstrap asymptotics, i.e. “ $\xrightarrow{p^*}$ ” and “ $\xrightarrow{d^*}$ ”, as well as the standard bootstrap stochastic order symbol “ $o_{p^*}(1)$ ” (cf. Chang and Park, 2003). To prove the asymptotic validity of the proposed bootstrap procedure, we first focus on the stochastic volatility part. Since  $L_n^*$  is maximized at  $\hat{\theta}_n^*$  its derivative is equal to zero:  $\frac{\partial L_n^*(\hat{\theta}_n^*)}{\partial \theta} = 0$ . A Taylor expansion around  $\hat{\theta}_n$  yields

$$0 = \sqrt{n} \frac{\partial L_n^*(\hat{\theta}_n^*)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\partial}{\partial \theta} \ell_t^*(\hat{\theta}_n) + \left( \frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) \right) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n)$$

with  $\check{\theta}_n$  between  $\hat{\theta}_n^*$  and  $\hat{\theta}_n$ . Lemma 6 in Appendix A.2 establishes  $\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) \xrightarrow{p^*} -2J$  almost surely. Since  $\frac{\partial}{\partial \theta} \ell_t^*(\theta) = \tilde{D}_t(\theta) \left( \frac{\epsilon_t^{*2}}{\sigma_t^2(\theta)} - 1 \right)$ , the first term on the right hand side reduces to  $\frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\eta_t^{*2} - 1)$ . Hence, we obtain

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) = \frac{1}{2} J^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\eta_t^{*2} - 1) + o_{p^*}(1) \quad (4.4)$$

almost surely with  $\frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\eta_t^{*2} - 1)$  converging in conditional distribution to  $N(0, (\kappa - 1)J)$  almost surely by Lemma 7 in Appendix A.2. The foregoing discussion can be summarized by the following intermediate result.

**Proposition 1.** *Suppose Assumptions 1–4, 5(i), 5(iii), 6, 7, 9 and 10 hold with*

$a = \pm 12$ ,  $b = 12$  and  $c = 6$ . Then, we have

$$\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d^*} N\left(0, \frac{\kappa - 1}{4} J^{-1}\right)$$

*almost surely.*

Proposition 1 establishes the asymptotic validity of the bootstrap for the volatility parameters. Next, we turn to the estimator of the quantile parameter associated with the VaR at level  $\alpha$ . Establishing the asymptotic validity of the bootstrap for the second part appears challenging since the bootstrap innovations are drawn from the discrete distribution  $\hat{\mathbb{F}}_n$ . To overcome this issue we rely on arguments employed by Bahadur (1966) and Berkes and Horváth (2003). Following the general steps of the proof of Francq and Zakoïan (2015, Theorem 4), we standardize equation (4.2) such that the bootstrap quantile estimator satisfies

$$\sqrt{n}(\hat{\xi}_{n,\alpha}^* - \hat{\xi}_{n,\alpha}) = \arg \min_{z \in \mathbb{R}} \underbrace{\sum_{t=1}^n \rho_\alpha\left(\hat{\eta}_t^* - \hat{\xi}_{n,\alpha} - \frac{z}{\sqrt{n}}\right) - \sum_{t=1}^n \rho_\alpha(\eta_t^* - \hat{\xi}_{n,\alpha})}_{Q_n^*(z)}.$$

Employing the identity of Koenker and Xiao (2006, Eq. (A.3)) we obtain<sup>4</sup>

$$Q_n^*(z) = zX_n^* + Y_n^* + I_n^*(z) + J_n^*(z) \tag{4.5}$$

---

<sup>4</sup>Note that the identity holds not only for  $u \neq 0$  but also for  $u = 0$ .

with

$$\begin{aligned}
X_n^* &= \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha), \\
Y_n^* &= \sum_{t=1}^n (\eta_t^* - \hat{\eta}_t^*) (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha), \\
I_n^*(z) &= \sum_{t=1}^n \int_0^{\frac{z}{\sqrt{n}}} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) ds, \\
J_n^*(z) &= \sum_{t=1}^n \int_{\frac{z}{\sqrt{n}}}^{\frac{z}{\sqrt{n}} + \eta_t^* - \hat{\eta}_t^*} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) ds.
\end{aligned}$$

Subsequently, we look at each term in turn while resorting to Lemmas 7 to 10 in Appendix A.2. Lemma 7 yields  $X_n^* \xrightarrow{d^*} N(0, \alpha(1 - \alpha))$  almost surely. Further, we notice that  $Y_n^*$  neither depends on  $z$  nor interacts with it; therefore it can be disregarded. The term  $I_n^*(z)$  converges in conditional probability to  $\frac{z^2}{2} f(\xi_\alpha)$  in probability by Lemma 8. Next, we analyze the asymptotic properties of  $J_n^*(z)$ , which can be split into  $J_n^*(z) = J_{n,1}^*(z) + J_{n,2}^*(z)$  with

$$J_{n,1}^*(z) = \sum_{t=1}^n \int_0^{\eta_t^* - \hat{\eta}_t^*} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} + s\}} - \mathbb{1}_{\{\eta_t^* - \hat{\xi}_{n,\alpha} - z/\sqrt{n} < 0\}}) ds \quad (4.6)$$

$$J_{n,2}^*(z) = \sum_{t=1}^n (\eta_t^* - \hat{\eta}_t^*) (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}). \quad (4.7)$$

Deviating from the proof of Francq and Zakoïan (2015), Lemma 9 shows that  $J_{n,1}^*(z)$  converges in conditional distribution to a random variable, which does not depend on  $z$ , in probability. We refer to Remark 6 in Appendix A.2 for more details on the difference of the proofs. Further, the second term is equal to  $J_{n,2}^*(z) = z \xi_\alpha f(\xi_\alpha) \Omega' \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) + o_{p^*}(1)$  in probability by Lemma 10. By the preceding discussion we obtain

$$Q_n^*(z) = \frac{z^2}{2} f(\xi_\alpha) + z \left( X_n^* + \xi_\alpha f(\xi_\alpha) \Omega' \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) \right) + J_{n,1}^*(z) + Y_n^* + o_{p^*}(1)$$

in probability. Employing Xiong and Li (2008, Theorem 3.3) and the basic corollary of Hjort and Pollard (2011), we obtain<sup>5</sup>

$$\sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*) = \xi_\alpha \Omega' \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) + \frac{1}{f(\xi_\alpha)} \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha) + o_{p^*}(1)$$

in probability. Together with (4.4) we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \\ \sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*) \end{pmatrix} = \begin{pmatrix} \frac{1}{2} J^{-1} & O_{r \times 1} \\ \frac{1}{2} \xi_\alpha \Omega' J^{-1} & \frac{1}{f(\xi_\alpha)} \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t (\eta_t^{*2} - 1) \\ \frac{1}{\sqrt{n}} \sum_{t=1}^n (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha) \end{pmatrix} + o_{p^*}(1).$$

Employing Lemma 7 leads to the paper's main result.

**Theorem 3.** (*Bootstrap consistency*) Suppose Assumptions 1–10 hold with  $a = \pm 12$ ,  $b = 12$  and  $c = 6$ . Then, we have

$$\begin{pmatrix} \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \\ \sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*) \end{pmatrix} \xrightarrow{d^*} N(0, \Sigma_\alpha)$$

in probability.

Theorem 3 is useful to validate the bootstrap for the conditional VaR estimator. For the asymptotic behavior of the conditional VaR estimator we refer to (3.9) and the text preceding it. The following corollary is established.

**Corollary 1.** Under the assumptions of Theorem 3 the conditional distribution of  $\sqrt{n}(\widehat{VaR}_{n,\alpha}^* - \widehat{VaR}_{n,\alpha})$  given  $\mathcal{F}_n$  and (3.9) given  $\mathcal{F}_n$  merge in probability.

---

<sup>5</sup>Matching notation, we take  $A_n(z) = Q_n^*(z)$ , which is convex, and set  $B_n(z) = \frac{z^2}{2}V + zU_n + C_n$ , where  $V = f(\xi_\alpha)$ ,  $U_n = X_n^* + \xi_\alpha f(\xi_\alpha) \Omega' \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)$  and  $C_n + r_n(z) = J_{n,1}^*(z) + Y_n^* + o_{p^*}(1)$  with  $r_n(z) \xrightarrow{p} 0$  for each  $z \in \mathbb{R}$ . The minimizers of  $A_n(z)$  and  $B_n(z)$  are  $\alpha_n = \sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*)$  and  $\beta_n = -V^{-1}U_n$ , respectively. The basic corollary of Hjort and Pollard (2011) states  $\alpha_n - \beta_n = o_p(1)$ , which implies  $\alpha_n - \beta_n = o_{p^*}(1)$  in probability (Xiong and Li, 2008, Theorem 3.3).

The proof of the corollary is deferred to Appendix A.2. Having proven first-order asymptotic validity of the bootstrap procedure described in Section 4.1, we turn to constructing bootstrap confidence intervals for VaR.

### 4.3 Bootstrap Confidence Intervals for VaR

Clearly, the VaR evaluation in (3.5) is subject to estimation risk that needs to be quantified. We propose the following algorithm to obtain approximately  $100(1 - \gamma)\%$  confidence intervals.

**Algorithm 2.** (*Fixed-design Bootstrap Confidence Intervals for VaR*)

1. Acquire a set of  $B$  bootstrap replicates, i.e.  $\widehat{VaR}_{n,\alpha}^{*(b)}$  for  $b = 1, \dots, B$ , by repeating Algorithm 1.

- 2.1. Obtain the *equal-tailed percentile* (EP) interval

$$\left[ \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(1 - \gamma/2), \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(\gamma/2) \right] \quad (4.8)$$

with  $\hat{G}_{n,B}^{*-1}(\cdot)$  being the quantile function (generalized inverse) of  $\hat{G}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\sqrt{n}(\widehat{VaR}_{n,\alpha}^{*(b)} - \widehat{VaR}_{n,\alpha}) \leq x\}}$ .

- 2.2. Calculate the *reversed-tails* (RT) interval

$$\left[ \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(\gamma/2), \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(1 - \gamma/2) \right]. \quad (4.9)$$

- 2.3. Compute the *symmetric* (SY) interval

$$\left[ \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{*-1}(1 - \gamma), \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{*-1}(1 - \gamma) \right] \quad (4.10)$$

with  $\hat{H}_{n,B}^{*-1}(\cdot)$  being the quantile function (generalized inverse) of  $\hat{H}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\sqrt{n}|\widehat{VaR}_{n,\alpha}^{*(b)} - \widehat{VaR}_{n,\alpha}| \leq x\}}$ .

The interval in (4.8) is obtained by the EP method, that is frequently encountered in the bootstrap literature. It is obtained from the (typically) infeasible equal-tailed confidence interval

$$\left[ \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} G_n^{-1}(1 - \gamma/2), \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} G_n^{-1}(\gamma/2) \right],$$

where  $G_n^{-1}$  is the (unknown) quantile function of  $\sqrt{n}(\widehat{VaR}_{n,\alpha} - VaR_{n,\alpha})$ , which is replaced by its bootstrap analogue  $\hat{G}_{n,B}^{*-1}$ . The same reasoning leads to the SY interval but with test statistic  $\sqrt{n}|\widehat{VaR}_{n,\alpha} - VaR_{n,\alpha}|$  instead of  $\sqrt{n}(\widehat{VaR}_{n,\alpha} - VaR_{n,\alpha})$  which makes it also clear that the interval in (4.10) presumes symmetry for rationalizing its construction. “Flipping around” its tails leads to the RT interval given in (4.9), which can be motivated by the results of Falk and Kaufmann (1991).<sup>6</sup> Clearly, the RT and the EP have equal length. Whereas (4.9) in its current form emphasizes the interval’s name, RT type intervals are frequently reported in their reduced form, i.e. the lower and upper bound of (4.9) simplify to the  $\gamma/2$  and  $1 - \gamma/2$  quantiles of  $\frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\widehat{VaR}_{n,\alpha}^{*(b)} \leq x\}}$ , respectively. RT intervals can either be motivated by the results of Falk and Kaufmann (1991)<sup>7</sup> or as the bootstrap analogue of the (uncentered) statistic  $\widehat{VaR}_{n,\alpha}$ . It is worth mentioning that RT type bootstrap intervals for the VaR are also constructed in reduced form by Christoffersen and Gonçalves (2005).

Regardless of whether we use an EP, RT or SY interval the meaning is always the

---

<sup>6</sup>In a random sample setting Falk and Kaufmann (1991) prove that the RT bootstrap interval for quantiles has asymptotically greater coverage than the corresponding EP bootstrap interval. For additional insights we refer to Hall and Martin (1988).

<sup>7</sup>In a random sample setting Falk and Kaufmann (1991) prove that the RT bootstrap interval for quantiles has asymptotically greater coverage than the corresponding EP bootstrap interval. For additional insights we refer to Hall and Martin (1988).

same: Given the past up to and including time  $n$  the probability that the conditional VaR for period  $n+1$  is contained in the intervals is approximately equal to  $100(1-\gamma)\%$ .

## 4.4 Bootstrap Extensions

The asymptotic normality result in Theorem 2 as well as the bootstrap consistency in Theorem 3 are derived, *inter alia*, under the assumption that the innovations are iid. In case this is not believed to be true – e.g. if the suggested specification tests mentioned in Section 3 indicate otherwise – asymptotic normality of  $\sqrt{n}(\hat{\theta}_n - \theta_0)$  can still be established under regularity assumptions. Escanciano (2009) studies the QML estimator under some dependence among the  $\eta_t$ 's while imposing slightly stronger (moment) conditions, whereas the related paper of Linton et al. (2010) investigates estimators in a GARCH(1,1) with dependent errors but under weaker moment conditions. A multivariate version of the dependence condition in Escanciano (2009) can be found in Francq and Zakoïan (2016).

Whereas the bootstrap method presented in Algorithm 1 is contingent on the iid assumption, alternative bootstrap techniques may be used if the iid condition is thought to be unrealistic. A variety of bootstrap methods exist that can capture dependence and non-identical random variables; see e.g. Lahiri (2003) for a broad overview. The wild or multiplier bootstrap (Mammen, 1993; Davidson and Flachaire, 2008) is particularly suited for dealing with non-identical variables, but does not capture dependence, unless it is properly modified (Shao, 2010; Friedrich et al., 2020).

Alternatively, one may go with a block bootstrap method which is appropriate in such settings. One possible choice is the moving block bootstrap (MBB) of Corradi and Iglesias (2008), who propose to resample the (pseudo-)likelihood in blocks. Although their method can in principle allow for dependence in the errors, we note

that their theory maintains the assumption of iid errors to establish asymptotic refinements. We consider a different variant of the MBB in the following algorithm, which is an extension of the fixed-design residual bootstrap.

**Algorithm 3.** (*Fixed-design moving block bootstrap*)

1. Build overlapping blocks of block size  $l \in \{1, \dots, n\}$  from the residuals and join together  $b = \lfloor n/l \rfloor$  blocks chosen randomly (with replacement), i.e.  $\{\eta_1^\diamond, \dots, \eta_b^\diamond\} = \{\hat{\eta}_{U_1}, \dots, \hat{\eta}_{U_1+l}, \dots, \hat{\eta}_{U_b}, \dots, \hat{\eta}_{U_b+l}\}$  with  $U_1, \dots, U_b \stackrel{iid}{\sim} \text{Uniform}\{1, \dots, n-l+1\}$ . Generate the bootstrap observation  $\epsilon_t^\diamond = \tilde{\sigma}_t(\hat{\theta}_n)\eta_t^\diamond$ .
2. - 4. Analogous to Algorithm 1 with  $*$  replaced by  $\diamond$

The advantage of this variant in our context is that it yields the fixed-design residual bootstrap of Algorithm 1 as a special case by taking  $l = 1$ . The choice of the block length  $l$  involves a trade-off between capturing the potential dependence structure and having a sufficient number of blocks for stable estimation. Although establishing the validity of the fixed-design MBB is beyond the scope of this paper, we apply it in our empirical application in Section 5.2 to compare it to the residual bootstrap.

## 5 Numerical Illustration

### 5.1 Monte Carlo Experiment

In order to evaluate the finite sample performance of the proposed bootstrap procedure a Monte Carlo experiment is conducted. We confine ourselves to four conditional



volatility specifications related to Examples 1 and 2 in Section 2. The first two are GARCH(1, 1) parameterizations with

$$(i) \text{ high persistence: } \theta_0 = (\omega_0, \alpha_0, \beta_0)' = (0.05 \times 20^2/252, 0.15, 0.8)';$$

$$(ii) \text{ low persistence: } \theta_0 = (\omega_0, \alpha_0, \beta_0)' = (0.05 \times 20^2/252, 0.4, 0.55)',$$

which are similar to the specifications of Gao and Song (2008, Section 4) and Spierdijk (2016, Section 4.2). In addition, we study two T-GARCH(1, 1) scenarios likewise associated with high and low persistence:

$$(iii) \text{ high persistence: } \theta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0)' = (0.05 \times 20/\sqrt{252}, 0.05, 0.10, 0.8)';$$

$$(iv) \text{ low persistence: } \theta_0 = (\omega_0, \alpha_0^+, \alpha_0^-, \beta_0)' = (0.05 \times 20/\sqrt{252}, 0.1, 0.3, 0.55)'.$$

Within the experiment the VaR level takes two values, i.e.  $\alpha \in \{0.01, 0.05\}$ , and there are two possible innovation distributions: the standard normal distribution and a Student- $t$  distribution with 6 degrees of freedom (df).<sup>8</sup> We consider four estimation sample sizes,  $n \in \{250; 500; 1,000; 5,000\}$ , whereas the number of bootstrap replicates is fixed and equal to  $B = 2,000$ . For each model version we simulate  $S = 2,000$  independent Monte Carlo trajectories. The combinations  $\alpha = 0.01$  and  $n = 250$  and  $n = 500$ , respectively, are included to see how the proposed bootstrap method works if we are looking at the tail of the distribution for relatively small  $n$  (see Remark 1).

All simulations are performed on a HP Z640 workstation with 16 cores using Matlab R2016a. The numerical optimization of the log-likelihood function is carried out employing the build-in function *fmincon* and running time is reduced by parallel computing using *parfor*. The code is available on the website of the third author.

---

<sup>8</sup>The Student- $t$  innovations are appropriately standardized to satisfy  $\mathbb{E}\eta_t^2 = 1$ .

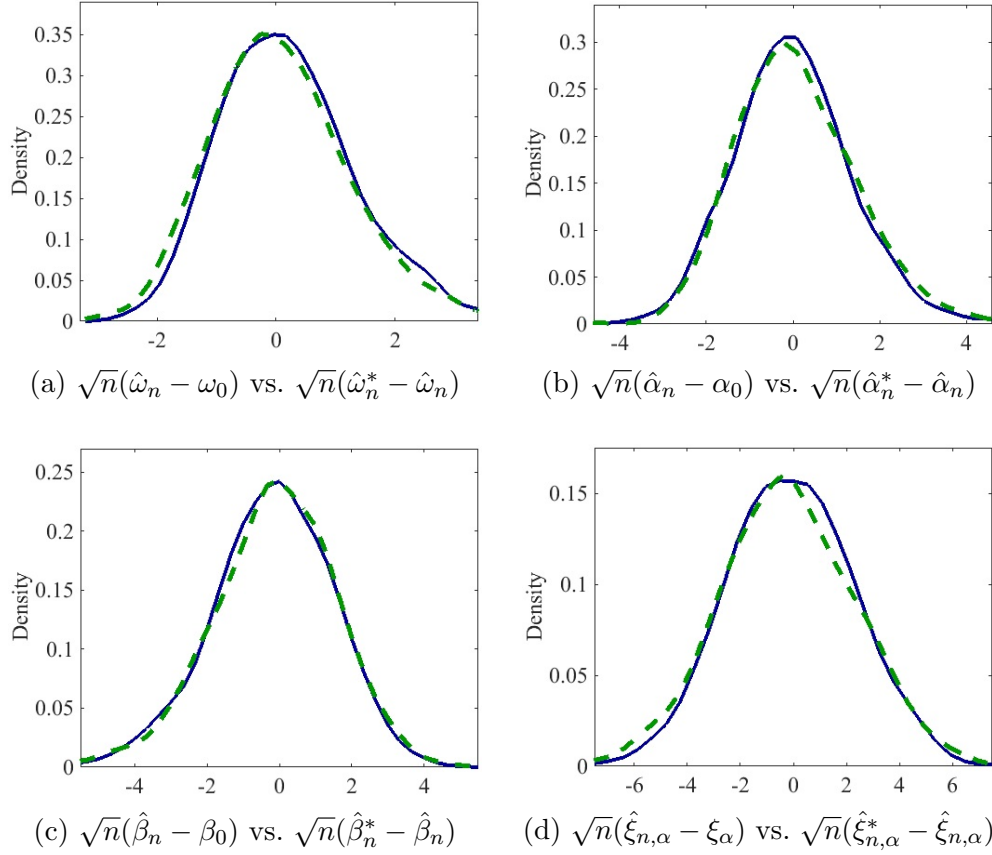


Figure 1: Density estimates for the distribution of the 2-step QMLE (full line) based on  $S = 2,000$  simulations and the fixed-design bootstrap distribution (dashed line) based on  $B = 2,000$  replications.  $\alpha$  is set to 0.05 and the DGP is a GARCH(1, 1) with  $\theta_0 = (0.08, 0.15, 0.8)'$ , sample size  $n = 5,000$  and (normalized) Student-t innovations (6 degrees of freedom).

Figure 1 displays the density of the distribution of the two-step QMLE estimator and the corresponding bootstrap distribution (given a particular sample) in the high persistence GARCH(1, 1) case for  $n = 5,000$ . Figures 1(a) to 1(c) indicate that the bootstrap distribution mimics adequately the finite sample distribution of the estimator of the volatility parameters. Besides, Figure 1(d) illustrates that the bootstrap approximation works as well for the distribution of the quantile estimator. Moreover, all density plots are roughly bell-shaped supporting the theoretical implications of Theorem 2 and 3.

Table 1 reports the results of the three 90%-bootstrap intervals for the 5%-VaR when the innovation distribution is Student-t (henceforth referred to as baseline). The results of the interval (3.10) based on asymptotic (AS) theory are included for comparison, where a Gaussian kernel is utilized together with a bandwidth following Silverman's (1986) rule-of-thumb. In the GARCH(1, 1) high persistence case (top right), we see that the average coverage varies around 90% across all sample sizes for the RT and the SY interval. In contrast, the EP and the AS interval fall short of the nominal 90% by 9.65 and 3.85 percentage points (pp), respectively, for small sample size ( $n = 250$ ). Nevertheless, their average coverage approaches the nominal value as the sample size increases. Remarkably, for all four intervals the average rate of the conditional VaR being below the interval is considerably less than the average rate of the conditional VaR being above the interval when the sample size is rather small ( $n \leq 500$ ). Regarding the intervals' length, we observe that the SY interval is on average larger than the EP/RT interval. As the sample size increases this gap diminishes and the intervals' average lengths shrink. Considering the low persistent case (top left) we find similar results regarding the intervals' average coverage, yet their average lengths turn out to be smaller compared to the high persistent case.

Sample Size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
GARCH(1,1)							
		low persistence			high persistence		
250	EP	81.10	7.30/11.60	0.569	80.35	7.75/11.90	0.776
	RT	90.30	3.15/6.55	0.569	90.20	3.70/6.10	0.776
	SY	87.90	4.25/7.85	0.592	88.80	3.60/7.60	0.807
	AS	86.10	3.75/10.15	0.577	86.15	4.25/9.60	0.774
500	EP	84.50	6.30/9.20	0.431	84.25	6.30/9.45	0.582
	RT	91.50	3.75/4.75	0.431	91.45	3.40/5.15	0.582
	SY	90.40	3.60/6.00	0.443	90.10	3.65/6.25	0.596
	AS	88.95	3.50/7.55	0.440	88.20	3.85/7.95	0.568
1,000	EP	87.05	5.05/7.90	0.305	86.45	6.05/7.50	0.417
	RT	91.55	3.75/4.70	0.305	91.05	4.50/4.45	0.417
	SY	91.15	3.55/5.30	0.310	90.30	4.75/4.95	0.424
	AS	89.40	4.00/6.60	0.314	89.25	4.45/6.30	0.410
5,000	EP	87.45	6.15/6.40	0.144	87.85	5.70/6.45	0.191
	RT	90.35	5.30/4.35	0.144	89.50	5.25/5.25	0.191
	SY	89.75	5.35/4.90	0.145	89.70	4.80/5.50	0.192
	AS	89.25	5.25/5.50	0.145	88.60	5.25/6.15	0.188
T-GARCH(1,1)							
		low persistence			high persistence		
250	EP	79.70	7.35/12.95	0.139	80.45	7.05/12.50	0.287
	RT	90.05	3.95/6.00	0.139	90.85	3.05/6.10	0.287
	SY	88.75	3.95/7.30	0.145	89.30	3.00/7.70	0.300
	AS	88.00	3.65/8.35	0.146	89.00	2.95/8.05	0.302
500	EP	82.80	6.10/11.10	0.104	82.35	6.25/11.40	0.214
	RT	90.20	4.20/5.60	0.104	91.30	3.50/5.20	0.214
	SY	89.15	4.05/6.80	0.107	90.10	2.95/6.95	0.219
	AS	89.15	3.65/7.20	0.108	89.80	3.05/7.15	0.221
1,000	EP	84.45	6.00/9.55	0.076	82.95	6.90/10.15	0.156
	RT	90.10	4.60/5.30	0.076	90.75	4.50/4.75	0.156
	SY	89.00	4.35/6.65	0.077	89.10	4.55/6.35	0.159
	AS	88.90	4.10/7.00	0.079	88.65	4.40/6.95	0.161
5,000	EP	88.40	5.35/6.25	0.035	88.30	4.95/6.75	0.073
	RT	90.35	5.20/4.45	0.035	90.45	4.80/4.75	0.073
	SY	90.75	4.70/4.55	0.035	89.75	4.45/5.80	0.074
	AS	91.30	4.25/4.45	0.036	90.40	4.40/5.20	0.075

Table 1 reports distinct features of the **fixed-design** bootstrap confidence intervals and the asymptotic interval for the conditional VaR at **level  $\alpha = 0.05$**  with **nominal coverage  $1 - \gamma = 90\%$** . For each interval type and different sample sizes ( $n$ ), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The bootstrap intervals are based on  $B = 2,000$  bootstrap replications and the averages are computed using  $S = 2,000$  simulations. The top presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) **Student-t innovations (6 df)**, whereas below the DGP is a Student-t T-GARCH(1,1).

This is intuitive as the conditional volatility tends to vary less in the low persistent case. Regarding the T-GARCH(1, 1), the overall picture is similar as in the GARCH case, however the under-coverage in small and medium-sized samples appears to be more extreme for the EP and reduced for the AS interval.

Next, we consider deviations from the baseline specification. In particular, we study a change in the innovation distribution  $F$  (Table 2), a change in the VaR level  $\alpha$  (Table 3) and a change in intervals' nominal coverage probability  $100(1 - \gamma)\%$  (Table 4). While Table 5 draws attention to the average coverage gap between the EP and the RT bootstrap interval, Table 6 permits a comparison of the fixed-design bootstrap with its recursive-design counterpart.

The simulation results for the scenario when the  $\eta_t$ 's follow a standard normal distribution are tabulated in Table 2. Although the error distribution underlying the QMLE is correctly specified in this case, the qualitative results stated above with regard to Table 1 persist: the RT and the SY intervals possess accurate coverage rates across sample sizes, whereas the EP and the AS interval exhibit under-coverage in samples of rather small size with different extent. Moreover, we observe that the intervals are on average shorter in the Gaussian case than in the baseline case. This seems partially driven by a smaller variance of  $\hat{\xi}_{n,\alpha}$ ; for  $\alpha = 0.05$  the asymptotic variance  $\zeta_\alpha$  in (3.6) is equal to 3.11 in the Gaussian case compared to 5.72 in the Student-t case with 6 degrees of freedom.

Table 3 focuses on the VaR at level  $\alpha = 0.01$  instead. In comparison to Table 1 it is striking that the EP and AS interval perform worse in terms of average coverage (especially for smaller sample sizes). Take note that this attribute is mainly driven by differences in the right tail of the bootstrap density. In contrast, the average coverage of the RT and the SY interval remain varying around 90% for  $n \geq 1,000$  while a small

Sample Size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
GARCH(1,1)							
		low persistence			high persistence		
250	EP	80.65	8.20/11.15	0.504	80.30	8.20/11.50	0.648
	RT	89.30	2.50/8.20	0.504	89.20	3.00/7.80	0.648
	SY	88.40	3.25/8.35	0.526	87.95	3.70/8.35	0.675
	AS	85.80	3.95/10.25	0.508	85.10	4.55/10.35	0.636
500	EP	85.10	6.75/8.15	0.384	83.10	7.75/9.15	0.472
	RT	91.45	3.10/5.45	0.384	89.70	3.60/6.70	0.472
	SY	90.85	3.50/5.65	0.396	88.65	4.20/7.15	0.482
	AS	89.15	4.10/6.75	0.391	87.20	4.50/8.30	0.459
1,000	EP	85.25	7.10/7.65	0.261	87.55	5.55/6.90	0.335
	RT	91.00	3.50/5.50	0.261	91.10	3.25/5.65	0.335
	SY	89.50	4.30/6.20	0.266	90.85	3.55/5.60	0.340
	AS	89.05	3.95/7.00	0.264	89.15	4.15/6.70	0.327
5,000	EP	87.50	5.30/7.20	0.121	87.85	5.55/6.60	0.149
	RT	90.20	4.35/5.45	0.121	89.30	4.85/5.85	0.149
	SY	89.75	4.30/5.95	0.122	89.15	4.95/5.90	0.150
	AS	89.10	4.40/6.50	0.121	88.95	4.95/6.10	0.147
T-GARCH(1,1)							
		low persistence			high persistence		
250	EP	81.50	6.60/11.90	0.116	80.65	7.70/11.65	0.238
	RT	90.20	2.25/7.55	0.116	90.00	2.15/7.85	0.238
	SY	88.65	2.80/8.55	0.121	89.10	2.55/8.35	0.248
	AS	88.50	2.50/9.00	0.119	88.85	2.60/8.55	0.247
500	EP	85.15	5.90/8.95	0.086	83.50	6.65/9.85	0.173
	RT	90.10	3.30/6.60	0.086	90.20	2.85/6.95	0.173
	SY	89.45	3.75/6.80	0.088	89.15	3.60/7.25	0.178
	AS	89.30	3.55/7.15	0.088	89.60	3.10/7.30	0.178
1,000	EP	84.80	5.95/9.25	0.061	84.60	6.60/8.80	0.125
	RT	90.05	3.85/6.10	0.061	90.90	3.25/5.85	0.125
	SY	89.50	3.85/6.65	0.062	89.55	4.05/6.40	0.128
	AS	89.25	3.75/7.00	0.063	89.50	3.70/6.80	0.128
5,000	EP	87.95	5.30/6.75	0.028	86.85	5.60/7.55	0.057
	RT	89.90	4.40/5.70	0.028	88.65	4.50/6.85	0.057
	SY	89.55	4.55/5.90	0.028	88.35	4.65/7.00	0.058
	AS	90.15	4.15/5.70	0.029	89.35	4.25/6.40	0.059

Table 2 reports distinct features of the **fixed-design** bootstrap confidence intervals and the asymptotic interval for the conditional VaR at **level  $\alpha = 0.05$**  with **nominal coverage  $1 - \gamma = 90\%$** . For each interval type and different sample sizes ( $n$ ), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The bootstrap intervals are based on  $B = 2,000$  bootstrap replications and the averages are computed using  $S = 2,000$  simulations. The top presents the results for the low and high persistence parametrization of a GARCH(1,1) with **Gaussian innovations**, whereas below the DGP is a Gaussian T-GARCH(1,1).

Sample Size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
GARCH(1,1)							
		low persistence			high persistence		
250	EP	72.95	8.15/18.90	1.288	71.75	9.05/19.20	1.715
	RT	87.05	1.25/11.70	1.288	86.55	1.20/12.25	1.715
	SY	85.35	1.70/12.95	1.302	85.25	1.70/13.05	1.730
	AS	78.20	2.90/18.90	1.133	78.05	2.75/19.20	1.500
500	EP	78.40	7.40/14.20	0.918	79.65	7.00/13.35	1.227
	RT	89.45	2.40/8.15	0.918	89.70	2.05/8.25	1.227
	SY	87.85	2.60/9.55	0.955	88.55	2.60/8.85	1.272
	AS	83.50	3.20/13.30	0.910	84.20	3.05/12.75	1.189
1,000	EP	81.45	5.75/12.80	0.657	82.00	5.60/12.40	0.886
	RT	90.40	2.30/7.30	0.657	89.90	3.05/7.05	0.886
	SY	88.95	2.85/8.20	0.679	88.80	3.20/8.00	0.914
	AS	85.85	2.80/11.35	0.644	85.75	3.25/11.00	0.841
5,000	EP	85.30	5.80/8.90	0.306	85.95	5.05/9.00	0.407
	RT	91.30	3.60/5.10	0.306	91.05	3.50/5.45	0.407
	SY	90.45	3.65/5.90	0.312	90.40	3.40/6.20	0.413
	AS	88.90	3.45/7.65	0.302	88.40	3.85/7.75	0.392
T-GARCH(1,1)							
		low persistence			high persistence		
250	EP	71.15	8.85/20.00	0.307	70.20	10.05/19.75	0.625
	RT	85.75	1.50/12.75	0.307	85.35	1.45/13.20	0.625
	SY	83.85	1.90/14.25	0.310	84.45	1.55/14.00	0.636
	AS	79.05	2.90/18.05	0.278	79.05	2.95/18.00	0.572
500	EP	77.95	7.00/15.05	0.219	77.70	7.70/14.60	0.449
	RT	88.35	2.20/9.45	0.219	88.65	1.70/9.65	0.449
	SY	86.65	2.60/10.75	0.228	88.10	1.95/9.95	0.467
	AS	84.55	2.65/12.80	0.220	84.70	2.25/13.05	0.448
1,000	EP	80.55	5.50/13.95	0.160	79.60	6.55/13.85	0.330
	RT	89.95	2.10/7.95	0.160	89.45	2.55/8.00	0.330
	SY	87.75	2.60/9.65	0.165	87.25	3.20/9.55	0.341
	AS	85.80	2.20/12.00	0.158	84.80	3.35/11.85	0.325
5,000	EP	86.25	4.85/8.90	0.074	85.50	5.55/8.95	0.155
	RT	91.40	3.70/4.90	0.074	91.80	3.70/4.50	0.155
	SY	90.20	3.60/6.20	0.075	90.25	3.75/6.00	0.157
	AS	89.80	3.40/6.80	0.074	89.25	4.15/6.60	0.154

Table 3 reports distinct features of the **fixed-design** bootstrap confidence intervals and the asymptotic interval for the conditional VaR at **level**  $\alpha = 0.01$  with **nominal coverage**  $1 - \gamma = 90\%$ . For each interval type and different sample sizes ( $n$ ), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The bootstrap intervals are based on  $B = 2,000$  bootstrap replications and the averages are computed using  $S = 2,000$  simulations. The top presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) **Student-t innovations (6 df)**, whereas below the DGP is a Student-t T-GARCH(1,1).

loss of accuracy occurs in shorter samples. Coherent with a value of  $\zeta_\alpha$  around 32 at  $\alpha = 0.01$  in the Student-t case, we find the intervals for the 1%-VaR to be on average considerably longer than the intervals for the 5%-VaR in the baseline case.

Increasing the intervals' nominal value from 90% to 95%, Table 4 presents the results of the intervals for the 5%-VaR. Again, the RT and the SY intervals perform well in terms of coverage: across sample sizes their average coverages are fairly close to 95%. Once more, the EP and AS interval fall short of the nominal coverage value, yet the discrepancies appear to be less in comparison to the baseline. For example in the high-persistent GARCH case with  $n = 500$ , the EP interval falls short by  $95\% - 90.25\% = 4.75\text{pp}$  compared to  $90\% - 84.25\% = 5.75\text{pp}$  (see Table 1).

While the small-sample-performance of the AS interval can be explained by its embodied density estimation, the question arises why the EP interval performs worse than the other bootstrap intervals, which seems counter-intuitive at first. Howbeit the results are in line with the theoretical findings of Falk and Kaufmann (1991, unnumbered Corollary, p. 488). In a random sample setting they prove that the RT bootstrap interval for quantiles has asymptotically greater coverage than the corresponding EP bootstrap interval. The emerging gap<sup>9</sup>

- (i) tends to be smaller for larger sample sizes,
- (ii) tends to be larger for more extreme quantiles, and
- (iii) tends to vary with the nominal coverage rate in a non-monotonic way.

Table 5 presents the average coverage gap between the EP and the RT bootstrap interval of the conditional VaR. For example, in the low persistence GARCH(1, 1) case

---

<sup>9</sup>We neglect their  $o(n^{-1/2})$  term. Take note that the theoretical results of Falk and Kaufmann (1991) are not directly applicable in our setting due to GARCH-type effects.



Sample Size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
GARCH(1,1)							
		low persistence			high persistence		
250	EP	87.20	4.15/8.65	0.682	85.80	4.50/9.70	0.929
	RT	94.75	1.45/3.80	0.682	95.10	1.15/3.75	0.929
	SY	93.90	1.70/4.40	0.719	94.00	1.40/4.60	0.982
	AS	91.65	1.65/6.70	0.688	91.95	1.50/6.55	0.923
500	EP	90.20	3.25/6.55	0.515	90.25	3.30/6.45	0.696
	RT	96.00	1.70/2.30	0.515	96.40	1.45/2.15	0.696
	SY	95.55	1.35/3.10	0.534	95.15	1.50/3.35	0.720
	AS	93.90	1.60/4.50	0.524	93.40	1.65/4.95	0.677
1,000	EP	92.65	2.45/4.90	0.364	91.80	3.45/4.75	0.498
	RT	96.10	2.05/1.85	0.364	95.65	2.20/2.15	0.498
	SY	95.75	1.40/2.85	0.373	95.30	2.00/2.70	0.510
	AS	94.85	1.45/3.70	0.374	93.85	2.30/3.85	0.488
5,000	EP	92.95	3.45/3.60	0.171	93.25	2.85/3.90	0.228
	RT	95.65	2.15/2.20	0.171	95.30	2.20/2.50	0.228
	SY	94.90	2.50/2.60	0.173	95.05	2.20/2.75	0.230
	AS	94.70	2.40/2.90	0.173	94.35	2.35/3.30	0.224
T-GARCH(1,1)							
		low persistence			high persistence		
250	EP	86.65	4.30/9.05	0.167	86.30	4.15/9.55	0.346
	RT	95.25	1.65/3.10	0.167	95.40	1.55/3.05	0.346
	SY	94.60	1.55/3.85	0.175	95.00	1.25/3.75	0.365
	AS	93.45	1.55/5.00	0.174	94.25	1.10/4.65	0.359
500	EP	88.70	3.50/7.80	0.125	88.45	3.75/7.80	0.256
	RT	95.60	1.90/2.50	0.125	96.25	1.30/2.45	0.256
	SY	94.40	1.60/4.00	0.129	94.85	1.45/3.70	0.266
	AS	94.00	1.25/4.75	0.129	94.55	1.00/4.45	0.264
1,000	EP	89.90	3.65/6.45	0.090	90.50	3.40/6.10	0.186
	RT	95.55	2.00/2.45	0.090	95.45	1.85/2.70	0.186
	SY	94.70	2.00/3.30	0.093	94.50	1.95/3.55	0.192
	AS	94.35	1.75/3.90	0.094	93.85	2.00/4.15	0.192
5,000	EP	93.70	2.65/3.65	0.042	93.55	2.30/4.15	0.087
	RT	95.50	2.50/2.00	0.042	95.65	2.40/1.95	0.087
	SY	95.20	2.30/2.50	0.042	95.45	2.00/2.55	0.088
	AS	95.15	2.20/2.65	0.043	95.75	1.90/2.35	0.090

Table 4 reports distinct features of the **fixed-design** bootstrap confidence intervals and the asymptotic interval for the conditional VaR at **level  $\alpha = 0.05$**  with **nominal coverage  $1 - \gamma = 95\%$** . For each interval type and different sample sizes ( $n$ ), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The bootstrap intervals are based on  $B = 2,000$  bootstrap replications and the averages are computed using  $S = 2,000$  simulations. The top presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) **Student-t innovations (6 df)**, whereas below the DGP is a Student-t T-GARCH(1,1).

of the baseline with  $n = 250$ , the average coverage gap amounts to  $90.30\% - 81.10\% = 9.20\text{pp}$  (see also Table 1). It is striking that all values are positive within Table 5, which highlights the superiority of the RT bootstrap interval over the EP bootstrap interval. Further, it is eminent that average coverage gap tends to decrease with increasing sample size, which supports (i). Comparing columns (1) and (3) we also find that the average coverage gap tends to be larger for the 1%-VaR than for the 5%-VaR, which gives rise to (ii). Regarding (iii), the result of Falk and Kaufmann (1991) suggests that the gap slightly decreases when increasing the nominal coverage from 90% to 95%. Such tendency is precisely observed when comparing columns (1) and (4) of Table 5.

Sample size	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Panel I: GARCH(1, 1)								
	low persistence				high persistence			
250	9.20	8.65	14.10	7.55	9.85	8.90	14.80	9.30
500	7.00	6.35	11.05	5.80	7.20	6.60	10.05	6.15
1,000	4.50	5.75	8.95	3.45	4.60	3.55	7.90	3.85
5,000	2.90	2.70	6.00	2.70	1.65	1.45	5.10	2.05
Panel II: T-GARCH(1, 1)								
	low persistence				high persistence			
250	10.35	8.70	14.60	8.60	10.40	9.35	15.15	9.10
500	7.40	4.95	10.40	6.90	8.95	6.70	10.95	7.80
1,000	5.65	5.25	9.40	5.65	7.80	6.30	9.85	4.95
5,000	1.95	1.95	5.15	1.80	2.15	1.80	6.30	2.10

Table 5 reports the average coverage gap between the RT and the EP fixed-design bootstrap interval in percentage points. For different sample sizes ( $n$ ) Panel I presents the results for the low and high persistence parameterization of a GARCH(1, 1), whereas Panel II displays the results for the corresponding T-GARCH(1, 1) processes.

(1) - Table 1: 5%-VaR, Student-t innovations and 90% nominal coverage (baseline)  
(2) - Table 2: 5%-VaR, Gaussian innovations and 90% nominal coverage  
(3) - Table 3: 1%-VaR, Student-t innovations and 90% nominal coverage  
(4) - Table 4: 5%-VaR, Student-t innovations and 95% nominal coverage

Sample size		Average coverage	Av. coverage below/above	Average length	Average coverage	Av. coverage below/above	Average length
GARCH(1,1)							
		low persistence				high persistence	
250	EP	81.30	5.95/12.75	0.591	80.70	6.30/13.00	0.835
	RT	89.95	3.95/6.10	0.591	89.95	4.15/5.90	0.835
	SY	89.55	3.40/7.05	0.623	90.80	3.15/6.05	0.885
500	EP	85.00	5.95/9.05	0.442	85.05	5.45/9.50	0.605
	RT	91.05	4.20/4.75	0.442	91.25	3.95/4.80	0.605
	SY	91.40	3.15/5.45	0.459	91.05	3.05/5.90	0.629
1,000	EP	87.00	4.50/8.50	0.309	86.50	5.55/7.95	0.425
	RT	91.60	4.00/4.40	0.309	91.20	4.45/4.35	0.425
	SY	91.70	3.15/5.15	0.317	91.00	4.05/4.95	0.436
5,000	EP	87.75	6.25/6.00	0.144	87.90	5.50/6.60	0.191
	RT	90.10	5.20/4.70	0.144	89.80	5.15/5.05	0.191
	SY	90.05	5.10/4.85	0.146	89.70	4.80/5.50	0.193
T-GARCH(1,1)							
		low persistence				high persistence	
250	EP	79.30	7.25/13.45	0.142	81.00	6.45/12.55	0.292
	RT	90.60	3.75/5.65	0.142	91.65	2.70/5.65	0.292
	SY	89.30	3.70/7.00	0.149	90.15	2.70/7.15	0.306
500	EP	82.90	5.65/11.45	0.106	82.65	6.00/11.35	0.216
	RT	89.80	4.60/5.60	0.106	91.45	3.50/5.05	0.216
	SY	89.50	3.90/6.60	0.110	90.50	2.90/6.60	0.224
1,000	EP	84.50	6.00/9.50	0.077	83.25	6.80/9.95	0.158
	RT	90.30	4.70/5.00	0.077	90.55	4.55/4.90	0.158
	SY	89.90	3.95/6.15	0.079	89.70	4.20/6.10	0.162
5,000	EP	88.15	5.45/6.40	0.035	88.40	4.80/6.80	0.074
	RT	90.25	5.35/4.40	0.035	90.10	5.15/4.75	0.074
	SY	90.50	4.90/4.60	0.036	90.50	4.15/5.35	0.075

Table 6 reports distinct features of the **recursive-design** bootstrap confidence intervals for the conditional VaR at level  $\alpha = 0.05$  with **nominal coverage**  $1 - \gamma = 90\%$ . For each interval type and different sample sizes ( $n$ ), the interval's average coverage rates (in %), the average rate of the conditional VaR being below/above the interval (in %) and the interval's average length are tabulated. The bootstrap intervals are based on  $B = 2,000$  bootstrap replications and the averages are computed using  $S = 2,000$  simulations. The top presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) **Student-t innovations (6 df)**, whereas below the DGP is a Student-t T-GARCH(1,1).

With regard to Remark 3 in Section 4.1, Table 6 reports the simulation results for the recursive-design bootstrap. We refer to Appendix B for computational details. In comparison to the fixed-design approach (see Table 1) we find that the recursive-design method performs similarly in terms of average coverage for each interval type, which corresponds to the simulation results of Cavaliere et al. (2018). It is striking, however, that the intervals' average lengths are larger in the recursive-design than in the fixed-design set-up. For example, in the high persistence GARCH case (Panel I, right) for  $n = 500$  the average length in the recursive-design approach is 0.605 for the EP/RT interval compared to 0.582 in the fixed-design. As the sample size increases this difference disappears. Regarding the running time, the fixed-design bootstrap scheme operates faster than its recursive-design counterpart, e.g. in the T-GARCH high persistence case for  $n = 500$ , applying Algorithm 2 with  $B = 2,000$  takes roughly 2.7 seconds whereas its recursive-design competitor takes about 2.9 seconds per simulation.

In summary, the simulations suggest that the RT and the SY bootstrap interval work well for both bootstrap designs and that they outperform in smaller samples the AS interval in terms of average coverage even though their tails are unequally represented. In contrast, for both bootstrap designs the EP interval falls short of its nominal coverage, which is in line with the theoretical findings of Falk and Kaufmann (1991). Since the fixed RT method leads on average to shorter intervals than the corresponding SY method and its recursive-design counterpart, this suggests to favor the fixed-design RT bootstrap interval in (4.9).

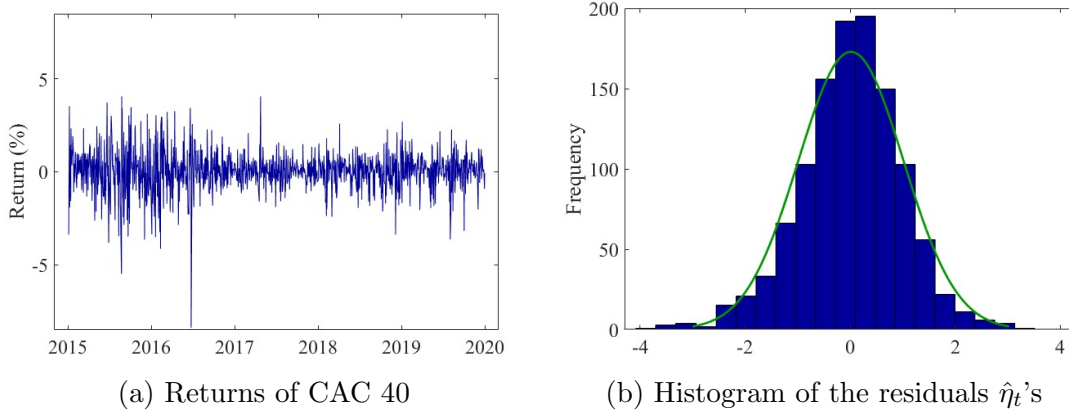


Figure 2: The returns of the French stock market index CAC 40 are plotted in (a) for the period January 1, 2015 – January 1, 2020. The histogram of the residuals is plotted in (b) after fitting a T-GARCH(1,1) model to the subperiod January 1, 2015 – July 1, 2019. A scaled normal density is superimposed.

## 5.2 Empirical Application

We analyze the French stock market index CAC 40 for the period January 1, 2015 – January 1, 2020. The index values for the period are retrieved from Yahoo Finance and daily (log-) returns (expressed in %) are computed using  $\epsilon_t = 100 \log(p_t/p_{t-1})$ , where  $p_t$  denotes the closing value of the index at trading day  $t$ . Figure 2(a) displays the resulting series of returns. We disregard the observations from July 1, 2019 onwards, which we leave for the out-of-sample evaluation, yielding  $n = 1,146$  remaining observations (i.e. January 1, 2015 - July 1, 2019). For the volatility process we consider the T-GARCH(1,1) model specified in Example 2.<sup>10</sup> Table 7 reports the corresponding point estimates with standard errors obtained by bootstrapping based on Algorithm 1. As documented in numerous studies we find that the volatility persistence is close to unity. In contrast, the point estimate  $\hat{\alpha}_n^+$  is rather small. Further,

<sup>10</sup>We also consider an Asymmetric Power GARCH model (Ding et al., 1993), i.e.  $\sigma_{t+1}^\delta = \omega_0 + \alpha_0^+(\epsilon_t^+)^{\delta} + \alpha_0^-(\epsilon_t^-)^{\delta} + \beta_0\sigma_t^\delta$  with  $\delta > 0$ , which nests the GARCH(1,1) model ( $\delta = 2$ ,  $\alpha_0^+ = \alpha_0^-$ ) and the T-GARCH(1,1) model ( $\delta = 1$ ) of Examples 1 and 2. In practice, the impact of the power  $\delta$  on the volatility is minor and the QML approach of Hamadeh and Zakoian (2011) suggests a  $\delta$  close to 1 in favor for the T-GARCH specification.

	$\hat{\omega}_n$	$\hat{\alpha}_n^+$	$\hat{\alpha}_n^-$	$\hat{\beta}_n$
point estimate	0.0292	0.0046	0.1798	0.9026
std. error	0.0109	0.0215	0.0339	0.0234

Table 7 T-GARCH(1,1) estimates for the subperiod January 1, 1998 – December 31, 2017. The standard errors are obtained by applying the fixed-design residual bootstrap with  $B = 2,000$  bootstrap replications.

we observe that  $\hat{\alpha}_n^-$  is considerably larger than  $\hat{\alpha}_n^+$  indicating a strong leverage effect, i.e. negative returns tend to increase volatility by more than positive returns of the same magnitude. Figure 2(b) plots the histogram of the residuals with the normal distribution superimposed. Further, we test the condition that the innovations are iid (see Assumption 5(i)) with the generalized run tests of Cho and White (2011).<sup>11</sup> These tests are particularly suitable in this case since they can be based on the residuals and are sensitive against a wide range of alternatives. The test statistic of the sup-norm based test is 0.40, which corresponds to a p-value of 0.27. consequently, one cannot reject the null hypothesis of iid innovations at any common significance level. Similarly, the generalized run test based on the  $L_1$ -norm cannot be rejected at a 10% significance level.

Next, we perform a rolling window analysis starting with subperiod January 1, 2015 – July 1, 2019 and ending with subperiod July 8, 2015 – January 1, 2020. We have 130 subperiods each consisting of 1,146 observations. For each rolling window period we fit a T-GARCH(1,1) model and estimate the one-period-ahead conditional VaR associated with level  $\alpha = 0.05$ . Further, we obtain the associated 95%-confidence intervals based on bootstrap and asymptotic normality. In addition to the RT intervals of the fixed- and residual-design residual bootstrap, we also compute a bootstrap interval based on moving blocks (see Algorithm 3)

<sup>11</sup>The implementation of the tests is available on the website of the first author.

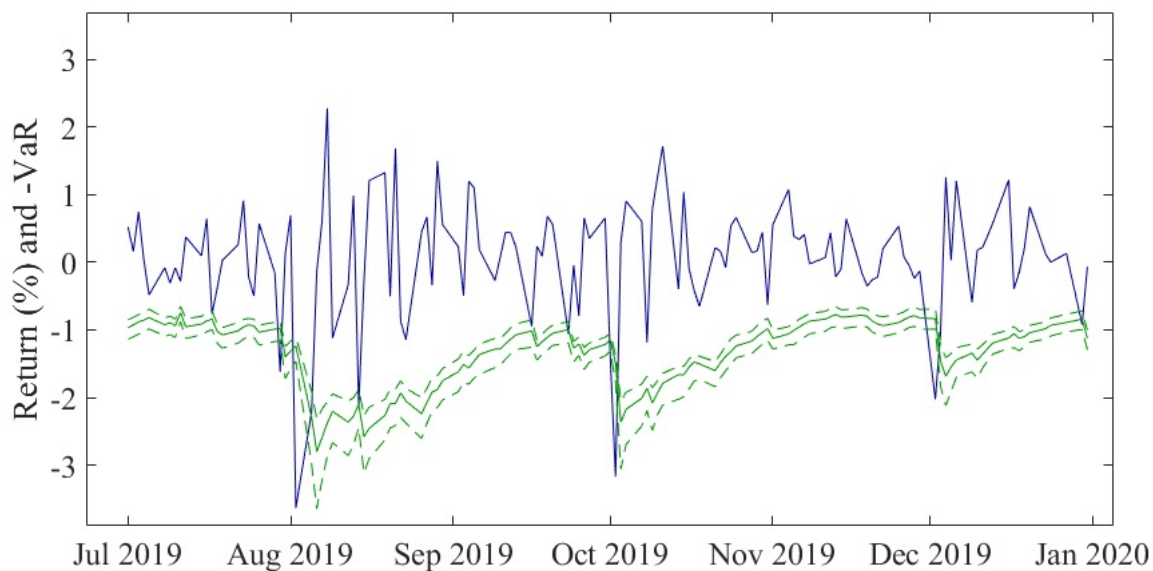


Figure 3: Returns and the estimated conditional VaR (solid) for the period June 2, 2019 – December 31, 2019. The estimation rests on the 1,146 preceding observations. Lower and upper bounds for the conditional VaR (dashed) are based on the fixed-design bootstrap scheme using the RT method with  $1 - \gamma = 95\%$ .

for which a block length of  $l = 40$  was selected. The corresponding intervals are  $[0.850, 1.136]$  (fixed-design),  $[0.834, 1.115]$  (recursive-design),  $[0.856, 1.134]$  (moving-block) and  $[0.828, 1.106]$  (asymptotic normality). Although the intervals are fairly similar, the asymptotic and recursive bootstrap intervals are shorter than the fixed-design intervals. Given its tendency to underestimate variability in finite samples, this result is unsurprising for the asymptotic interval, although for the recursive bootstrap this contrasts the simulation findings. Note that the fixed-design iid and block bootstraps produce very similar interval, which is not surprising as our conducted specification tests did not indicate any violation of the iid assumption on the innovations.

The results of the rolling window analysis are visualized in Figure 3. It plots the realized return together with (the opposite of) the estimated conditional VaR. For clarity we only indicate the lower and upper bound of the 95% RT fixed-design bootstrap interval. We observe that in more turbulent times (e.g. August, 2019), the

estimated VaR amplifies. In such volatile periods we expect the estimation risk to increase and, accordingly, we find wider bootstrap confidence intervals.

## 6 Concluding Remarks

In this paper we study the two-step estimation procedure of Francq and Zakoïan (2015) associated with the conditional VaR. In the first step, the conditional volatility parameters are estimated by QMLE, while the second step corresponds to approximating the quantile of the innovations' distribution by the empirical quantile of the residuals. A fixed-design residual bootstrap method is proposed to mimic the finite sample distribution of the two-step estimator and its consistency is proven under mild assumptions. In addition, an algorithm is provided for the construction of bootstrap intervals for the conditional VaR to take into account the uncertainty induced by estimation. Three interval types are suggested and a large-scale simulation study is conducted to investigate their performance in finite samples. We find that the equal-tailed percentile interval based on the fixed-design residual bootstrap tends to fall short of its nominal value, whereas the corresponding interval based on reversed tails yields accurate average coverage combined with the shortest average length. Although the result seems counter-intuitive at first, it is in line with the theoretical findings of Falk and Kaufmann (1991). In the simulation study we also consider the recursive-design residual bootstrap. It turns out that the recursive-design and the fixed-design bootstrap perform similar in terms of average coverage. Yet in smaller samples the fixed-design scheme leads on average to shorter intervals. Further, the interval estimation by means of the fixed-design residual bootstrap is illustrated in an empirical application to daily returns of the French stock index CAC 40.



Natural extensions of this work are encompassing other risk measures such as Expected Shortfall (Heinemann and Telg, 2018) and developing a bootstrap procedure for the one-step estimator of Francq and Zakoïan (2015). Further, it is worthwhile to consider a smoothed bootstrap version in the spirit of Hall et al. (1989), which offers potential gains in accuracy. The latter two extensions are together with the fixed-design moving block bootstrap left for future research.

## A Auxiliary Results and Proofs

### A.1 Non-bootstrap Lemmas

In analogy to  $D_t(\theta)$  and  $\hat{D}_t$  we write  $H_t(\theta) = \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'}$  and  $\hat{H}_t = \tilde{H}_t(\hat{\theta}_n)$  with  $\tilde{H}_t(\theta) = \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'}$ . Further, we introduce

$$\begin{aligned} S_t &= \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)}, & T_t &= \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)}, \\ U_t &= \sup_{\theta \in \mathcal{V}(\theta_0)} \|D_t(\theta)\|, & V_t &= \sup_{\theta \in \mathcal{V}(\theta_0)} \|H_t(\theta)\|, \end{aligned} \tag{A.1}$$

and stress that  $\{S_t\}$ ,  $\{T_t\}$ ,  $\{U_t\}$  and  $\{V_t\}$  are strictly stationary and ergodic processes (cf. Francq and Zakoïan, 2011, p. 182/405).

**Lemma 1.** *Suppose Assumptions 1, 2, 3, 4(i), 5(i), 6 and 9(i) hold with  $a = -1$ . Then, we have  $\sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}_n(x) - F(x)| \xrightarrow{a.s.} 0$ .*

*Proof.* The proof follows Berkes and Horváth (2003, Theorem 2.1 & Lemma 5.1) and consists of three parts. First, we show that for any  $\varepsilon > 0$  there is a  $\tau > 0$  such that

$$\begin{aligned} \limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x \tilde{\sigma}_t(\theta) / \sigma_t(\theta_0)\}} - F(x) \right| \\ \leq 2 \left( F(x + \varepsilon|x|) - F(x - \varepsilon|x|) \right) \end{aligned} \tag{A.2}$$

almost surely for any  $x \in \mathbb{R}$ , where  $\mathcal{V}_\tau(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| \leq \tau\}$ . In the second step, we show  $\hat{\mathbb{F}}_n(x) \xrightarrow{a.s.} F(x)$  for any  $x \in \mathbb{R}$  using (A.2) and thereafter prove  $\sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}_n(x) - F(x)| \xrightarrow{a.s.} 0$ .

Let  $\varepsilon > 0$  and note that  $\sigma_t \geq \underline{\omega}$  by Assumption 3. Together with Assumption 4(i),

there exists a random variable  $n_0$  such that  $C_1\rho^t/\sigma_t(\theta_0) \leq \varepsilon$  for all  $t > n_0$ . Then

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\tilde{\sigma}_t(\theta)/\sigma_t(\theta_0)\}} &\leq \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta)/\sigma_t(\theta_0) + |x|C_1\rho^t/\sigma_t(\theta_0)\}} \\ &\leq \frac{n_0}{n} + \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}} \end{aligned}$$

holds almost surely. Let  $\tau > 0$  (to be specified); for any  $\theta \in \mathcal{V}_\tau(\theta_0)$  we get

$$\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}} \leq \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}}$$

almost surely. The uniform ergodic theorem for strictly stationary sequences (cf. Francq and Zakoïan, 2011, p. 181), henceforth called the uniform ergodic theorem, and Assumptions 2, 3 and 5(i) yield

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}} &\xrightarrow{a.s.} \mathbb{E} \mathbb{1}_{\{\eta_t \leq \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}} \\ &= \mathbb{E} F\left(\sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\right). \end{aligned}$$

Further, Assumptions 3 and 9(i) with  $a = -1$  imply  $\lim_{\tau \rightarrow 0} \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) = x$  almost surely. Thus, the dominated convergence theorem entails

$$\lim_{\tau \rightarrow 0} \mathbb{E} F\left(\sup_{\theta \in \mathcal{V}_\tau(\theta_0)} x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\right) = F(x + \varepsilon|x|).$$

Putting the results together, we get that for every  $\varepsilon > 0$ , there is a  $\tau > 0$  such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\tilde{\sigma}_t(\theta)/\sigma_t(\theta_0)\}} \leq F(x) + 2\left(F(x + \varepsilon|x|) - F(x)\right)$$

almost surely for any  $x \in \mathbb{R}$ . Similarly it can be shown that for every  $\varepsilon > 0$ , there is

a  $\tau > 0$  such that

$$\liminf_{n \rightarrow \infty} \inf_{\theta \in \mathcal{V}_\tau(\theta_0)} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x \tilde{\sigma}_t(\theta) / \sigma_t(\theta_0)\}} \geq F(x) - 2 \left( F(x) - F(x - \varepsilon|x|) \right).$$

almost surely for any  $x \in \mathbb{R}$ . Combining both results, we establish (A.2).

Next, we show  $\hat{\mathbb{F}}_n(x) \xrightarrow{a.s.} F(x)$  for any  $x \in \mathbb{R}$ . Let  $\delta > 0$ ; by continuity of  $F$  (see Assumption 5(i)), there is a  $\varepsilon > 0$  such that  $|F(x + \varepsilon|x|) - F(x - \varepsilon|x|)| < \delta/2$ . Employing equation (A.2), there are  $\tau > 0$  and a random variable  $n_1$  such that

$$\sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x \tilde{\sigma}_t(\theta) / \sigma_t(\theta_0)\}} - F(x) \right| < \delta$$

for all  $n \geq n_1$ . Since  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  by Theorem 1 there is a random variable  $n_2$  such that  $\hat{\theta}_n \in \mathcal{V}_\tau(\theta_0)$  for all  $n \geq n_2$ . Thus,

$$|\hat{\mathbb{F}}_n(x) - F(x)| \leq \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x \tilde{\sigma}_t(\theta) / \sigma_t(\theta_0)\}} - F(x) \right| < \delta$$

for all  $n \geq \max\{n_1, n_2\}$ , which establishes  $\hat{\mathbb{F}}_n(x) \xrightarrow{a.s.} F(x)$  for any  $x \in \mathbb{R}$ . Using Pólya's lemma (cf. Roussas, 1997, p. 206), we establish  $\sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}_n(x) - F(x)| \xrightarrow{a.s.} 0$  completing the proof.  $\square$

**Lemma 2.** *Suppose Assumptions 1–3, 4(i) and 5(i) hold.*

- (i) *If in addition Assumptions 4(ii) and 9(ii) hold with  $b = 1$ , then  $\hat{\Omega}_n \xrightarrow{a.s.} \Omega$ .*
- (ii) *If in addition Assumptions 4(ii) and 9(ii) hold with  $b = 2$ , then  $\hat{J}_n \xrightarrow{a.s.} J$ .*
- (iii) *If in addition Assumptions 4(ii) and 9(iii) hold with  $c = 1$ , then  $\frac{1}{n} \sum_{t=1}^n \hat{H}_t \xrightarrow{a.s.} \mathbb{E}[H_t]$ .*

(iv) If in addition Assumptions 5(iii) and 9(i) hold with  $a = 4$ , then we have

$$\frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^m \mathbb{1}_{\{l \leq \hat{\eta}_t < u\}} \xrightarrow{a.s.} \mathbb{E}[\eta_t^m \mathbb{1}_{\{l \leq \eta_t < u\}}] \text{ for } m \in \{0, 1, 2, 3, 4\} \text{ and } l < u.$$

(v) If in addition Assumptions 4 and 9(i)-(ii) hold with  $a = \pm 2$  and  $b = 4$ , then

$$\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{l \leq \sqrt{n}(\tilde{\psi}_t - 1) < u\}} \left( \sqrt{n}(\tilde{\psi}_t - 1) \right)^m \xrightarrow{a.s.} \mathbb{E} \left[ \mathbb{1}_{\{l \leq D'_t(v_1 - v_2) < u\}} (D'_t(v_1 - v_2))^m \right]$$

for  $v_1, v_2 \in \mathbb{R}^r$ ,  $m \in \{0, 1, 2, 3, 4\}$  and  $l < u$  with  $\tilde{\psi}_t = \frac{\tilde{\sigma}_t(\hat{\theta}_n + n^{-1/2}v_1)}{\tilde{\sigma}_t(\hat{\theta}_n + n^{-1/2}v_2)}$ .

*Proof.* Consider the first statement and expand

$$\frac{1}{n} \sum_{t=1}^n \hat{D}_t = \underbrace{\frac{1}{n} \sum_{t=1}^n D_t(\hat{\theta}_n)}_I + \underbrace{\frac{1}{n} \sum_{t=1}^n (\tilde{D}_t(\hat{\theta}_n) - D_t(\hat{\theta}_n))}_{II}.$$

Focusing on  $I$ , we take  $\varepsilon > 0$  and let  $e_1, \dots, e_r$  denote the unit vectors spanning  $\mathbb{R}^r$ .

Since  $D_t(\theta)$  is continuous in  $\theta$  we can take  $\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$  such that

$$\mathbb{E}[e'_i D_t] - \varepsilon < \mathbb{E} \left[ \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] \leq \mathbb{E} \left[ \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] < \mathbb{E}[e'_i D_t] + \varepsilon$$

for all  $i = 1, \dots, r$ . Since  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  (Theorem 1), we have  $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$  almost surely.

Together with the uniform ergodic theorem we obtain

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) &\stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \xrightarrow{a.s.} \mathbb{E} \left[ \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] < \mathbb{E}[e'_i D_t] + \varepsilon \\ \frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) &\stackrel{a.s.}{\geq} \frac{1}{n} \sum_{t=1}^n \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \xrightarrow{a.s.} \mathbb{E} \left[ \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] > \mathbb{E}[e'_i D_t] - \varepsilon. \end{aligned}$$

Taking  $\varepsilon \searrow 0$  establishes  $\frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) \xrightarrow{a.s.} \mathbb{E}[e'_i D_t]$  for all  $i$  yielding  $I \xrightarrow{a.s.} \mathbb{E}[D_t] = \Omega$ .

Regarding  $II$ , we note that for each  $\theta \in \Theta$ , Assumption 4 implies

$$\begin{aligned}
\|\tilde{D}_t(\theta) - D_t(\theta)\| &= \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} - \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \\
&= \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \left( \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} - \frac{\partial \sigma_t(\theta)}{\partial \theta} \right) + \frac{\sigma_t(\theta) - \tilde{\sigma}_t(\theta)}{\tilde{\sigma}_t(\theta)} \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \\
&\leq \frac{1}{\tilde{\sigma}_t(\theta)} \left\| \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} - \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| + \frac{|\sigma_t(\theta) - \tilde{\sigma}_t(\theta)|}{\tilde{\sigma}_t(\theta)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \\
&\leq \frac{C_1 \rho^t}{\underline{\omega}} + \frac{C_1 \rho^t}{\underline{\omega}} \|D_t(\theta)\| = \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|D_t(\theta)\|).
\end{aligned} \tag{A.3}$$

We obtain

$$\|II\| \leq \frac{1}{n} \sum_{t=1}^n \|\tilde{D}_t(\hat{\theta}_n) - D_t(\hat{\theta}_n)\| \leq \frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + \|D_t(\hat{\theta}_n)\|) \stackrel{a.s.}{\leq} \frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + U_t).$$

For each  $\varepsilon > 0$ , Markov's inequality entails

$$\sum_{t=1}^{\infty} \mathbb{P}[\rho^t (1 + U_t) > \varepsilon] \leq \sum_{t=1}^{\infty} \rho^t \frac{1 + \mathbb{E}[U_t]}{\varepsilon} = \frac{1 + \mathbb{E}[U_t]}{\varepsilon(1 - \rho)} < \infty$$

since  $\rho \in (0, 1)$  and  $\mathbb{E}[U_t] < \infty$  by Assumption 9(ii). The Borel-Cantelli lemma implies

$$0 = \mathbb{P} \left[ \lim_{t \rightarrow \infty} \bigcup_{s=t}^{\infty} \left\{ \rho^s (1 + U_s) > \varepsilon \right\} \right] \geq \mathbb{P} \left[ \lim_{t \rightarrow \infty} \rho^t (1 + U_t) > \varepsilon \right] \tag{A.4}$$

and hence  $\rho^t (1 + U_t) \rightarrow 0$  almost surely. Cesàro's lemma yields  $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + U_t) \xrightarrow{a.s.} 0$  and hence  $\|II\| \xrightarrow{a.s.} 0$ , which validates the first statement.

Consider the second statement and expand

$$\frac{1}{n} \sum_{t=1}^n \hat{D}_t \hat{D}_t' = \underbrace{\frac{1}{n} \sum_{t=1}^n D_t(\hat{\theta}_n) D_t'(\hat{\theta}_n)}_{III} + \underbrace{\frac{1}{n} \sum_{t=1}^n \left( \tilde{D}_t(\hat{\theta}_n) \tilde{D}_t'(\hat{\theta}_n) - D_t(\hat{\theta}_n) D_t'(\hat{\theta}_n) \right)}_{IV}.$$

We focus on *III* and let  $\varepsilon > 0$ . Since  $D_t(\theta)D_t(\theta)'$  is continuous in  $\theta$  we can take  $\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$  such that

$$\begin{aligned} \mathbb{E}[e'_i D_t D'_t e_j] - \varepsilon &< \mathbb{E} \left[ \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j \right] \\ &\leq \mathbb{E} \left[ \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j \right] < \mathbb{E}[e'_i D_t D'_t e_j] + \varepsilon \end{aligned}$$

for all  $i, j = 1, \dots, r$ . Since  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  by Theorem 1, we have  $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$  almost surely.

Together with the uniform ergodic theorem we obtain

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) e_j &\leq \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j \\ &\xrightarrow{a.s.} \mathbb{E} \left[ \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j \right] < \mathbb{E}[e'_i D_t D'_t e_j] + \varepsilon \\ \frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) e_j &\geq \frac{1}{n} \sum_{t=1}^n \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j \\ &\xrightarrow{a.s.} \mathbb{E} \left[ \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) D'_t(\theta) e_j \right] > \mathbb{E}[e'_i D_t D'_t e_j] - \varepsilon \end{aligned}$$

Taking  $\varepsilon \searrow 0$  establishes  $\frac{1}{n} \sum_{t=1}^n e'_i D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) e_j \xrightarrow{a.s.} \mathbb{E}[e'_i D_t D'_t e_j]$  for all pairs  $(i, j)$  yielding *III*  $\xrightarrow{a.s.} \mathbb{E}[D_t D'_t] = J$ . Consider *IV*; using (A.3) and the elementary inequality

$$\|xx' - yy'\| \leq \|x - y\|^2 + 2\|x - y\| \|y\| \quad (\text{A.5})$$

for all  $x, y \in \mathbb{R}^m$  with  $m \in \mathbb{N}$ , we obtain for  $\theta \in \Theta$

$$\begin{aligned}
& \left\| \tilde{D}_t(\theta) \tilde{D}'_t(\theta) - D_t(\theta) D'_t(\theta) \right\| \\
& \leq \left\| \tilde{D}_t(\theta) - D_t(\theta) \right\|^2 + 2 \left\| \tilde{D}_t(\theta) - D_t(\theta) \right\| \left\| D_t(\theta) \right\| \\
& \leq \frac{C_1^2}{\underline{\omega}^2} \rho^{2t} \left( 1 + \left\| D_t(\theta) \right\| \right)^2 + \frac{2C_1}{\underline{\omega}} \rho^t \left( 1 + \left\| D_t(\theta) \right\| \right) \left\| D_t(\theta) \right\| \\
& \leq \frac{C_1^2}{\underline{\omega}^2} \rho^t \left( 1 + \left\| D_t(\theta) \right\| \right)^2 + \frac{2C_1}{\underline{\omega}} \rho^t \left( 1 + \left\| D_t(\theta) \right\| \right)^2 \\
& = \left( \frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \rho^t \left( 1 + \left\| D_t(\theta) \right\| \right)^2.
\end{aligned} \tag{A.6}$$

Hence, we get

$$\begin{aligned}
\|IV\| & \leq \frac{1}{n} \sum_{t=1}^n \left\| \tilde{D}_t(\hat{\theta}_n) \tilde{D}'_t(\hat{\theta}_n) - D_t(\hat{\theta}_n) D'_t(\hat{\theta}_n) \right\| \leq \left( \frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \frac{1}{n} \sum_{t=1}^n \rho^t \left( 1 + \left\| D_t(\hat{\theta}_n) \right\| \right)^2 \\
& \stackrel{a.s.}{\leq} \left( \frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \frac{1}{n} \sum_{t=1}^n \rho^t (1 + U_t)^2.
\end{aligned} \tag{A.7}$$

For each  $\varepsilon > 0$ , Markov's inequality yields

$$\sum_{t=1}^{\infty} \mathbb{P} \left[ \rho^t (1 + U_t)^2 > \varepsilon \right] \leq \sum_{t=1}^{\infty} \rho^{t/2} \frac{1 + \mathbb{E}[U_t]}{\sqrt{\varepsilon}} = \frac{1 + \mathbb{E}[U_t]}{\sqrt{\varepsilon}(1 - \sqrt{\rho})} < \infty$$

and  $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + U_t)^2 \xrightarrow{a.s.} 0$  follows from combining the Borel-Cantelli lemma with Cesàro's lemma. Hence,  $\|IV\| \xrightarrow{a.s.} 0$ , which validates the second statement.

Consider the third statement and expand

$$\frac{1}{n} \sum_{t=1}^n \hat{H}_t = \underbrace{\frac{1}{n} \sum_{t=1}^n H_t(\hat{\theta}_n)}_V + \underbrace{\frac{1}{n} \sum_{t=1}^n \left( \tilde{H}_t(\hat{\theta}_n) - H_t(\hat{\theta}_n) \right)}_{VI}$$

We focus on  $V$  and let  $\varepsilon > 0$ . Since  $H_t(\theta)$  is continuous in  $\theta$  we can take  $\mathcal{V}_\varepsilon(\theta_0) \subseteq$



$\mathcal{V}(\theta_0)$  such that

$$\mathbb{E}[e'_i H_t e_j] - \varepsilon < \mathbb{E}\left[\inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j\right] \leq \mathbb{E}\left[\sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j\right] < \mathbb{E}[e'_i H_t e_j] + \varepsilon$$

for all  $i, j \in \{1, \dots, r\}$ . As  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  by Theorem 1, we have  $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$  almost surely.

Together with the uniform ergodic theorem we obtain

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n e'_i H_t(\hat{\theta}_n) e_j &\leq \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j \xrightarrow{a.s.} \mathbb{E}\left[\sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j\right] < \mathbb{E}[e'_i H_t e_j] + \varepsilon \\ \frac{1}{n} \sum_{t=1}^n e'_i H_t(\hat{\theta}_n) e_j &\geq \frac{1}{n} \sum_{t=1}^n \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j \xrightarrow{a.s.} \mathbb{E}\left[\inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i H_t(\theta) e_j\right] > \mathbb{E}[e'_i H_t e_j] - \varepsilon \end{aligned}$$

Taking  $\varepsilon \searrow 0$  establishes  $\frac{1}{n} \sum_{t=1}^n e'_i H_t(\hat{\theta}_n) e_j \xrightarrow{a.s.} \mathbb{E}[e'_i H_t e_j]$  for all pairs  $(i, j)$  yielding  $V \xrightarrow{a.s.} \mathbb{E}[H_t]$ . Regarding  $VI$ , we note that

$$\begin{aligned} \|\tilde{H}_t(\theta) - H_t(\theta)\| &= \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} - \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| \\ &= \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \left( \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right) + \frac{\sigma_t(\theta) - \tilde{\sigma}_t(\theta)}{\tilde{\sigma}_t(\theta)} \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| \\ &\leq \frac{1}{\tilde{\sigma}_t(\theta)} \left\| \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| + \frac{|\sigma_t(\theta) - \tilde{\sigma}_t(\theta)|}{\tilde{\sigma}_t(\theta)} \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| \\ &\leq \frac{C_1 \rho^t}{\underline{\omega}} + \frac{C_1 \rho^t}{\underline{\omega}} \|H_t(\theta)\| = \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|H_t(\theta)\|) \end{aligned} \tag{A.8}$$

for each  $\theta \in \Theta$ . We obtain

$$\|VI\| \leq \frac{1}{n} \sum_{t=1}^n \|\tilde{H}_t(\hat{\theta}_n) - H_t(\hat{\theta}_n)\| \leq \frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + \|H_t(\hat{\theta}_n)\|) \stackrel{a.s.}{\leq} \frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + V_t).$$

For each  $\varepsilon > 0$ , Markov's inequality yields

$$\sum_{t=1}^{\infty} \mathbb{P}\left[\rho^t (1 + V_t) > \varepsilon\right] \leq \sum_{t=1}^{\infty} \rho^t \frac{1 + \mathbb{E}[V_t]}{\varepsilon} = \frac{1 + \mathbb{E}[V_t]}{\varepsilon(1 - \rho)} < \infty$$

and  $\frac{1}{n} \sum_{t=1}^n \rho^t(1 + V_t) \xrightarrow{a.s.} 0$  follows from combining the Borel-Cantelli lemma with Cesàro's lemma. Hence,  $\|VI\| \xrightarrow{a.s.} 0$ , which validates the third statement.

Consider the fourth statement; let  $m \in \{0, 1, 2, 3, 4\}$  and take  $l, u \in \mathbb{R}$  such that  $l < u$ . We employ the partial integration formula

$$G(u-)H(u-) - G(l-)H(l-) = \int_{[l,u)} G(t-) dH(t) + \int_{[l,u)} H(s) dG(s) \quad (\text{A.9})$$

with  $G$  and  $H$  both right-continuous functions being locally of bounded variation to expand

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^m \mathbb{1}_{\{l \leq \hat{\eta}_t < u\}} - \mathbb{E}[\eta_t^m \mathbb{1}_{\{l \leq \eta_t < u\}}] &= \int_{[l,u)} x^m d\hat{\mathbb{F}}_n(x) - \int_{[l,u)} x^m dF(x) \\ &= u^m (\hat{\mathbb{F}}_n(u-) - F(u)) - l^m (\hat{\mathbb{F}}_n(l-) - F(l)) + \int_{[l,u)} (\hat{\mathbb{F}}_n(x) - F(x)) dx^m. \end{aligned}$$

Lemma 1 implies  $\hat{\mathbb{F}}_n(u-) \xrightarrow{a.s.} F(u)$  and  $\hat{\mathbb{F}}_n(l-) \xrightarrow{a.s.} F(l)$  and together with the dominated convergence theorem yields  $\int_{[l,u)} (\hat{\mathbb{F}}_n(x) - F(x)) dx^m \xrightarrow{a.s.} 0$ . Thus,

$$\frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^m \mathbb{1}_{\{l \leq \hat{\eta}_t < u\}} \xrightarrow{a.s.} \mathbb{E}[\eta_t^m \mathbb{1}_{\{l \leq \eta_t < u\}}]$$

for  $m \in \{0, 1, 2, 3, 4\}$  and  $l, u \in \mathbb{R}$ . Since  $\mathbb{E}[|\eta_t|^m] < \infty$  and  $\mathbb{E}[\eta_t^m \mathbb{1}_{\{l \leq \eta_t < u\}}] = \int_l^u x^m f(x) dx$  is continuous in  $l$  and  $u$  it is easy to see that the result extends to  $l = -\infty$  and  $u = \infty$ , which validates the fourth statement.

Consider the fifth statement, whose proof follows the general steps of the proof of Lemma 1 and the fourth statement. Define

$$\hat{\mathbb{G}}_n(x) = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\sqrt{n}(\tilde{\psi}_t - 1) \leq x\}} \quad \text{and} \quad G(x) = \mathbb{P}[D'_t(v_1 - v_2) \leq x].$$

First, we show that for any  $\varepsilon > 0$  there is a  $\tau > 0$  such that almost surely

$$\limsup_{n \rightarrow \infty} \sup_{\theta_1, \theta_2 \in \mathcal{V}_\tau(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} (\tilde{D}'_t(\theta_1)v_1 - \tilde{D}'_t(\theta_2)v_2) \leq x \right\}} - G(x) \right| \leq 2(G(x + \Delta\varepsilon) - G(x - \Delta\varepsilon)) \quad (\text{A.10})$$

for any  $x \in \mathbb{R}$ , where  $\Delta = |x| + \|v_1\| + \|v_2\|$  and  $\mathcal{V}_\tau(\theta_0) = \{\theta \in \Theta : \|\theta - \theta_0\| \leq \tau\}$ . Then, we show  $\hat{\mathbb{G}}_n(x) \xrightarrow{a.s.} G(x)$  for any  $x \in \mathbb{R}$  and  $\sup_{x \in \mathbb{R}} |\hat{\mathbb{G}}_n(x) - G(x)| \xrightarrow{a.s.} 0$ . Last, we prove  $\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\sqrt{n}(\tilde{\psi}_t - 1) < u\}} (\sqrt{n}(\tilde{\psi}_t - 1))^m \xrightarrow{a.s.} \mathbb{E} [\mathbb{1}_{\{l \leq D'_t(v_1 - v_2) < u\}} (D'_t(v_1 - v_2))^m]$ .

Let  $\varepsilon > 0$  and set  $\tau > 0$  sufficiently small such that  $\mathcal{V}_\tau(\theta_0) \subset \mathcal{V}(\theta_0)$ . Regarding the initial conditions Assumption 4(i) implies

$$\begin{aligned} \left| \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} - \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right| &= \left| \frac{\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \frac{\sigma_t(\theta_2) - \tilde{\sigma}_t(\theta_2)}{\tilde{\sigma}_t(\theta_2)} \right| \\ &\leq \frac{|\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)|}{\tilde{\sigma}_t(\theta_2)} + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \frac{|\sigma_t(\theta_2) - \tilde{\sigma}_t(\theta_2)|}{\tilde{\sigma}_t(\theta_2)} \\ &\leq \frac{C_1 \rho^t}{\underline{\omega}} + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \frac{C_1 \rho^t}{\underline{\omega}} = \frac{C_1 \rho^t}{\underline{\omega}} \left( 1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right) \end{aligned} \quad (\text{A.11})$$

for any  $\theta_1, \theta_2 \in \Theta$  and together with (A.3) we find

$$\begin{aligned} &\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} (\tilde{D}'_t(\theta_1)v_1 - \tilde{D}'_t(\theta_2)v_2) \leq x \right\}} \\ &= \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ D'_t(\theta_1)v_1 - D'_t(\theta_2)v_2 - x \frac{\sigma_t(\theta_2)}{\sigma_t(\theta_1)} \leq x \left( \frac{\tilde{\sigma}_t(\theta_2)}{\tilde{\sigma}_t(\theta_1)} - \frac{\sigma_t(\theta_2)}{\sigma_t(\theta_1)} \right) + (D_t(\theta_1) - \tilde{D}_t(\theta_1))' v_1 + (\tilde{D}_t(\theta_2) - D_t(\theta_2))' v_2 \right\}} \\ &\leq \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ D'_t(\theta_1)v_1 - D'_t(\theta_2)v_2 - x \frac{\sigma_t(\theta_2)}{\sigma_t(\theta_1)} \leq |x| \frac{C_1 \rho^t}{\underline{\omega}} \left( 1 + \frac{\sigma_t(\theta_2)}{\sigma_t(\theta_1)} \right) + \|v_1\| \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|D_t(\theta_1)\|) + \|v_2\| \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|D_t(\theta_2)\|) \right\}} \\ &\leq \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ D'_t(\theta_1)u - D'_t(\theta_2)v - x \frac{\sigma_t(\theta_2)}{\sigma_t(\theta_1)} \leq |x| \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) + (\|v_1\| + \|v_2\|) \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right\}} \end{aligned}$$

for all  $\theta_1, \theta_2 \in \mathcal{V}_\tau(\theta_0)$ . We have  $\rho^t(1 + U_t) \xrightarrow{a.s.} 0$  by (A.4). Further, for each  $\varepsilon > 0$ ,

Markov's and Hölder's inequality together with Assumption 9(i) entail

$$\sum_{t=1}^{\infty} \mathbb{P}[\rho^t(1 + S_t T_t) > \varepsilon] \leq \sum_{t=1}^{\infty} \rho^t \frac{1 + \mathbb{E}[S_t T_t]}{\varepsilon} \leq \frac{1 + \mathbb{E}[S_t^2]^{\frac{1}{2}} \mathbb{E}[T_t^2]^{\frac{1}{2}}}{\varepsilon(1 - \rho)} < \infty.$$

The Borel-Cantelli lemma implies  $\rho^t(1 + S_t T_t) \xrightarrow{a.s.} 0$ . Hence, there exists a random variable  $n_0$  such that  $\frac{C_1 \rho^t}{\underline{\omega}}(1 + U_t) \leq \varepsilon$  and  $\frac{C_1 \rho^t}{\underline{\omega}}(1 + S_t T_t) \leq \varepsilon$  for all  $t > n_0$ . It follows that almost surely

$$\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} (\tilde{D}'_t(\theta_1)v_1 - \tilde{D}'_t(\theta_2)v_2) \leq x \right\}} \leq \frac{n_0}{n} + \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \inf_{\theta_1, \theta_2 \in \mathcal{V}_\tau(\theta_0)} (D'_t(\theta_1)v_1 - D'_t(\theta_2)v_2 - x \frac{\sigma_t(\theta_2)}{\sigma_t(\theta_1)}) \leq \Delta\varepsilon \right\}}$$

for all  $\theta_1, \theta_2 \in \mathcal{V}_\tau(\theta_0)$ . The uniform ergodic theorem and Assumptions 2 and 3 yield

$$\begin{aligned} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \inf_{\theta_1, \theta_2 \in \mathcal{V}_\tau(\theta_0)} (D'_t(\theta_1)v_1 - D'_t(\theta_2)v_2 - x \frac{\sigma_t(\theta_2)}{\sigma_t(\theta_1)}) \leq \Delta\varepsilon \right\}} \\ \xrightarrow{a.s.} \mathbb{E} \left[ \mathbb{1}_{\left\{ \inf_{\theta_1, \theta_2 \in \mathcal{V}_\tau(\theta_0)} (D'_t(\theta_1)v_1 - D'_t(\theta_2)v_2 - x \frac{\sigma_t(\theta_2)}{\sigma_t(\theta_1)}) \leq \Delta\varepsilon \right\}} \right]. \end{aligned}$$

The dominated convergence theorem entails

$$\lim_{\tau \rightarrow 0} \mathbb{E} \left[ \mathbb{1}_{\left\{ \inf_{\theta_1, \theta_2 \in \mathcal{V}_\tau(\theta_0)} (D'_t(\theta_1)v_1 - D'_t(\theta_2)v_2 - x \frac{\sigma_t(\theta_2)}{\sigma_t(\theta_1)}) \leq \Delta\varepsilon \right\}} \right] = \mathbb{E}[\mathbb{1}_{\{D'_t(v_1 - v_2) - x \leq \Delta\varepsilon\}}] = G(x + \Delta\varepsilon).$$

Putting the results together, we get that for every  $\varepsilon > 0$ , there is a  $\tau > 0$  such that

$$\limsup_{n \rightarrow \infty} \sup_{\theta_1, \theta_2 \in \mathcal{V}_\tau(\theta_0)} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} (\tilde{D}'_t(\theta_1)v_1 - \tilde{D}'_t(\theta_2)v_2) \leq x \right\}} \leq G(x) + 2(G(x + \Delta\varepsilon) - G(x))$$

almost surely for any  $x \in \mathbb{R}$ . Similarly it can be shown that for every  $\varepsilon > 0$ , there is

a  $\tau > 0$  such that

$$\liminf_{n \rightarrow \infty} \sup_{\theta_1, \theta_2 \in \mathcal{V}_\tau(\theta_0)} \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} (\tilde{D}'_t(\theta_1)v_1 - \tilde{D}'_t(\theta_2)v_2) \leq x \right\}} \geq G(x) - 2(G(x) - G(x - \Delta\varepsilon))$$

almost surely for any  $x \in \mathbb{R}$ . Combining both results establishes (A.10).

Next, we show  $\hat{\mathbb{G}}_n(x) \xrightarrow{a.s.} G(x)$  for any  $x \in \mathbb{R}$ . Let  $\delta > 0$ ; by continuity of  $G$ , there is a  $\varepsilon > 0$  such that  $|G(x + \Delta\varepsilon) - G(x - \Delta\varepsilon)| < \delta/2$ . Employing equation (A.10), there are  $\tau > 0$  and a random variable  $n_1$  such that

$$\sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} (\tilde{D}'_t(\theta_1)v_1 - \tilde{D}'_t(\theta_2)v_2) \leq x \right\}} - G(x) \right| < \delta$$

for all  $n \geq n_1$ . In addition, the mean value theorem implies

$$\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\sqrt{n}(\tilde{\psi}_t - 1) \leq x\}} = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\tilde{\sigma}_t(\dot{\theta}_n)}{\tilde{\sigma}_t(\ddot{\theta}_n)} (\tilde{D}'_t(\dot{\theta}_n)v_1 - \tilde{D}'_t(\ddot{\theta}_n)v_2) \leq x \right\}} \quad (\text{A.12})$$

with  $\dot{\theta}_n$  lying between  $\hat{\theta}_n$  and  $\hat{\theta}_n + n^{-1/2}v_1$  and  $\ddot{\theta}_n$  lying between  $\hat{\theta}_n$  and  $\hat{\theta}_n + n^{-1/2}v_2$ . Since  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  by Theorem 1 there is a random variable  $n_2$  such that  $\dot{\theta}_n, \ddot{\theta}_n \in \mathcal{V}_\tau(\theta_0)$  for all  $n \geq n_2$ . Thus,

$$|\hat{\mathbb{G}}_n(x) - G(x)| \leq \sup_{\theta \in \mathcal{V}_\tau(\theta_0)} \left| \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\left\{ \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} (\tilde{D}'_t(\theta_1)v_1 - \tilde{D}'_t(\theta_2)v_2) \leq x \right\}} - G(x) \right| < \delta$$

for all  $n \geq \max\{n_1, n_2\}$ , which establishes  $\hat{\mathbb{G}}_n(x) \xrightarrow{a.s.} G(x)$  for any  $x \in \mathbb{R}$ . Using Pólya's lemma (cf. Roussas, 1997, p. 206), we establish  $\sup_{x \in \mathbb{R}} |\hat{\mathbb{G}}_n(x) - G(x)| \xrightarrow{a.s.} 0$ .

Next, let  $l, u \in \mathbb{R}$  with  $l < u$ . We use the partial integration formula (A.9) to expand

$$\begin{aligned}
& \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{l \leq \sqrt{n}(\tilde{\psi}_t - 1) < u\}} \left( \sqrt{n}(\tilde{\psi}_t - 1) \right)^m - \mathbb{E} \left[ \mathbb{1}_{\{l \leq D'_t(v_1 - v_2) < u\}} (D'_t(v_1 - v_2))^m \right] \\
&= \int_{[l, u)} x^m d\hat{\mathbb{G}}_n(x) - \int_{[l, u)} x^m dG(x) \\
&= u^m (\hat{\mathbb{G}}_n(u-) - G(u)) - l^m (\hat{\mathbb{G}}_n(l-) - G(l)) + \int_{[l, u)} (\hat{\mathbb{G}}_n(x) - G(x)) dx^m.
\end{aligned}$$

We have  $\hat{\mathbb{G}}_n(u-) \xrightarrow{a.s.} G(u)$  and  $\hat{\mathbb{G}}_n(l-) \xrightarrow{a.s.} G(l)$  and together with the dominated convergence theorem yields  $\int_{[l, u)} (\hat{\mathbb{G}}_n(x) - G(x)) dx^m \xrightarrow{a.s.} 0$ . Thus, we establish

$$\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{l \leq \sqrt{n}(\tilde{\psi}_t - 1) < u\}} \left( \sqrt{n}(\tilde{\psi}_t - 1) \right)^m \xrightarrow{a.s.} \mathbb{E} \left[ \mathbb{1}_{\{l \leq D'_t(v_1 - v_2) < u\}} (D'_t(v_1 - v_2))^m \right].$$

Let  $g(x)$  be the corresponding density of  $G(x)$ . As  $\mathbb{E}[|D'_t(v_1 - v_2)|^m] \leq \|v_1 - v_2\|^m \mathbb{E}[U_t^m] < \infty$  and  $\mathbb{E}[\mathbb{1}_{\{l \leq D'_t(v_1 - v_2) < u\}} (D'_t(v_1 - v_2))^m] = \int_l^u x^m g(x) dx$  is continuous in  $l$  and  $u$  it is easy to see that the result extends to  $l = -\infty$  and  $u = \infty$ , which validates the fifth statement and completes the proof.  $\square$

**Lemma 3.** *Suppose Assumptions 1–9 hold with  $a = \pm 6$ ,  $b = 6$  and  $c = 2$  and let  $\mathcal{I}_n = (\xi_\alpha - a_n, \xi_\alpha + a_n)$  with  $a_n \sim n^{-\varrho} \log n$  for some  $\varrho \in (0, 1)$ . Then, we have*

$$\sup_{x, y \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \xrightarrow{P} 0.$$

*Replacing any  $\hat{\mathbb{F}}_n(\cdot)$  by  $\hat{\mathbb{F}}_n(\cdot -)$  does not alter the result.*

*Proof.* We follow Berkes and Horváth (2003) and define

$$\begin{aligned}
\tilde{\gamma}_t(u) &= \tilde{\sigma}_t(\theta_0 + n^{-1/2}u) / \sigma_t(\theta_0) \\
\gamma_t(u) &= \sigma_t(\theta_0 + n^{-1/2}u) / \sigma_t(\theta_0) \\
\zeta_t(x, u) &= \mathbb{1}_{\{\eta_t \leq x\tilde{\gamma}_t(u)\}} - F(x\tilde{\gamma}_t(u)) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \\
S_n(x, u) &= \sum_{t=1}^n \zeta_t(x, u) \\
\mathbb{F}_n(x) &= \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\}}.
\end{aligned}$$

Let  $A > 0$  and write  $\mathcal{V}(\xi_\alpha)$  to denote the neighborhood around  $\xi_\alpha$  on which  $f$  is continuous; see Assumption 5(ii). Since  $\xi_\alpha < 0$ , we can take a compact neighborhood  $\mathcal{X} = [\underline{x}, \bar{x}] \subset \mathcal{V}(\xi_\alpha)$  such that  $\xi_\alpha \in \mathcal{X}$  and  $\bar{x} < 0$ . We establish the result in seven steps:

$$\text{Step 1: } \mathbb{E}[|S_n(x, u)|^4] = O(n) \text{ for all } x \in \mathcal{X} \text{ and for all } u \in \{u \in \mathbb{R}^r : \|u\| \leq A\};$$

$$\text{Step 2: } \sup_{x \in \mathcal{X}} |S_n(x, u)| = o_p(\sqrt{n}) \text{ for all } u \in \{u \in \mathbb{R}^r : \|u\| \leq A\};$$

$$\text{Step 3: } \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} |S_n(x, u)| = o_p(\sqrt{n});$$

$$\text{Step 4: } \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n (F(x\tilde{\gamma}_t(u)) - F(x)) - xf(x)\Omega'u \right| = o_p(1);$$

$$\text{Step 5: } \sup_{x \in \mathcal{X}} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \mathbb{F}_n(x)) - xf(x)\Omega'\sqrt{n}(\hat{\theta}_n - \theta_0) \right| = o_p(1);$$

$$\text{Step 6: } \sup_{x, y \in \mathcal{I}_n} \left| \sqrt{n}(\mathbb{F}_n(x) - \mathbb{F}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| = O(n^{-\varrho/2} \log n) \text{ a.s.};$$

$$\text{Step 7: } \sup_{x, y \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \xrightarrow{p} 0.$$

*Step 1* to *Step 5* are similar to the proofs of Berkes and Horváth (2003), whereas *Step 6* resembles Bahadur (1966, Lemma 1).

Throughout *Step 1* to *Step 4* we take  $\delta \in (0, 1/2)$  such that  $\mathcal{X}_\delta = [\underline{x}(1+2\delta), \bar{x}(1-2\delta)]$  satisfies  $\mathcal{X} \subset \mathcal{X}_\delta \subset \mathcal{V}(\xi_\alpha)$ . Because  $f$  is continuous on  $\mathcal{X}_\delta$  and  $\mathcal{X}_\delta$  is compact,  $f$  is uniformly continuous on  $\mathcal{X}_\delta$  and there exists a finite  $M > 0$  such that

$$\sup_{x \in \mathcal{X}_\delta} f(x) \leq M. \quad (\text{A.13})$$

Consider *Step 1*; let  $\mathcal{F}_t$  be the  $\sigma$ -algebra generated by  $\zeta_t, \zeta_{t-1}, \dots$  and note that  $\{S_t(x, u), \mathcal{F}_t\}$  is a martingale given  $x$  and  $u$ . Theorem 2.11 of Hall and Heyde (1980) yields

$$\mathbb{E}[|S_n(x, u)|^4] \leq C \left( \mathbb{E} \left[ \max_{1 \leq t \leq n} \zeta_t^4(x, u) \right] + \mathbb{E} \left[ \left( \sum_{t=1}^n \mathbb{E}_{t-1} [\zeta_t^2(x, u)] \right)^2 \right] \right),$$

for some absolute constant  $C > 0$  independent of  $x$  and  $u$ , where  $\mathbb{E}_{t-1} = \mathbb{E}[\cdot | \mathcal{F}_{t-1}]$  is the expectation given  $\mathcal{F}_{t-1}$ . As  $|\zeta_t(x, u)| \leq 2$  for all  $t$  such that  $\mathbb{E}[\max_{1 \leq t \leq n} \zeta_t^4(x, u)] \leq 16$ , it suffices to show that

$$\mathbb{E} \left[ \left( \sum_{t=1}^n \mathbb{E}_{t-1} [\zeta_t^2(x, u)] \right)^2 \right] = O(n). \quad (\text{A.14})$$

First, we focus on the inner part  $\mathbb{E}_{t-1} [\zeta_t^2(x, u)]$  and decompose  $\zeta_t(x, u)$  into

$$\zeta_t(x, u) = \zeta_{t,1}(x, u) + \zeta_{t,2}(x, u)$$

with

$$\begin{aligned} \zeta_{t,1}(x, u) &= \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} - F(x \tilde{\gamma}_t(u)) - \mathbb{1}_{\{\eta_t \leq x \gamma_t(u)\}} + F(x \gamma_t(u)) \\ \zeta_{t,2}(x, u) &= \mathbb{1}_{\{\eta_t \leq x \gamma_t(u)\}} - F(x \gamma_t(u)) - \mathbb{1}_{\{\eta_t \leq x\}} + F(x). \end{aligned}$$



The elementary inequality

$$\left(\sum_{i=1}^m x_i\right)^2 \leq m \sum_{i=1}^m x_i^2 \quad (\text{A.15})$$

for all  $x_1, \dots, x_m \in \mathbb{R}$  with  $m \in \mathbb{N}$  implies that

$$\mathbb{E}_{t-1}[\zeta_t^2(x, u)] \leq 2\left(\mathbb{E}_{t-1}[\zeta_{t,1}^2(x, u)] + \mathbb{E}_{t-1}[\zeta_{t,2}^2(x, u)]\right).$$

Moreover, the inequality  $\text{Var}[\mathbb{1}_{\{X \leq y\}} - \mathbb{1}_{\{X \leq z\}}] \leq |F_X(y) - F_X(z)|$  for  $y, z \in \mathbb{R}$  and  $X \sim F_X$  gives

$$\begin{aligned} \mathbb{E}_{t-1}[\zeta_{t,1}^2(x, u)] &= \text{Var}_{t-1}[\mathbb{1}_{\{\eta_t \leq x\tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x\gamma_t(u)\}}] \leq |F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u))| \\ \mathbb{E}_{t-1}[\zeta_{t,2}^2(x, u)] &= \text{Var}_{t-1}[\mathbb{1}_{\{\eta_t \leq x\gamma_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x\}}] \leq |F(x\gamma_t(u)) - F(x)|. \end{aligned}$$

Combining results, it follows that

$$\mathbb{E}_{t-1}[\zeta_t^2(x, u)] \leq 2\left(|F(x\gamma_t(u)) - F(x)| + |F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u))|\right). \quad (\text{A.16})$$

Employing (A.16), we obtain that the left-hand side in (A.14) is bounded by

$$\begin{aligned} &4\mathbb{E}\left[\left(\sum_{t=1}^n |F(x\gamma_t(u)) - F(x)| + \sum_{t=1}^n |F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u))|\right)^2\right] \\ &\leq 8\left(\underbrace{\mathbb{E}\left[\left(\sum_{t=1}^n |F(x\gamma_t(u)) - F(x)|\right)^2\right]}_I + \underbrace{\mathbb{E}\left[\left(\sum_{t=1}^n |F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u))|\right)^2\right]}_{II}\right), \end{aligned}$$

where the last inequality follows from applying (A.15) once more. It suffices to show

that both terms are  $O(n)$ . Consider  $I$ ; The Cauchy-Schwarz inequality yields

$$\begin{aligned} I &= \sum_{t=1}^n \sum_{\tau=1}^n \mathbb{E} \left[ \left| F(x\gamma_t(u)) - F(x) \right| \left| F(x\gamma_\tau(u)) - F(x) \right| \right] \\ &\leq \sum_{t=1}^n \sum_{\tau=1}^n \left( \mathbb{E} \left[ \left( F(x\gamma_t(u)) - F(x) \right)^2 \right] \right)^{\frac{1}{2}} \left( \mathbb{E} \left[ \left( F(x\gamma_\tau(u)) - F(x) \right)^2 \right] \right)^{\frac{1}{2}}. \end{aligned} \quad (\text{A.17})$$

Henceforth, we take  $n$  sufficiently large such that  $\{\theta : \|\theta - \theta_0\| \leq A/\sqrt{n}\} \subseteq \mathcal{V}(\theta_0)$ .

The mean value theorem implies

$$\begin{aligned} \sup_{\|u\| \leq A} |\gamma_t(u) - 1| &= \sup_{\|u\| \leq A} \left| \frac{\sigma_t(\theta_0 + u/\sqrt{n}) - \sigma_t(\theta_0)}{\sigma_t(\theta_0)} \right| \\ &= \sup_{\|u\| \leq A} \left| \frac{1}{\sigma_t(\theta_0)} \frac{\partial \sigma_t(\bar{\theta}_n)}{\partial \theta'} \frac{1}{\sqrt{n}} u \right| = \frac{1}{\sqrt{n}} \sup_{\|u\| \leq A} \left| \frac{\sigma_t(\bar{\theta}_n)}{\sigma_t(\theta_0)} D'_t(\bar{\theta}_n) u \right| \\ &\leq \frac{1}{\sqrt{n}} \sup_{\|\theta - \theta_0\| \leq An^{-1/2}} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \sup_{\|\theta - \theta_0\| \leq An^{-1/2}} \|D_t(\theta)\| \sup_{\|u\| \leq A} \|u\| \leq \frac{A}{\sqrt{n}} T_t U_t, \end{aligned} \quad (\text{A.18})$$

where  $T_t$  and  $U_t$  are defined in (A.1) and  $\bar{\theta}_n$  lies between  $\theta_0$  and  $\theta_0 + u/\sqrt{n}$ . Define the event

$$\mathcal{A}_{n,t} = \left\{ \frac{A}{\sqrt{n}} T_t U_t \leq \delta \right\}, \quad (\text{A.19})$$

where  $\delta$  is given in the text preceding (A.13). The inner term of (A.17) can be bounded by

$$\begin{aligned} \mathbb{E} \left[ \left( F(x\gamma_t(u)) - F(x) \right)^2 \right] &= \mathbb{E} \left[ \underbrace{\left( F(x\gamma_t(u)) - F(x) \right)^2}_{\leq 1} (\mathbb{1}_{\{\mathcal{A}_{n,t}^c\}} + \mathbb{1}_{\{\mathcal{A}_{n,t}\}}) \right] \\ &\leq \underbrace{\mathbb{P}[\mathcal{A}_{n,t}^c]}_{I_1} + \underbrace{\mathbb{E} \left[ \left( F(x\gamma_t(u)) - F(x) \right)^2 \mathbb{1}_{\{\mathcal{A}_{n,t}\}} \right]}_{I_2}, \end{aligned} \quad (\text{A.20})$$

where the superscript  $c$  denotes the event's complement. Using Markov's inequality,

the Cauchy-Schwarz inequality and Assumption 9,  $I_1$  can be bounded by

$$I_1 = \mathbb{P} \left[ \frac{A}{\sqrt{n}} T_t U_t > \delta \right] \leq \frac{A^2}{n\delta^2} \mathbb{E}[T_t^2 U_t^2] \leq \frac{A^2}{n\delta^2} \underbrace{\left( \mathbb{E}[T_t^4] \right)^{\frac{1}{2}}}_{<\infty} \underbrace{\left( \mathbb{E}[U_t^4] \right)^{\frac{1}{2}}}_{<\infty} \quad (\text{A.21})$$

and, thus,  $I_1 = O(n^{-1})$ . Regarding  $I_2$ , the mean value theorem implies

$$I_2 = \mathbb{E} \left[ x^2 f^2(x\bar{\gamma}_t) (\gamma_t(u) - 1)^2 \mathbb{1}_{\{\mathcal{A}_{n,t}\}} \right]$$

with  $\bar{\gamma}_t$  being between  $\gamma_t(u)$  and 1. Since  $|\bar{\gamma}_t - 1| \leq |\gamma_t(u) - 1| \leq \delta$  in the event of  $\mathcal{A}_{n,t}$ , we have  $x\bar{\gamma}_t \in \mathcal{X}_\delta$ . Employing (A.13), (A.18), the Cauchy-Schwarz inequality and Assumption 9, we establish

$$I_2 \leq \mathbb{E} \left[ \underline{x}^2 M^2 \frac{A^2}{n} T_t^2 U_t^2 \mathbb{1}_{\{\mathcal{A}_{n,t}\}} \right] \leq \frac{\underline{x}^2 M^2 A^2}{n} \underbrace{\left( \mathbb{E}[T_t^4] \right)^{\frac{1}{2}}}_{<\infty} \underbrace{\left( \mathbb{E}[U_t^4] \right)^{\frac{1}{2}}}_{<\infty} = O(n^{-1}). \quad (\text{A.22})$$

Combining (A.20) to (A.22) yields

$$\mathbb{E} \left[ \left( F(x\gamma_t(u)) - F(x) \right)^2 \right] \leq I_1 + I_2 = O(n^{-1})$$

and, together with (A.17), we get

$$I \leq \sum_{t=1}^n \sum_{r=1}^n O(n^{-1/2}) O(n^{-1/2}) = O(n).$$

Next, we consider  $II$ , which can be bounded analogously to (A.17) by

$$II \leq \sum_{t=1}^n \sum_{\tau=1}^n \left( \mathbb{E} \left[ \left( F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u)) \right)^2 \right] \right)^{\frac{1}{2}} \times \left( \mathbb{E} \left[ \left( F(x\tilde{\gamma}_\tau(u)) - F(x\gamma_\tau(u)) \right)^2 \right] \right)^{\frac{1}{2}}. \quad (\text{A.23})$$

Assumption 4(i) gives

$$\sup_{||u|| \leq A} |\tilde{\gamma}_t(u) - \gamma_t(u)| = \sup_{||u|| \leq A} \frac{|\tilde{\sigma}_t(\theta_0 + n^{-1/2}u) - \sigma_t(\theta_0 + n^{-1/2}u)|}{\sigma_t(\theta_0)} \leq \rho^t \frac{C_1}{\underline{\omega}}. \quad (\text{A.24})$$

We define the events

$$\mathcal{B}_t = \left\{ \rho^t \frac{C_1}{\underline{\omega}} \leq \delta \rho^{t/2} \right\} \quad \text{and} \quad \mathcal{C}_{n,t} = \mathcal{A}_{n,t} \cap \mathcal{B}_t. \quad (\text{A.25})$$

In analogy to (A.20), the inner part of (A.23) can be bounded by

$$\mathbb{E} \left[ \left( F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u)) \right)^2 \right] \leq \underbrace{\mathbb{P}[\mathcal{C}_{n,t}^c]}_{II_1} + \underbrace{\mathbb{E} \left[ \left( F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u)) \right)^2 \mathbb{1}_{\{\mathcal{C}_{n,t}\}} \right]}_{II_2}.$$

Employing (A.21) and Markov's inequality yields

$$\begin{aligned} II_1 &= \mathbb{P}[\mathcal{A}_{n,t}^c \cup \mathcal{B}_t^c] \leq \mathbb{P}[\mathcal{A}_{n,t}^c] + \mathbb{P}[\mathcal{B}_t^c] = \mathbb{P}[\mathcal{A}_{n,t}^c] + \mathbb{P} \left[ \rho^{t/2} \frac{C_1}{\underline{\omega}} > \delta \right] \\ &\leq \frac{A^2}{n\delta^2} \left( \mathbb{E}[T_t^4] \right)^{\frac{1}{2}} \left( \mathbb{E}[U_t^4] \right)^{\frac{1}{2}} + (\rho^{s/2})^t \frac{\mathbb{E}[C_1^s]}{\delta^s \underline{\omega}^s} = O(n^{-1}) + O((\rho^{s/2})^t). \end{aligned} \quad (\text{A.26})$$

Regarding  $II_2$ , the mean value theorem implies

$$II_2 = \mathbb{E} \left[ x^2 f^2(x\tilde{\gamma}_t) (\tilde{\gamma}_t(u) - \gamma_t(u))^2 \mathbb{1}_{\{\mathcal{C}_{n,t}\}} \right]$$

with  $\check{\gamma}_t$  between  $\tilde{\gamma}_t(u)$  and  $\gamma_t(u)$ . Since

$$|\check{\gamma}_t - 1| \leq |\check{\gamma}_t - \gamma_t(u)| + |\gamma_t(u) - 1| \leq |\tilde{\gamma}_t(u) - \gamma_t(u)| + |\gamma_t(u) - 1| \leq 2\delta$$

in the event of  $\mathcal{C}_{n,t} = \mathcal{A}_{n,t} \cap \mathcal{B}_t$ , we have  $x\check{\gamma}_t \in \mathcal{X}_\delta$ . Employing (A.13) and (A.24) we obtain

$$II_2 \leq \mathbb{E} \left[ \underbrace{\underline{x}^2 M^2 \left( \rho^t \frac{C_1}{\underline{\omega}} \right)^2 \mathbb{1}_{\{\mathcal{C}_{n,t}\}}}_{\leq \delta^2 \rho^t} \right] \leq \underline{x}^2 M^2 \delta^2 \rho^t = O(\rho^t). \quad (\text{A.27})$$

Equations (A.26) and (A.27) imply

$$\mathbb{E} \left[ \left( F(x\check{\gamma}_t(u)) - F(x\gamma_t(u)) \right)^2 \right] \leq C(n^{-1} + \rho^t + (\rho^{s/2})^t)$$

for some constant  $C > 0$ . Inserting this result into (A.23), we conclude

$$II \leq C \sum_{t=1}^n \sum_{\tau=1}^n \left( n^{-1} + \rho^t + (\rho^{s/2})^t \right)^{\frac{1}{2}} \left( n^{-1} + \rho^\tau + (\rho^{s/2})^\tau \right)^{\frac{1}{2}} = O(n),$$

which completes *Step 1*.

In *Step 2* we divide  $\mathcal{X}$  into intervals with the points  $\underline{x} = x_1 < x_2 < \dots < x_N < x_{N+1} = \bar{x}$  satisfying  $0.5 n^{-3/4} \leq x_{j+1} - x_j \leq n^{-3/4}$  for all  $j = 1, \dots, N$  and  $N \in \mathbb{N}$ . It follows that  $N = O(n^{3/4})$ . We obtain

$$\begin{aligned} \sup_{x \in \mathcal{X}} |S_n(x, u)| &= \max_{1 \leq j \leq N} \sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u)| \\ &\leq \max_{1 \leq j \leq N} \sup_{x_j \leq x \leq x_{j+1}} \left( |S_n(x_{j+1}, u)| + |S_n(x, u) - S_n(x_{j+1}, u)| \right) \\ &\leq \max_{1 \leq j \leq N} |S_n(x_{j+1}, u)| + \max_{1 \leq j \leq N} \sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u) - S_n(x_{j+1}, u)|. \end{aligned} \quad (\text{A.28})$$

We bound the second term using the elementary inequality

$$|x - y| \leq \max\{x, y\} \quad (\text{A.29})$$

for all  $x, y \geq 0$ . For  $j = 1 \dots, N$ , we have

$$\begin{aligned}
& \sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u) - S_n(x_{j+1}, u)| \\
&= \sup_{x_j \leq x \leq x_{j+1}} \left| \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - \mathbb{1}_{\{\eta_t \leq x\}} + F(x_{j+1} \tilde{\gamma}_t(u)) - F(x \tilde{\gamma}_t(u)) \right) \right. \\
&\quad \left. - \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_{j+1} \tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} + F(x_{j+1}) - F(x) \right) \right| \\
&\leq \sup_{x_j \leq x \leq x_{j+1}} \max \left\{ \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - \mathbb{1}_{\{\eta_t \leq x\}} + F(x_{j+1} \tilde{\gamma}_t(u)) - F(x \tilde{\gamma}_t(u)) \right), \right. \\
&\quad \left. \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_{j+1} \tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} + F(x_{j+1}) - F(x) \right) \right\} \quad (\text{A.30}) \\
&\leq \max \left\{ \underbrace{\sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - \mathbb{1}_{\{\eta_t \leq x_j\}} + F(x_{j+1} \tilde{\gamma}_t(u)) - F(x_j \tilde{\gamma}_t(u)) \right)}_{=A_n}, \right. \\
&\quad \left. \underbrace{\sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_{j+1} \tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x_j \tilde{\gamma}_t(u)\}} + F(x_{j+1}) - F(x_j) \right)}_{=B_n} \right\}.
\end{aligned}$$

Note that  $A_n$  and  $B_n$  are positive, where the later can be rewritten as

$$\begin{aligned}
B_n &= \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_{j+1} \tilde{\gamma}_t(u)\}} - F(x_{j+1} \tilde{\gamma}_t(u)) - \mathbb{1}_{\{\eta_t \leq x_{j+1}\}} + F(x_{j+1}) \right) \\
&\quad - \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_j \tilde{\gamma}_t(u)\}} - F(x_j \tilde{\gamma}_t(u)) - \mathbb{1}_{\{\eta_t \leq x_j\}} + F(x_j) \right) \\
&\quad + \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - \mathbb{1}_{\{\eta_t \leq x_j\}} + F(x_{j+1} \tilde{\gamma}_t(u)) - F(x_j \tilde{\gamma}_t(u)) \right) \\
&= S_n(x_{j+1}, u) - S_n(x_j, u) + A_n.
\end{aligned} \quad (\text{A.31})$$

It follows from (A.30) and (A.31) that

$$\sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u) - S_n(x_{j+1}, u)| \leq |S_n(x_{j+1}, u)| + |S_n(x_j, u)| + A_n. \quad (\text{A.32})$$

Moreover,  $A_n$  expands as follows:

$$\begin{aligned} A_n = & \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - F(x_{j+1}) - \mathbb{1}_{\{\eta_t \leq x_j\}} + F(x_j) \right) + n(F(x_{j+1}) - F(x_j)) \\ & + \sum_{t=1}^n \left( F(x_{j+1} \tilde{\gamma}_t(u)) - F(x_j \tilde{\gamma}_t(u)) \right) \end{aligned} \quad (\text{A.33})$$

Using equations (A.28), (A.32) and (A.33), we establish

$$\sup_{x \in \mathcal{X}} |S_n(x, u)| \leq 3III + IV + V + VI + 2VII, \quad (\text{A.34})$$

where

$$\begin{aligned} III &= \max_{1 \leq j \leq N+1} |S_n(x_j, u)| \\ IV &= \max_{1 \leq j \leq N} n(F(x_{j+1}) - F(x_j)) \\ V &= \max_{1 \leq j \leq N} \left| \sum_{t=1}^n (\mathbb{1}_{\{\eta_t \leq x_{j+1}\}} - F(x_{j+1})) - \sum_{t=1}^n (\mathbb{1}_{\{\eta_t \leq x_j\}} - F(x_j)) \right| \\ VI &= \max_{1 \leq j \leq N} \sum_{t=1}^n \left( F(x_{j+1} \gamma_t(u)) - F(x_j \gamma_t(u)) \right) \\ VII &= \max_{1 \leq j \leq N+1} \sum_{t=1}^n \left| F(x_j \tilde{\gamma}_t(u)) - F(x_j \gamma_t(u)) \right|. \end{aligned}$$

We look at each term in turn. For each  $\varepsilon > 0$ , Markov's inequality implies

$$\begin{aligned}\mathbb{P}[III \geq \sqrt{n}\varepsilon] &= \mathbb{P}\left[\max_{1 \leq j \leq N+1} |S_n(x_j, u)|^4 \geq n^2 \varepsilon^4\right] \leq \frac{1}{n^2 \varepsilon^4} \mathbb{E}\left[\max_{1 \leq j \leq N+1} |S_n(x_j, u)|^4\right] \\ &\leq \sum_{j=1}^{N+1} \frac{1}{n^2 \varepsilon^4} \mathbb{E}\left[|S_n(x_j, u)|^4\right] \rightarrow 0\end{aligned}$$

as  $N = O(n^{3/4})$  and  $\mathbb{E}[|S_n(x, u)|^4] = O(n)$  by *Step 1*. Thus, we have  $III = o_p(\sqrt{n})$ .

Regarding *IV*, the mean value theorem and (A.13) yield

$$F(x_{j+1}) - F(x_j) = f(\check{x}_j)(x_{j+1} - x_j) \leq Mn^{-3/4}, \quad (\text{A.35})$$

where  $\check{x}_j \in (x_j, x_{j+1})$ . It follows that

$$IV \leq nMn^{-3/4} = Mn^{1/4}$$

yielding  $IV = O(n^{1/4})$ . Further, Theorem 4.3.1 of Csörgő and Révész (1981) implies that there exists a sequence of Brownian bridges  $\{B_n(y) : 0 \leq y \leq 1\}$  such that

$$\begin{aligned}V/\sqrt{n} &= \max_{1 \leq j \leq N} \left| \sqrt{n}(\mathbb{F}_n(x_{j+1}) - F(x_{j+1})) - \sqrt{n}(\mathbb{F}_n(x_j) - F(x_j)) \right| \\ &\leq \max_{1 \leq j \leq N} \left| B_n(F(x_{j+1})) - B_n(F(x_j)) \right| + \max_{1 \leq j \leq N} \left| \sqrt{n}(\mathbb{F}_n(x_j) - F(x_j)) - B_n(F(x_j)) \right| \\ &\quad + \max_{1 \leq j \leq N} \left| \sqrt{n}(\mathbb{F}_n(x_{j+1}) - F(x_{j+1})) - B_n(F(x_{j+1})) \right| \\ &\leq \max_{1 \leq j \leq N} \left| B_n(F(x_{j+1})) - B_n(F(x_j)) \right| + 2 \sup_{x \in \mathbb{R}} \left| \sqrt{n}(\mathbb{F}_n(x) - F(x)) - B_n(F(x)) \right| \\ &\stackrel{a.s.}{=} \max_{1 \leq j \leq N} \underbrace{\left| B_n(F(x_{j+1})) - B_n(F(x_j)) \right|}_{Z_{n,j}} + o(1).\end{aligned}$$

Next, we show that  $\max_{1 \leq j \leq N} |Z_{n,j}| = o_p(1)$ . By the definition of a Brownian bridge



(cf. Csörgő and Révész, 1981, p. 41),  $Z_{n,j}$  is Gaussian with mean 0 and variance

$$\text{Var}[Z_{n,j}] = (F(x_{j+1}) - F(x_j)) \underbrace{\left(1 - (F(x_{j+1}) - F(x_j))\right)}_{\leq 1} \leq Mn^{-3/4}$$

by (A.35). In addition, we have  $\mathbb{E}[Z_{n,j}^4] = 3(\text{Var}[Z_{n,j}])^2 \leq 3M^2n^{-3/2}$ . Thus, for each  $\varepsilon > 0$ , Markov's inequality implies

$$\begin{aligned} \mathbb{P}\left[\max_{1 \leq j \leq N} |Z_{n,j}| \geq \varepsilon\right] &= \mathbb{P}\left[\max_{1 \leq j \leq N} Z_{n,j}^4 \geq \varepsilon^4\right] \leq \frac{1}{\varepsilon^4} \mathbb{E}\left[\max_{1 \leq j \leq N} Z_{n,j}^4\right] \\ &\leq \frac{1}{\varepsilon^4} \mathbb{E}\left[\sum_{j=1}^N Z_{n,j}^4\right] \leq \frac{1}{\varepsilon^4} \sum_{j=1}^N 3M^2n^{-3/2} = \frac{3M^2}{\varepsilon^4} n^{-3/2} N \rightarrow 0 \end{aligned}$$

as  $N = O(n^{3/4})$  and we conclude  $\max_{1 \leq j \leq N} |Z_{n,j}| = o_p(1)$ . Thus,  $V = o_p(\sqrt{n})$ . In analogy to (A.20), we bound  $VI$  by

$$VI \leq \underbrace{\sum_{t=1}^n \mathbb{1}_{\{\mathcal{A}_{n,t}^c\}}}_{VI_1} + \underbrace{\max_{1 \leq j \leq N} \sum_{t=1}^n \left(F(x_{j+1}\gamma_t(u)) - F(x_j\gamma_t(u))\right) \mathbb{1}_{\{\mathcal{A}_{n,t}\}}}_{VI_2}. \quad (\text{A.36})$$

Concerning the first subterm, for each  $\varepsilon > 0$ , Markov's inequality and (A.21) lead to

$$\begin{aligned} \mathbb{P}[VI_1 \geq \sqrt{n}\varepsilon] &\leq \frac{1}{\sqrt{n}\varepsilon} \mathbb{E}\left[\sum_{t=1}^n \mathbb{1}_{\{\mathcal{A}_{n,t}^c\}}\right] = \frac{1}{\sqrt{n}\varepsilon} \sum_{t=1}^n \mathbb{P}[\mathcal{A}_{n,t}^c] \\ &\leq \frac{A^2}{\sqrt{n}\varepsilon\delta^2} \left(\mathbb{E}[T_t^4]\right)^{\frac{1}{2}} \left(\mathbb{E}[U_t^4]\right)^{\frac{1}{2}} = O(n^{-1/2}). \end{aligned} \quad (\text{A.37})$$

Thus, we have  $VI_1 = o_p(\sqrt{n})$ . Regarding  $VI_2$ , the mean value theorem implies

$$VI_2 = \max_{1 \leq j \leq N} \sum_{t=1}^n \gamma_t(u) f(\tilde{x}_j \gamma_t(u)) (x_{j+1} - x_j) \mathbb{1}_{\{\mathcal{A}_{n,t}\}},$$

where  $\tilde{x}_j$  lies between  $x_j$  and  $x_{j+1}$ . Since  $|\gamma_t(u) - 1| \leq \delta$  in the event of  $\mathcal{A}_{n,t}$ , we have

$\tilde{x}_j \gamma_t(u) \in \mathcal{X}_\delta$ . Employing (A.13) and (A.18), we get

$$VI_2 \leq \sum_{t=1}^n \left( 1 + \frac{A}{\sqrt{n}} T_t U_t \right) M n^{-3/4} = M n^{1/4} + \frac{A}{n^{1/4}} \frac{1}{n} \sum_{t=1}^n T_t U_t$$

Whereas the first term is of order  $O(n^{1/4})$ , the second term vanishes almost surely as

$$\frac{1}{n} \sum_{t=1}^n T_t U_t \leq \underbrace{\left( \frac{1}{n} \sum_{t=1}^n T_t^2 \right)^{\frac{1}{2}}}_{\xrightarrow{a.s.} \mathbb{E}[T_t^2] < \infty} \underbrace{\left( \frac{1}{n} \sum_{t=1}^n U_t^2 \right)^{\frac{1}{2}}}_{\xrightarrow{a.s.} \mathbb{E}[U_t^2] < \infty} \quad (\text{A.38})$$

by Markov's inequality, the uniform ergodic theorem and Assumption 9. Hence,  $VI_2 = O(n^{1/4})$  almost surely. Next, we show

$$VII^\diamond = \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| F(x \tilde{\gamma}_t(u)) - F(x \gamma_t(u)) \right| = O_p(1), \quad (\text{A.39})$$

which implies  $VII = O_p(1)$ . Similar to (A.20), we bound  $VII^\diamond$  by

$$VII^\diamond \leq \underbrace{\sum_{t=1}^n \mathbb{1}_{\{\mathcal{C}_{n,t}^c\}}}_{VII_1^\diamond} + \underbrace{\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| F(x_j \tilde{\gamma}_t(u)) - F(x_j \gamma_t(u)) \right| \mathbb{1}_{\{\mathcal{C}_{n,t}\}}}_{VII_2^\diamond}$$

where the event  $\mathcal{C}_{n,t} = \mathcal{A}_{n,t} \cap \mathcal{B}_t$  is defined in (A.25). We show that both terms are  $O_p(1)$ . Employing Markov's inequality and (A.26), we have for each  $C > 0$

$$\begin{aligned} \mathbb{P}[VII_1^\diamond \geq C] &\leq \frac{1}{C} \mathbb{E}[VII_1^\diamond] = \frac{1}{C} \sum_{t=1}^n \mathbb{P}[\mathcal{C}_{n,t}^c] \leq \frac{1}{C} \sum_{t=1}^n \left( \mathbb{P}[\mathcal{A}_{n,t}^c] + \mathbb{P}[\mathcal{B}_t^c] \right) \\ &\leq \frac{1}{C} \sum_{t=1}^n \left( \frac{A^2}{n \delta^2} \left( \mathbb{E}[T_t^4] \right)^{\frac{1}{2}} \left( \mathbb{E}[U_t^4] \right)^{\frac{1}{2}} + (\rho^{s/2})^t \frac{\mathbb{E}[C_1^s]}{\delta^s \underline{\omega}^s} \right) \\ &\leq \frac{1}{C} \left( \frac{A^2}{\delta^2} \left( \mathbb{E}[T_t^4] \right)^{\frac{1}{2}} \left( \mathbb{E}[U_t^4] \right)^{\frac{1}{2}} + \frac{\mathbb{E}[C_1^s]}{\underline{\omega}^s \delta^s (1 - \rho^{s/2})} \right). \end{aligned} \quad (\text{A.40})$$

Choosing  $C$  sufficiently large,  $\mathbb{P}[VII_1^\diamond \geq C]$  can be made sufficiently small and we conclude  $V_1^\diamond = O_p(1)$ . Analogously to (A.27) we obtain

$$\begin{aligned} VII_2^\diamond &= \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| x f(x \check{\gamma}_t) (\check{\gamma}_t(u) - \gamma_t(u)) \right| \mathbb{1}_{\{\mathcal{C}_{n,t}\}} \\ &\leq \sum_{t=1}^n |x| M \underbrace{\frac{C_1 \rho^t}{\underline{\omega}} \mathbb{1}_{\{\mathcal{C}_{n,t}\}}}_{\leq \delta \rho^{t/2}} \leq \sum_{t=1}^n |x| M \delta \rho^{t/2} \leq \frac{2|x| M \delta}{(1 - \sqrt{\rho})^2} = O(1) \end{aligned} \quad (\text{A.41})$$

and we conclude  $VII^\diamond = O_p(1)$ . *Step 2* is completed.

In *Step 3* we divide the (hyper-)cube  $[-A, A]^r$  into  $L = (2N)^r$  cubes with side length  $A/N$  and  $N \in \mathbb{N}$ . In case of a cube  $\ell$ ,  $u_\bullet(\ell)$  and  $u^\bullet(\ell)$  denote the lower left and upper right vertex of  $\ell$ .<sup>12</sup> Similar to (A.28), we obtain

$$\begin{aligned} \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} |S_n(x, u)| &\leq \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} |S_n(x, u^\bullet(\ell))| \\ &\quad + \max_{1 \leq \ell \leq L} \sup_{u_\bullet(\ell) \leq u \leq u^\bullet(\ell)} \sup_{x \in \mathcal{X}} |S_n(x, u) - S_n(x, u^\bullet(\ell))|. \end{aligned} \quad (\text{A.42})$$

We focus on the second term. Fix  $\ell \in \{1 \dots, L\}$  and consider  $u$  satisfying  $u_\bullet(\ell) \leq u \leq u^\bullet(\ell)$  (element-by-element comparison). Assumption 8 implies  $\check{\gamma}_t(u_\bullet(\ell)) \leq \check{\gamma}_t(u) \leq$

---

<sup>12</sup>Lower left (right) vertex means that all coordinates of  $u_\bullet(\ell)$  ( $u^\bullet(\ell)$ ) are less (larger) than or equal to the corresponding coordinates of any elements of  $\ell$ .

$\tilde{\gamma}_t(u^\bullet(\ell))$ . Since  $x < 0$  for all  $x \in \mathcal{X}$ , the elementary inequality (A.29) implies

$$\begin{aligned}
& |S_n(x, u) - S_n(x, u^\bullet(\ell))| \\
&= \left| \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} - F(x \tilde{\gamma}_t(u)) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \right) \right. \\
&\quad \left. - \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^\bullet(\ell))\}} - F(x \tilde{\gamma}_t(u^\bullet(\ell))) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \right) \right| \\
&= \left| \underbrace{\sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^\bullet(\ell))\}} \right)}_{\geq 0} - \underbrace{\sum_{t=1}^n \left( F(x \tilde{\gamma}_t(u)) - F(x \tilde{\gamma}_t(u^\bullet(\ell))) \right)}_{\geq 0} \right| \quad (\text{A.43}) \\
&\leq \max \left\{ \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^\bullet(\ell))\}} \right), \sum_{t=1}^n \left( F(x \tilde{\gamma}_t(u)) - F(x \tilde{\gamma}_t(u^\bullet(\ell))) \right) \right\} \\
&\leq \max \left\{ \underbrace{\sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u_\bullet(\ell))\}} - \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^\bullet(\ell))\}} \right)}_{=C_n}, \underbrace{\sum_{t=1}^n \left( F(x \tilde{\gamma}_t(u_\bullet(\ell))) - F(x \tilde{\gamma}_t(u^\bullet(\ell))) \right)}_{=D_n} \right\}.
\end{aligned}$$

Note that  $C_n$  can be written as

$$\begin{aligned}
C_n &= \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u_\bullet(\ell))\}} - F(x \tilde{\gamma}_t(u_\bullet(\ell))) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \right) \\
&\quad - \sum_{t=1}^n \left( \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u^\bullet(\ell))\}} - F(x \tilde{\gamma}_t(u^\bullet(\ell))) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x)) \right) \\
&\quad + \sum_{t=1}^n \left( F(x \tilde{\gamma}_t(u_\bullet(\ell))) - F(x \tilde{\gamma}_t(u^\bullet(\ell))) \right) \quad (\text{A.44}) \\
&= S_n(x, u_\bullet(\ell)) - S_n(x, u^\bullet(\ell)) + D_n.
\end{aligned}$$

Combining (A.43) and (A.44), we find

$$|S_n(x, u) - S_n(x, u^\bullet(\ell))| \leq |S_n(x, u_\bullet(\ell))| + |S_n(x, u^\bullet(\ell))| + |D_n|. \quad (\text{A.45})$$

Moreover,  $D_n$  expands as follows:

$$\begin{aligned}
D_n = & \sum_{t=1}^n \left( F(x\gamma_t(u_{\bullet}(\ell))) - F(x\gamma_t(u^{\bullet}(\ell))) \right) \\
& + \sum_{t=1}^n \left( F(x\tilde{\gamma}_t(u_{\bullet}(\ell))) - F(x\gamma_t(u_{\bullet}(\ell))) \right) \\
& - \sum_{t=1}^n \left( F(x\tilde{\gamma}_t(u^{\bullet}(\ell))) - F(x\gamma_t(u^{\bullet}(\ell))) \right)
\end{aligned} \tag{A.46}$$

Equations (A.42) and (A.46) lead to

$$\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} |S_n(x, u)| \leq 2VIII + IX + X + XI + XII \tag{A.47}$$

with

$$\begin{aligned}
VIII &= \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} |S_n(x, u^{\bullet}(\ell))| \\
IX &= \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} |S_n(x, u_{\bullet}(\ell))| \\
X &= \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| F(x\tilde{\gamma}_t(u_{\bullet}(\ell))) - F(x\gamma_t(u_{\bullet}(\ell))) \right| \\
XI &= \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left| F(x\tilde{\gamma}_t(u^{\bullet}(\ell))) - F(x\gamma_t(u^{\bullet}(\ell))) \right| \\
XII &= \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left( F(x\gamma_t(u_{\bullet}(\ell))) - F(x\gamma_t(u^{\bullet}(\ell))) \right).
\end{aligned}$$

$VIII$  and  $IX$  are  $o_p(\sqrt{n})$  for fixed  $L$  by *Step 2* whereas  $X = O_p(1)$  and  $XI = O_p(1)$  by (A.39). In analogy to (A.20), we bound  $XII$  by

$$XII \leq \underbrace{\sum_{t=1}^n \mathbb{1}_{\{\mathcal{A}_{n,t}^c\}}}_{XII_1} + \underbrace{\max_{1 \leq j \leq N} \sup_{x \in \mathcal{X}} \sum_{t=1}^n \left( F(x\gamma_t(u_{\bullet}(\ell))) - F(x\gamma_t(u^{\bullet}(\ell))) \right) \mathbb{1}_{\{\mathcal{A}_{n,t}\}}}_{XII_2}. \tag{A.48}$$

We have  $XII_1 = o_p(\sqrt{n})$  by (A.37). Regarding  $XII_2$ , the mean value theorem implies

$$XII_2 = \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} \sum_{t=1}^n x f(x \bar{\gamma}_t) (\gamma_t(u_{\bullet}(\ell)) - \gamma_t(u^{\bullet}(\ell))) \mathbb{1}_{\{\mathcal{A}_{n,t}\}}$$

with  $\bar{\gamma}_t$  lying between  $\gamma_t(u_{\bullet}(\ell))$  and  $\gamma_t(u^{\bullet}(\ell))$ . Since  $|\bar{\gamma}_t - 1| \leq 2\delta$  in the event of  $\mathcal{A}_{n,t}$ , we have  $x \bar{\gamma}_t \in \mathcal{X}_{\delta}$  for all  $x \in \mathcal{X}$ . Taking  $n$  sufficiently large such that  $\{\theta : \|\theta - \theta_0\| \leq A/\sqrt{n}\} \subseteq \mathcal{V}(\theta_0)$ , (A.13) and the mean value theorem imply

$$\begin{aligned} XII_2 &\leq |\underline{x}| M \max_{1 \leq \ell \leq L} \sup_{x \in \mathcal{X}} \sum_{t=1}^n (\gamma_t(u^{\bullet}(\ell)) - \gamma_t(u_{\bullet}(\ell))) \\ &= |\underline{x}| M \max_{1 \leq \ell \leq L} \sum_{t=1}^n \frac{\sigma_t(\theta_0 + n^{-1/2} u^{\bullet}(\ell)) - \sigma_t(\theta_0 + n^{-1/2} u_{\bullet}(\ell))}{\sigma_t(\theta_0)} \\ &= |\underline{x}| M \max_{1 \leq \ell \leq L} \sum_{t=1}^n \frac{1}{\sigma_t(\theta_0)} \frac{\partial \sigma_t(\bar{\theta}_n)}{\partial \theta'} \frac{1}{\sqrt{n}} (u^{\bullet}(\ell) - u_{\bullet}(\ell)) \\ &\leq \frac{|\underline{x}| M}{\sqrt{n}} \max_{1 \leq \ell \leq L} \sum_{t=1}^n \frac{\sigma_t(\bar{\theta}_n)}{\sigma_t(\theta_0)} \left\| \frac{1}{\sigma_t(\bar{\theta}_n)} \frac{\partial \sigma_t(\bar{\theta}_n)}{\partial \theta} \right\| \|u^{\bullet}(\ell) - u_{\bullet}(\ell)\| \\ &\leq \frac{rA|\underline{x}|M}{\sqrt{n}N} \sum_{t=1}^n \sup_{\|\theta - \theta_0\| \leq A/\sqrt{n}} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \sup_{\|\theta - \theta_0\| \leq A/\sqrt{n}} \|D_t(\theta)\| \\ &\leq \frac{rA|\underline{x}|M}{\sqrt{n}N} \sum_{t=1}^n T_t U_t, \end{aligned}$$

where  $\theta_0 + n^{-1/2} u_{\bullet}(\ell) \leq \bar{\theta}_n \leq \theta_0 + n^{-1/2} u^{\bullet}(\ell)$  (componentwise). Employing (A.38), we obtain  $XII_2 = O(\sqrt{n})/N$  almost surely, where the  $O(\sqrt{n})$  term does not depend on  $N$ . Choosing  $N$  large, we obtain  $XII_2 = o(\sqrt{n})$  almost surely and we conclude that  $XII = o_p(\sqrt{n})$ . *Step 3* is completed.

Regarding *Step 4* we establish the following bound:

$$\begin{aligned}
& \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( F(\tilde{\gamma}_t(u)x) - F(x) \right) - xf(x)\Omega'u \right| \\
& \leq \underbrace{\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( F(x\tilde{\gamma}_t(u)) - F(x\gamma_t(u)) \right) \right|}_{=XIII} \\
& \quad + \underbrace{\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| xf(x) \frac{1}{n} \sum_{t=1}^n D'_t u - xf(x)\Omega'u \right|}_{=XIV} \\
& \quad + \underbrace{\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( F(x\gamma_t(u)) - F(x) \right) - xf(x) \frac{1}{n} \sum_{t=1}^n D'_t u \right|}_{=XV},
\end{aligned} \tag{A.49}$$

where  $XIII = O_p(n^{-1/2})$  by (A.39). Further, (A.13) and the ergodic theorem imply

$$XIV \leq \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} |x|f(x) \left\| \frac{1}{n} \sum_{t=1}^n D_t - \Omega \right\| \|u\| \leq A|\underline{x}|M \left\| \frac{1}{n} \sum_{t=1}^n D_t - \Omega \right\| \xrightarrow{a.s.} 0.$$

Regarding the last term, we use the mean value theorem and (A.13) to obtain

$$\begin{aligned}
XV &= \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( xf(x\bar{\gamma}_t)(\gamma_t(u) - 1) - xf(x) \frac{1}{\sqrt{n}} D'_t u \right) \right| \\
&\leq \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( xf(x)(\gamma_t(u) - 1) - xf(x) \frac{1}{\sqrt{n}} D'_t u \right) \right| \\
&\quad + \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( xf(x\bar{\gamma}_t)(\gamma_t(u) - 1) - xf(x)(\gamma_t(u) - 1) \right) \right| \\
&\leq \underbrace{\frac{|\underline{x}|M}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} \left| (\gamma_t(u) - 1) - \frac{1}{\sqrt{n}} D'_t u \right|}_{XV_1} \\
&\quad + \underbrace{\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n x(f(x\bar{\gamma}_t) - f(x))(\gamma_t(u) - 1) \right|}_{XV_2}
\end{aligned}$$

with  $\bar{\gamma}_t$  being between  $\gamma_t(u)$  and 1. For  $n$  sufficiently large such that  $\{\theta : \|\theta - \theta_0\| \leq A/\sqrt{n}\} \subseteq \mathcal{V}(\theta_0)$ , a second-order Taylor expansion gives

$$\begin{aligned} XV_1 &= \frac{|\underline{x}|M}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} \frac{1}{\sigma_t(\theta_0)} \left| \sigma_t(\theta_0 + n^{-1/2}u) - \sigma_t(\theta_0) - \frac{1}{\sqrt{n}} \frac{\partial \sigma_t(\theta_0)}{\partial \theta'} u \right| \\ &= \frac{|\underline{x}|M}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} \frac{1}{\sigma_t(\theta_0)} \left| \frac{1}{2n} u' \frac{\partial^2 \sigma_t(\bar{\theta}_n)}{\partial \theta \partial \theta'} u \right| \leq \frac{A^2 |\underline{x}|M}{2n^{3/2}} \sum_{t=1}^n \frac{\sigma_t(\bar{\theta}_n)}{\sigma_t(\theta_0)} \left\| \frac{1}{\sigma_t(\bar{\theta}_n)} \frac{\partial^2 \sigma_t(\bar{\theta}_n)}{\partial \theta \partial \theta'} \right\| \\ &\leq \frac{A^2 |\underline{x}|M}{2n^{3/2}} \sum_{t=1}^n \sup_{\|\theta - \theta_0\| \leq A/\sqrt{n}} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \sup_{\|\theta - \theta_0\| \leq A/\sqrt{n}} \|H_t(\theta)\| \leq \frac{A^2 |\underline{x}|M}{2n^{3/2}} \sum_{t=1}^n T_t V_t \end{aligned}$$

with  $\bar{\theta}_n$  being between  $\theta_0$  and  $\theta_0 + n^{-1/2}u$ . The Cauchy-Schwarz inequality, the uniform ergodic theorem and Assumption 9 yield

$$\frac{1}{n} \sum_{t=1}^n T_t V_t \leq \underbrace{\left( \frac{1}{n} \sum_{t=1}^n T_t^2 \right)^{\frac{1}{2}}}_{\xrightarrow{a.s.} \mathbb{E}[T_t^2] < \infty} \underbrace{\left( \frac{1}{n} \sum_{t=1}^n V_t^2 \right)^{\frac{1}{2}}}_{\xrightarrow{a.s.} \mathbb{E}[V_t^2] < \infty}$$

and we conclude that  $XV_1 = O(n^{-1/2})$  almost surely. Before turning to  $XV_2$ , we establish two auxiliary results:

- (i)  $\frac{1}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} |\gamma_t(u) - 1| = O(1)$  almost surely;
- (ii)  $\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \max_{1 \leq t \leq n} |f(x\bar{\gamma}_t) - f(x)| = o_p(1)$ .

Statement (i) follows from (A.18) and (A.38) as

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} |\gamma_t(u) - 1| \leq \frac{A}{n} \sum_{t=1}^n T_t U_t \leq A \underbrace{\left( \frac{1}{n} \sum_{t=1}^n T_t^2 \right)^{\frac{1}{2}}}_{\xrightarrow{a.s.} \mathbb{E}[T_t^2] < \infty} \underbrace{\left( \frac{1}{n} \sum_{t=1}^n U_t^2 \right)^{\frac{1}{2}}}_{\xrightarrow{a.s.} \mathbb{E}[U_t^2] < \infty}.$$

To show (ii), we note that the Cauchy-Schwarz inequality and Assumption 9 yield  $\mathbb{E}[(T_t U_t)^3] \leq \mathbb{E}[T_t^6]^{\frac{1}{2}} \mathbb{E}[U_t^6]^{\frac{1}{2}} < \infty$ . For every  $\varepsilon > 0$  and for  $n$  sufficiently large such



that  $\{\theta : \|\theta - \theta_0\| \leq A/\sqrt{n}\} \subseteq \mathcal{V}(\theta_0)$ , we have

$$\begin{aligned} & \mathbb{P} \left[ \sup_{\|u\| \leq A} \max_{1 \leq t \leq n} |\gamma_t(u) - 1| \geq \varepsilon \right] \leq \mathbb{P} \left[ A \max_{1 \leq t \leq n} T_t U_t \geq \varepsilon \sqrt{n} \right] \\ & \leq \mathbb{P} \left[ A^3 \max_{1 \leq t \leq n} (T_t U_t)^3 \geq \varepsilon^3 n^{3/2} \right] \leq \frac{A^3}{n^{3/2} \varepsilon^3} \mathbb{E} \left[ \max_{1 \leq t \leq n} (T_t U_t)^3 \right] \leq \frac{A^3}{\sqrt{n} \varepsilon^3} \mathbb{E} [(T_t U_t)^3], \end{aligned}$$

which converges to 0, and thus we obtain  $\sup_{\|u\| \leq A} \max_{1 \leq t \leq n} |\gamma_t(u) - 1| = o_p(1)$ . Because  $\bar{\gamma}_t$  lies between  $\gamma_t(u)$  and 1, it follows that  $\sup_{\|u\| \leq A} \max_{1 \leq t \leq n} |\bar{\gamma}_t - 1| = o_p(1)$ . Thus, for sufficiently large  $n$ , we have  $x\bar{\gamma}_t \in \mathcal{X}_\delta$  with probability close to one. Then, statement (ii) follows from the fact that  $f$  is uniformly continuous on  $\mathcal{X}_\delta$ . Employing both auxiliary results, we obtain

$$\begin{aligned} XV_2 & \leq \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \frac{1}{\sqrt{n}} \sum_{t=1}^n |x| |f(x\bar{\gamma}_t) - f(x)| |\gamma_t(u) - 1| \\ & \leq |\underline{x}| \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \max_{1 \leq t \leq n} |f(x\bar{\gamma}_t) - f(x)| \frac{1}{\sqrt{n}} \sum_{t=1}^n \sup_{\|u\| \leq A} |\gamma_t(u) - 1| = o_p(1). \end{aligned}$$

Thus  $XV$  is  $o_p(1)$ , which completes *Step 4*.

Concerning *Step 5* we obtain for each  $\varepsilon > 0$

$$\begin{aligned} & \mathbb{P} \left[ \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t \leq x\}} - \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\}} - x f(x) \Omega' \sqrt{n} (\hat{\theta}_n - \theta_0) \right| \geq \varepsilon \right] \\ & \leq \mathbb{P} \left[ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq \tilde{\gamma}_t(u)x\}} - \frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbb{1}_{\{\eta_t \leq x\}} - x f(x) \Omega' u \right| \geq \varepsilon \right] \\ & \quad + \mathbb{P} \left[ \sqrt{n} \|\hat{\theta}_n - \theta_0\| > A \right] \\ & \leq \mathbb{P} \left[ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( F(\tilde{\gamma}_t(u)x) - F(x) \right) - x f(x) \Omega' u \right| \geq \frac{\varepsilon}{2} \right] \\ & \quad + \mathbb{P} \left[ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} |S_n(x, u)/\sqrt{n}| \geq \frac{\varepsilon}{2} \right] + \mathbb{P} \left[ \sqrt{n} \|\hat{\theta}_n - \theta_0\| > A \right]. \end{aligned}$$

Since  $\sqrt{n}||\hat{\theta}_n - \theta_0|| = O_p(1)$  by Theorem 2, the third term can be made arbitrarily small for large  $n$  by choosing  $A$  sufficiently large. Given  $A$ , the first two terms converge to zero by *Step 3* and *Step 4*, which completes *Step 5*.

Regarding *Step 6* we refer to Bahadur (1966, Lemma 1). Replacing  $\xi$  by  $\xi_\alpha$  in the proof and choosing the sequences  $a_n$  and  $b_n$  to satisfy  $a_n \sim n^{-\varrho} \log n$  and  $b_n \sim n^\psi$  as  $n \rightarrow \infty$ , where  $\psi = (1 - \varrho)/2$ , it follows that

$$\mathbb{H}_{n,\alpha} = \sup_{x \in \mathcal{I}_n} \left| (\mathbb{F}_n(x) - \mathbb{F}_n(\xi_\alpha)) - (F(x) - F(\xi_\alpha)) \right| = O(n^{-(\varrho+\psi)} \log n)$$

almost surely as  $n \rightarrow \infty$ . Inserting the definition of  $\psi$  and inflating the term by  $\sqrt{n}$  leads to  $\sqrt{n} \mathbb{H}_{n,\alpha} = O(n^{-\varrho/2} \log n)$  almost surely as  $n \rightarrow \infty$ . Together with the triangle inequality, we establish

$$\sup_{x,y \in \mathcal{I}_n} \left| \sqrt{n}(\mathbb{F}_n(x) - \mathbb{F}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \leq 2\sqrt{n} \mathbb{H}_{n,\alpha} = O(n^{-\varrho/2} \log n),$$

which completes *Step 6*.

Regarding *Step 7* we bound

$$\begin{aligned} & \sup_{x,y \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \\ & \leq 2 \sup_{x \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \mathbb{F}_n(x)) - xf(x)\Omega' \sqrt{n}(\hat{\theta}_n - \theta_0) \right| \\ & \quad + \sup_{x,y \in \mathcal{I}_n} \left| \sqrt{n}(\mathbb{F}_n(x) - \mathbb{F}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \\ & \quad + \sup_{x,y \in \mathcal{I}_n} \left| (xf(x) - yf(y))\Omega' \sqrt{n}(\hat{\theta}_n - \theta_0) \right|. \end{aligned}$$

Taking  $n$  sufficiently large such that  $\mathcal{I}_n \subset \mathcal{X}$ , the first term on the right-hand side vanishes in probability by *Step 5*. The second term vanishes almost surely by *Step 6*.

The last term can be bounded as follows:

$$\sup_{x,y \in \mathcal{I}_n} \left| (xf(x) - yf(y))\Omega' \sqrt{n}(\hat{\theta}_n - \theta_0) \right| \leq \sup_{x,y \in \mathcal{I}_n} |xf(x) - yf(y)| \|\Omega\| \sqrt{n} \|\hat{\theta}_n - \theta_0\|.$$

Since  $f(x)$ , and hence  $xf(x)$ , is continuous in a neighborhood of  $\xi_\alpha$  by Assumption 5(ii) and  $\mathcal{I}_n$  shrinks to  $\xi_\alpha$  we have  $\sup_{x,y \in \mathcal{I}_n} |xf(x) - yf(y)| \rightarrow 0$ . Together with  $\sqrt{n} \|\hat{\theta}_n - \theta_0\| = O_p(1)$  (Theorem 2) we find that the last term converges in probability to 0, which completes *Step 7*.

To verify that replacing any  $\hat{\mathbb{F}}_n(\cdot)$  by  $\hat{\mathbb{F}}_n(\cdot -)$  does not alter the result, we note that  $\hat{\mathbb{F}}_n(x - n^{-1}) \leq \hat{\mathbb{F}}_n(x-) \leq \hat{\mathbb{F}}_n(x) \leq \hat{\mathbb{F}}_n(x + n^{-1})$  for all  $x \in \mathcal{I}_n$  (similarly for  $y$ ). Setting  $\bar{\mathcal{I}}_n = (\xi_\alpha - \bar{a}_n, \xi_\alpha + \bar{a}_n)$  with  $\bar{a}_n = a_n + n^{-1}$ , we can bound  $\sup_{x,y \in \mathcal{I}_n} |\sqrt{n}(\hat{\mathbb{F}}_n(x-) - \hat{\mathbb{F}}_n(y)) - \sqrt{n}(F(x) - F(y))|$  and  $\sup_{x,y \in \mathcal{I}_n} |\sqrt{n}(\hat{\mathbb{F}}_n(x-) - \hat{\mathbb{F}}_n(y-)) - \sqrt{n}(F(x) - F(y))|$  by

$$\begin{aligned} & \sup_{x,y \in \bar{\mathcal{I}}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y)) - \sqrt{n}(F(x) - F(y)) \right| \\ & + 2 \sup_{y \in \mathcal{I}_n} \sqrt{n} \left( F(y + n^{-1}) - F(y - n^{-1}) \right). \end{aligned} \tag{A.50}$$

The first term in (A.50) vanishes in probability by *Step 7* as  $\bar{a}_n \sim a_n$ . Regarding the second term, the mean value theorem implies

$$2 \sup_{y \in \mathcal{I}_n} \sqrt{n} \left( F\left(y + \frac{1}{n}\right) - F\left(y - \frac{1}{n}\right) \right) = \frac{4}{\sqrt{n}} \sup_{y \in \mathcal{I}_n} f(y + \varepsilon_n),$$

where  $|\varepsilon_n| \leq n^{-1}$ . Since  $\frac{4}{\sqrt{n}} \rightarrow 0$  and  $\sup_{y \in \mathcal{I}_n} f(y + \varepsilon_n) \rightarrow f(\xi_\alpha)$  the term vanishes, which completes the proof.  $\square$

*Remark 5.* *Step 5* is closely related to Lemma 3.2 of Gao and Song (2008) with  $\Omega$  corresponding to their  $\mathbf{e}/2$ . Whereas in *Step 5* we establish the uniformity over a

compact neighborhood of  $\xi_\alpha$ , they claim –without formal proof– uniform convergence in probability over  $\mathbb{R}$  assuming differentiability of  $f$  and  $\sup_{x \in \mathbb{R}} x^2 |f'(x)| < \infty$ .

## A.2 Bootstrap Lemmas

Henceforth we use  $\mathbb{P}^*$ ,  $\mathbb{E}^*$ ,  $\text{Var}^*$  and  $\text{Cov}^*$  to denote the probability, expectation, variance and covariance conditional on  $\mathcal{F}_n$ .

**Lemma 4.** *Suppose Assumptions 1–3, 4(i), 5(i) and 5(iii) hold.*

(i) *If in addition Assumption 9(i) holds with  $a = 4$ , then  $\mathbb{E}^*[\eta_t^{*m}] \xrightarrow{a.s.} \mathbb{E}[\eta_t^m]$  for  $m \in \{1, 2, 3, 4\}$ .*

(ii) *If in addition Assumptions 6, 7 and 9(i) hold with  $a = -1, 4$ , then we have  $\mathbb{E}^*[\eta_t^{*m} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \xrightarrow{a.s.} \mathbb{E}[\eta_t^m \mathbb{1}_{\{\eta_t < \xi_\alpha\}}]$  for  $m \in \{0, 1, 2, 3, 4\}$ .*

*Proof.* Lemma 2 gives  $\mathbb{E}^*[\eta_t^{*m} \mathbb{1}_{\{\eta_t^* < u\}}] = \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^m \mathbb{1}_{\{\hat{\eta}_t < u\}} \xrightarrow{a.s.} \mathbb{E}[\eta_t^m \mathbb{1}_{\{\eta_t < u\}}]$ . Taking  $u = \infty$  proves the first claim, whereas the second claim follows from  $\mathbb{E}[\eta_t^m \mathbb{1}_{\{\eta_t < u\}}]$  being continuous in  $u$  and  $\hat{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha$  by Theorem 1.  $\square$

**Lemma 5.** *Suppose Assumptions 1–3, 4(i), 5(i), 5(iii), 6 and 9(i)–(ii) hold with  $a = \pm 4$ . Then, we have  $\hat{\theta}_n^* \xrightarrow{p^*} \theta_0$  almost surely.*

*Proof.* The proof is inspired by Francq and Zakoïan (2004, Theorem 2.1). Let  $\nu > 0$  and set  $\mathcal{B} = \{\theta \in \Theta : \|\theta - \theta_0\| \geq \nu\}$ ; We establish the result in three steps:

*Step 1:* we obtain  $L_n^*(\theta) - L_n^*(\hat{\theta}_n) = \frac{1}{2n} \sum_{t=1}^n \left( 1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \eta_t^{*2} + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right) + R_n^*(\theta)$  with  $\sup_{\theta \in \Theta} |R_n^*(\theta)| \xrightarrow{p^*} 0$  almost surely;

*Step 2:* There exists a  $\zeta < 0$  such that  $\sup_{\theta \in \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) < \zeta/2 + S_n^*$  with  $S_n^* \xrightarrow{p^*} 0$  almost surely;

*Step 3:* we show  $\mathbb{P}^*[\hat{\theta}_n^* \in \mathcal{B}] \xrightarrow{a.s.} 0$ .

Regarding *Step 1* we find

$$L_n^*(\theta) - L_n^*(\hat{\theta}_n) = \frac{1}{2n} \sum_{t=1}^n \left\{ \eta_t^{*2} - \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \eta_t^{*2} + \log \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \right\},$$

where  $\frac{1}{n} \sum_{t=1}^n \eta_t^{*2} \xrightarrow{p^*} 1$  almost surely since

$$\mathbb{E}^* \left[ \frac{1}{n} \sum_{t=1}^n \eta_t^{*2} \right] = \mathbb{E}^*[\eta_t^{*2}] \xrightarrow{a.s.} 1 \quad \text{and} \quad \mathbb{V}\text{ar}^* \left[ \frac{1}{n} \sum_{t=1}^n \eta_t^{*2} \right] = \frac{1}{n} \mathbb{V}\text{ar}^*[\eta_t^{*2}] \xrightarrow{a.s.} 0$$

by Lemma 4. It remains to show the negligibility of the initial conditions, i.e.

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left\{ \log \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} - \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right\} \right| \xrightarrow{a.s.} 0 \quad (\text{A.51})$$

and

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left( \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} - \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \right) \eta_t^{*2} \right| \xrightarrow{p^*} 0 \quad (\text{A.52})$$

almost surely. The inequality  $\log(1+x) \leq x$  for all  $x > -1$  and Assumption 4(i) yield

$$\begin{aligned} & \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left( \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} - \log \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \right) \right| = \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left( \log \frac{\tilde{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)} - \log \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\sigma_t^2(\hat{\theta}_n)} \right) \right| \\ & \leq \sup_{\theta \in \Theta} \frac{2}{n} \sum_{t=1}^n \left| \log \frac{\tilde{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)} \right| = \sup_{\theta \in \Theta} \frac{4}{n} \sum_{t=1}^n \left| \log \frac{\tilde{\sigma}_t(\theta)}{\sigma_t(\theta)} \right| = \sup_{\theta \in \Theta} \frac{4}{n} \sum_{t=1}^n \left| \log \left( 1 + \frac{\tilde{\sigma}_t(\theta) - \sigma_t(\theta)}{\sigma_t(\theta)} \right) \right| \\ & \leq \frac{4}{n} \sum_{t=1}^n \log \left( 1 + \frac{C_1 \rho^t}{\underline{\omega}} \right) \leq \frac{4}{n} \sum_{t=1}^n \frac{C_1 \rho^t}{\underline{\omega}} \leq \frac{4C_1}{\underline{\omega}(1-\rho)n} \xrightarrow{a.s.} 0 \end{aligned}$$

verifying (A.51). Further, Assumption 4(i) and (A.5) imply

$$\begin{aligned}
& \sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^n \left( \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right) \eta_t^{*2} \right| \leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \left| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right| \eta_t^{*2} \\
&= \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \left| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n) - \sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\hat{\theta}_n)} + \frac{\sigma_t^2(\theta) - \tilde{\sigma}_t^2(\theta)}{\sigma_t^2(\theta)} \right| \eta_t^{*2} \\
&\leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \left( \frac{|\tilde{\sigma}_t^2(\hat{\theta}_n) - \sigma_t^2(\hat{\theta}_n)|}{\sigma_t^2(\hat{\theta}_n)} + \frac{|\sigma_t^2(\theta) - \tilde{\sigma}_t^2(\theta)|}{\sigma_t^2(\theta)} \right) \eta_t^{*2} \\
&\leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\theta)} \left( \frac{|\tilde{\sigma}_t(\hat{\theta}_n) - \sigma_t(\hat{\theta}_n)|^2}{\sigma_t^2(\hat{\theta}_n)} + 2 \frac{|\tilde{\sigma}_t(\hat{\theta}_n) - \sigma_t(\hat{\theta}_n)|}{\sigma_t(\hat{\theta}_n)} \right. \\
&\quad \left. + \frac{|\sigma_t(\theta) - \tilde{\sigma}_t(\theta)|^2}{\sigma_t^2(\theta)} + 2 \frac{|\sigma_t(\theta) - \tilde{\sigma}_t(\theta)|}{\sigma_t(\theta)} \right) \eta_t^{*2} \\
&\leq \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\underline{\omega}^2} \left( \frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} + 2 \frac{C_1 \rho^t}{\underline{\omega}} + \frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} + 2 \frac{C_1 \rho^t}{\underline{\omega}} \right) \eta_t^{*2} \\
&\leq \left( \frac{2C_1^2}{\underline{\omega}^4} + \frac{4C_1}{\underline{\omega}^3} \right) \frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \eta_t^{*2}.
\end{aligned}$$

To verify (A.52) we are left to show that  $\frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \eta_t^{*2} \xrightarrow{p^*} 0$  almost surely. For every  $\varepsilon > 0$ , Markov's inequality implies

$$\mathbb{P}^* \left[ \frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \eta_t^{*2} \geq \varepsilon \right] \leq \frac{1}{\varepsilon} \frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \mathbb{E}^* [\eta_t^{*2}]$$

As  $\mathbb{E}^* [\eta_t^{*2}] \xrightarrow{a.s.} 1$  (Lemma 4), it remains to show that  $\frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \xrightarrow{a.s.} 0$ . We have

$$\frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) = \frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\theta_0) \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta_0)} \leq \left( \frac{1}{n} \sum_{t=1}^n \rho^{2t} \sigma_t^4(\theta_0) \right)^{\frac{1}{2}} \left( \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^4(\hat{\theta}_n)}{\sigma_t^4(\theta_0)} \right)^{\frac{1}{2}}$$

by the Cauchy-Schwarz inequality. Since  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  (Theorem 1) such that  $\hat{\theta}_n \in \mathcal{V}(\theta_0)$

almost surely, the uniform ergodic theorem and Assumption 9(i) result in

$$\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^4(\hat{\theta}_n)}{\sigma_t^4(\theta_0)} \stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n T_t^4 \stackrel{a.s.}{\rightarrow} \mathbb{E}[T_t^4] < \infty.$$

In addition, we have for  $\delta > 0$

$$\sum_{t=1}^{\infty} \mathbb{P}[\rho^{2t} \sigma_t^4(\theta_0) > \delta] \leq \sum_{t=1}^{\infty} \frac{\rho^{st/2} \mathbb{E}[\sigma_t^s(\theta_0)]}{\delta^{s/(4)}} = \frac{\mathbb{E}[\sigma_t^s(\theta_0)]}{\delta^{s/(4)}(1 - \rho^{s/2})} < \infty$$

such that the Borel-Cantelli Lemma implies  $\rho^{2t} \sigma_t^4(\theta_0) \xrightarrow{a.s.} 0$  as  $t \rightarrow \infty$ . Therefore,  $\frac{1}{n} \sum_{t=1}^n \rho^{2t} \sigma_t^4(\theta_0) \xrightarrow{a.s.} 0$  follows by Cesàro's lemma. Combining results, we establish  $\frac{1}{n} \sum_{t=1}^n \rho^t \sigma_t^2(\hat{\theta}_n) \xrightarrow{a.s.} 0$ , which verifies (A.52) and completes *Step 1*.

Consider *Step 2*; by compactness of  $\mathcal{B}$  the Heine-Borel theorem entails that there exists a finite number of neighborhoods of size smaller than  $1/k$ , i.e.  $\mathcal{V}_k(\theta_1), \dots, \mathcal{V}_k(\theta_K)$  with  $K = K(k) \in \mathbb{N}$ , covering  $\mathcal{B}$ . We have

$$\sup_{\theta \in \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) = \max_{i=1, \dots, K} \sup_{\theta \in \mathcal{V}_k(\theta_i) \cap \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n).$$

Next, we fix  $i \in \{1, \dots, K\}$ . With regard to *Step 1*, we obtain for each  $M > 1$

$$\begin{aligned}
& L_n^*(\theta) - L_n^*(\hat{\theta}_n) \\
&= \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{\frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} > M\right\}} \underbrace{\left(1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \eta_t^{*2} + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)}\right)}_{\geq 0} \\
&\quad + \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{\frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \leq M\right\}} \left(1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \eta_t^{*2} + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)}\right) + R_n^*(\theta) \\
&\leq \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{\frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} > M\right\}} \left(1 + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)}\right) + \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{\frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \leq M\right\}} \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} (1 - \eta_t^{*2}) \\
&\quad + \frac{1}{2n} \sum_{t=1}^n \mathbb{1}_{\left\{\frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \leq M\right\}} \left(1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)}\right) + R_n^*(\theta)
\end{aligned}$$

such that

$$\begin{aligned}
& \sup_{\theta \in \mathcal{V}_k(\theta_i) \cap \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) \\
&\stackrel{a.s.}{\leq} \underbrace{\frac{1}{2} \frac{1}{n} \sum_{t=1}^n \sup_{\substack{\|\dot{\theta} - \theta_0\| \leq 1/k \\ \|\theta - \theta_i\| \leq 1/k}} \mathbb{1}_{\left\{\frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} > M\right\}} \left(1 + \log \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)}\right)}_I \\
&\quad + \underbrace{\frac{1}{2} \frac{1}{n} \sum_{t=1}^n \sup_{\substack{\|\dot{\theta} - \theta_0\| \leq 1/k \\ \|\theta - \theta_i\| \leq 1/k}} \mathbb{1}_{\left\{\frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} \leq M\right\}} \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} (1 - \eta_t^{*2})}_{II} \\
&\quad + \underbrace{\frac{1}{2} \frac{1}{n} \sum_{t=1}^n \sup_{\substack{\|\dot{\theta} - \theta_0\| \leq 1/k \\ \|\theta - \theta_i\| \leq 1/k}} \mathbb{1}_{\left\{\frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} \leq M\right\}} \left(1 - \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} + \log \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)}\right)}_{III} + \underbrace{\sup_{\theta \in \Theta} |R_n^*(\theta)|}_{IV}.
\end{aligned}$$

Subsequently, we consider each term in turn. Regarding *I*, take  $k$  sufficiently large such that  $\dot{\theta}$  satisfying  $\|\dot{\theta} - \theta_0\| \leq 1/k$  yields  $\dot{\theta} \in \mathcal{V}(\theta_0)$ . The uniform ergodic theorem,



the inequality  $\log(x) \leq x$  for all  $x > 0$  and the Cauchy-Schwarz inequality imply

$$\begin{aligned}
I &\xrightarrow{a.s.} \mathbb{E} \left[ \sup_{\substack{\|\dot{\theta} - \theta_0\| \leq 1/k \\ \|\theta - \theta_i\| \leq 1/k}} \mathbb{1}_{\left\{ \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} > M \right\}} \left( 1 + \log \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} \right) \right] \leq \mathbb{E} \left[ \mathbb{1}_{\{\sigma_t^2 T_t^2 > M \underline{\omega}^2\}} \left( 1 + \log \frac{\sigma_t^2 T_t^2}{\underline{\omega}^2} \right) \right] \\
&= \mathbb{E} \left[ \mathbb{1}_{\{\sigma_t^2 T_t^2 > M \underline{\omega}^2\}} \left( 1 - 2 \log \underline{\omega} + \frac{4}{s} \log \sigma_t^{s/2} + 2 \log T_t \right) \right] \\
&\leq \mathbb{E} \left[ \mathbb{1}_{\{\sigma_t^2 T_t^2 > M \underline{\omega}^2\}} \left( 1 - 2 \log \underline{\omega} + \frac{4}{s} \sigma_t^{s/2} + 2 T_t \right) \right] \\
&\leq \underbrace{\left( \mathbb{E} \left[ \left( 1 - 2 \log \underline{\omega} + \frac{4}{s} \sigma_t^{s/2} + 2 T_t \right)^2 \right] \right)^{\frac{1}{2}}}_{I_1} \underbrace{\left( \mathbb{P} \left[ \sigma_t^2 T_t^2 > M \underline{\omega}^2 \right] \right)^{\frac{1}{2}}}_{I_2}
\end{aligned}$$

with  $\sigma_t = \sigma_t(\theta_0)$ . Employing (A.15) we find that

$$I_1 \leq 4 \left( 1 + (2 \log \underline{\omega})^2 + \frac{16}{s^2} \mathbb{E}[\sigma_t^s] + 4 \mathbb{E}[T_t^2] \right) < \infty$$

and using Markov's inequality the second subterm can be bounded by

$$I_2 \leq \mathbb{P} \left[ T_t^2 > M \underline{\omega}^2 / 2 \right] + \mathbb{P} \left[ \sigma_t^2 > M \underline{\omega}^2 / 2 \right] \leq \frac{2}{M \underline{\omega}^2} \mathbb{E}[T_t^2] + \left( \frac{2}{\sqrt{M \underline{\omega}}} \right)^s \mathbb{E}[\sigma_t^s].$$

Since  $I_1$  can be made arbitrarily small by the choice of  $M$  we get  $I = o(1)$  almost surely. Further, for given  $M$ , Lemma 4 entails

$$\left| \mathbb{E}^*[II] \right| \leq M \left| 1 - \mathbb{E}^*[\eta_t^{*2}] \right| \xrightarrow{a.s.} 0 \quad \text{and} \quad \mathbb{V}\text{ar}^*[II] \leq \frac{M^2}{n} \mathbb{V}\text{ar}^*[\eta_t^{*2}] \xrightarrow{a.s.} 0$$

such that  $II \xrightarrow{P^*} 0$  almost surely. Consider  $III$ ; the uniform ergodic theorem yields

$$III \xrightarrow{a.s.} \mathbb{E} \left[ \sup_{\substack{\|\dot{\theta} - \theta_0\| \leq 1/k \\ \|\theta - \theta_i\| \leq 1/k}} \mathbb{1}_{\left\{ \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} \leq M \right\}} \left( 1 - \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} + \log \frac{\sigma_t^2(\dot{\theta})}{\sigma_t^2(\theta)} \right) \right]$$

and the right-hand side approaches

$$\mathbb{E} \left[ 1 - \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_i)} + \log \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_i)} \right] \quad (\text{A.53})$$

as  $M$  and  $k$  grow large. Thus, almost surely, *III* can be made arbitrarily close to (A.53) by choosing  $M$  and  $k$  sufficiently large. Further, since  $\theta_i \in \mathcal{B}$ , we have  $\theta_i \neq \theta_0$  and Assumption 3 implies  $\frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta_i)} \neq 1$  almost surely. The elementary inequality  $1 - x + \log x \leq 0$  for  $x > 0$ , which holds with equality if and only if  $x = 1$ , implies that (A.53) is strictly smaller than 0. We conclude that there exists a  $\zeta_i < 0$  such that  $\text{III} < \zeta_i$  holds for sufficiently large  $M$  and  $k$  and  $n$  almost surely. Set  $\zeta = \max_{i=1, \dots, K} \zeta_i$ , which satisfies  $\zeta < 0$ . Combining results we complete *Step 2*.

Consider *Step 3*; if  $\hat{\theta}_n^* \in \mathcal{B}$ , then (4.1) yields

$$\sup_{\theta \in \mathcal{B}} L_n^*(\theta) = L_n^*(\hat{\theta}_n^*) \geq L_n^*(\hat{\theta}_n).$$

and by monotonicity of the probability measure  $\mathbb{P}^*$  we obtain

$$\mathbb{P}^*[\hat{\theta}_n^* \in \mathcal{B}] \leq \mathbb{P}^* \left[ \sup_{\theta \in \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) \geq 0 \right].$$

Together with *Step 2* we obtain

$$\mathbb{P}^*[\hat{\theta}_n^* \in \mathcal{B}] \leq \mathbb{P}^*[\zeta/2 + S_n^* > 0] + o(1) \leq \mathbb{P}^*[|S_n^*| > -\zeta/2] + o(1) = o(1)$$

almost surely, which completes *Step 3* and establishes the lemma's claim.  $\square$

**Lemma 6.** *Suppose Assumptions 1–4, 5(i), 5(iii), 6 and 9 hold with  $a = \pm 12$ ,  $b = 12$  and  $c = 6$ . Then, we have  $\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) \xrightarrow{P^*} -2J$  almost surely for  $\check{\theta}_n$  between  $\hat{\theta}_n^*$  and  $\hat{\theta}_n$ .*

*Proof.* We have

$$\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) = \underbrace{\frac{1}{n} \sum_{t=1}^n \left( \frac{\epsilon_t^{*2}}{\tilde{\sigma}_t^2(\check{\theta}_n)} - 1 \right) \tilde{H}_t(\check{\theta}_n)}_I - \underbrace{\frac{1}{n} \sum_{t=1}^n \left( 3 \frac{\epsilon_t^{*2}}{\tilde{\sigma}_t^2(\check{\theta}_n)} - 1 \right) \tilde{D}_t(\check{\theta}_n) \tilde{D}'_t(\check{\theta}_n)}_{II}.$$

Employing  $\epsilon_t^* = \tilde{\sigma}_t(\hat{\theta}_n) \eta_t^*$  the first term can be expanded as follows:

$$I = \underbrace{\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} H_t(\check{\theta}_n) \eta_t^{*2}}_{I_1} + \underbrace{\frac{1}{n} \sum_{t=1}^n \left( \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \tilde{H}_t(\check{\theta}_n) - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} H_t(\check{\theta}_n) \right) \eta_t^{*2}}_{I_2} - \underbrace{\frac{1}{n} \sum_{t=1}^n \tilde{H}_t(\check{\theta}_n)}_{I_3}.$$

Consider  $I_1$ ; we take  $\varepsilon > 0$  and denote the unit vectors spanning  $\mathbb{R}^r$  by  $e_1, \dots, e_r$ .

Since  $\frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} H_t(\theta_2)$  is continuous in  $\theta_1$  and  $\theta_2$  we can take  $\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$  such that

$$\begin{aligned} \mathbb{E}[e'_i H_t e_j] - \varepsilon &< \mathbb{E} \left[ \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i H_t(\theta_2) e_j \right] \\ &\leq \mathbb{E} \left[ \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i H_t(\theta_2) e_j \right] < \mathbb{E}[e'_i H_t e_j] + \varepsilon \end{aligned}$$

for all  $i, j = 1, \dots, r$ . Since  $\check{\theta}_n$  lies between  $\hat{\theta}_n^*$  and  $\hat{\theta}_n$ , Theorem 1 and Lemma 5 imply  $\check{\theta}_n \xrightarrow{p^*} \theta_0$  almost surely. Since  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  and  $\check{\theta}_n \xrightarrow{p^*} \theta_0$  almost surely, we have  $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$  almost surely and  $\check{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$  with conditional probability close to one almost surely.

In such case, we have for all pairs  $(i, j)$

$$L_n^*(i, j) \leq \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} e'_i H_t(\check{\theta}_n) e_j \eta_t^{*2} \leq U_n^*(i, j)$$

with

$$L_n^*(i, j) = \frac{1}{n} \sum_{t=1}^n \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i H_t(\theta_2) e_j \eta_t^{*2}$$

$$U_n^*(i, j) = \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i H_t(\theta_2) e_j \eta_t^{*2}.$$

Using the uniform ergodic theorem, the conditional mean of the upper bound satisfies

$$\mathbb{E}^*[U_n^*(i, j)] = \mathbb{E}^*[\eta_t^{*2}] \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i H_t(\theta_2) e_j$$

$$\xrightarrow{a.s.} \mathbb{E} \left[ \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i H_t(\theta_2) e_j \right] < \mathbb{E}[e'_i H_t e_j] + \varepsilon.$$

whereas its conditional variance vanishes:

$$\mathbb{V}\text{ar}^*[U_n^*(i, j)] = \mathbb{V}\text{ar}^*[\eta_t^{*2}] \frac{1}{n^2} \sum_{t=1}^n \left( \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i H_t(\theta_2) e_j \right)^2 \leq \mathbb{V}\text{ar}^*[\eta_t^{*2}] \frac{1}{n^2} \sum_{t=1}^n S_t^4 T_t^4 V_t^2$$

$$\leq \mathbb{V}\text{ar}^*[\eta_t^{*2}] \frac{1}{n} \underbrace{\left( \frac{1}{n} \sum_{t=1}^n S_t^{12} \right)}_{\xrightarrow{a.s.} \mathbb{E}[S_t^{12}] < \infty} \underbrace{\left( \frac{1}{n} \sum_{t=1}^n T_t^{12} \right)}_{\xrightarrow{a.s.} \mathbb{E}[T_t^{12}] < \infty} \underbrace{\left( \frac{1}{n} \sum_{t=1}^n V_t^6 \right)}_{\xrightarrow{a.s.} \mathbb{E}[V_t^6] < \infty} \xrightarrow{a.s.} 0.$$

Similarly, we obtain for the lower bound

$$\mathbb{E}^*[L_n^*(i, j)] \xrightarrow{a.s.} \mathbb{E} \left[ \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i H_t(\theta_2) e_j \right] > \mathbb{E}[e'_i H_t e_j] - \varepsilon$$

and  $\mathbb{V}\text{ar}^*[L_n^*(i, j)] \xrightarrow{a.s.} 0$ . Taking  $\varepsilon \searrow 0$  subsequently, we get  $\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} e'_i H_t(\check{\theta}_n) e_j \eta_t^{*2} \xrightarrow{p^*} \mathbb{E}[e'_i H_t e_j]$  almost surely for all pairs  $(i, j)$ , which in turn yields  $I_1 \xrightarrow{p^*} \mathbb{E}[H_t]$  almost surely. Regarding  $I_2$ , we combine (A.11) and the elementary inequalities (A.5) with

$m = 1$ , which yields

$$\begin{aligned}
& \left| \frac{\tilde{\sigma}_t^2(\theta_1)}{\tilde{\sigma}_t^2(\theta_2)} - \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} \right| \leq \left| \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} - \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right|^2 + 2 \left| \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} - \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right| \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \\
& \leq \frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} \left( 1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right)^2 + \frac{2C_1 \rho^t}{\underline{\omega}} \left( 1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right) \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \\
& \leq \left( \frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \rho^t \left( 1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right)^2 \leq \left( \frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left( 1 + \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} \right)
\end{aligned} \tag{A.54}$$

for any  $\theta_1, \theta_2 \in \Theta$ . It follows that

$$\begin{aligned}
\|I_2\| & \leq \frac{1}{n} \sum_{t=1}^n \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \tilde{H}_t(\check{\theta}_n) - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} H_t(\check{\theta}_n) \right\| \eta_t^{*2} \\
& = \frac{1}{n} \sum_{t=1}^n \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \left( \tilde{H}_t(\check{\theta}_n) - H_t(\check{\theta}_n) \right) + \left( \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) H_t(\check{\theta}_n) \right\| \eta_t^{*2} \\
& \leq \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \left\| \tilde{H}_t(\check{\theta}_n) - H_t(\check{\theta}_n) \right\| + \left| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right| \left\| H_t(\check{\theta}_n) \right\| \right\} \eta_t^{*2} \\
& \leq \frac{1}{n} \sum_{t=1}^n \left\{ \left( \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} + \left( \frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left( 1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \right) \frac{C_1 \rho^t}{\underline{\omega}} \left( 1 + \left\| H_t(\check{\theta}_n) \right\| \right) \right. \\
& \quad \left. + \left( \frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left( 1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \left\| H_t(\check{\theta}_n) \right\| \right\} \eta_t^{*2} \\
& \leq \left( \frac{5C_1}{\underline{\omega}} + \frac{6C_1^2}{\underline{\omega}^2} + \frac{2C_1^3}{\underline{\omega}^3} \right) \frac{1}{n} \sum_{t=1}^n \rho^t \left( 1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \left( 1 + \left\| H_t(\check{\theta}_n) \right\| \right) \eta_t^{*2},
\end{aligned}$$

where the third inequality comes from (A.8) and (A.54). In the case of  $\hat{\theta}_n \in \mathcal{V}(\theta_0)$

and  $\check{\theta}_n \in \mathcal{V}(\theta_0)$ , we get

$$\frac{1}{n} \sum_{t=1}^n \rho^t \left( 1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \left( 1 + \left\| H_t(\check{\theta}_n) \right\| \right) \eta_t^{*2} \leq \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t) \eta_t^{*2}.$$

For any  $\delta > 0$  we find

$$\mathbb{P}^* \left[ \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t) \eta_t^{*2} \geq \delta \right] = \frac{\mathbb{E}^*[\eta_t^{*2}]}{\delta} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t).$$

using Markov's inequality. Moreover, for  $\varepsilon > 0$  we have

$$\begin{aligned} \sum_{t=1}^{\infty} \mathbb{P} \left[ \rho^t (1 + S_t^2 T_t^2) (1 + V_t) > \varepsilon \right] &\leq \sum_{t=1}^{\infty} \rho^t \frac{\mathbb{E}[(1 + S_t^2 T_t^2)(1 + V_t)]}{\varepsilon} \\ &= \frac{\mathbb{E}[(1 + S_t^2 T_t^2)(1 + V_t)]}{\varepsilon(1 - \rho)} < \infty \end{aligned}$$

such that the Borel-Cantelli Lemma implies  $\rho^t (1 + S_t^2 T_t^2) (1 + V_t) \xrightarrow{a.s.} 0$  as  $t \rightarrow \infty$ . Therefore,  $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t) \xrightarrow{a.s.} 0$  follows by Césaro's lemma and we get  $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + V_t) \eta_t^{*2} \xrightarrow{p^*} 0$  almost surely. Combining results gives  $\|I_2\| \xrightarrow{p^*} 0$  almost surely. Similar to the proof of Lemma 2(iii), we establish  $I_3 \xrightarrow{p^*} \mathbb{E}[H_t]$  almost surely using  $\check{\theta}_n \xrightarrow{p^*} \theta_0$  almost surely. Combining results we establish that  $I = I_1 + I_2 - I_3 \xrightarrow{p^*} 0$  almost surely. Consider the second term and expand

$$\begin{aligned} II = & 3 \underbrace{\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \eta_t^{*2}}_{II_1} + 3 \underbrace{\frac{1}{n} \sum_{t=1}^n \left( \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} D_t(\check{\theta}_n) D'_t(\check{\theta}_n) - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right) \eta_t^{*2}}_{II_2} \\ & - \underbrace{\frac{1}{n} \sum_{t=1}^n D_t(\check{\theta}_n) D'_t(\check{\theta}_n)}_{II_3}. \end{aligned}$$

We treat the subterms of  $II$  analogously to the subterms of  $I$ . We begin with  $II_1$  and take  $\varepsilon > 0$ . Since  $\frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} D_t(\theta_2) D'_t(\theta_2)$  is continuous in  $\theta_1$  and  $\theta_2$  we can take

$\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$  such that

$$\begin{aligned} \mathbb{E}[e'_i D_t D'_t e_j] - \varepsilon &< \mathbb{E} \left[ \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \right] \\ &\leq \mathbb{E} \left[ \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \right] < \mathbb{E}[e'_i D_t D'_t e_j] + \varepsilon \end{aligned}$$

for all  $i, j = 1, \dots, r$ . Since  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  and  $\check{\theta}_n \xrightarrow{p^*} \theta_0$  almost surely, we have  $\hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$  almost surely and  $\check{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0)$  with conditional probability close to one almost surely. In such case, we have for all pairs  $(i, j)$

$$\bar{L}_n^*(i, j) \leq \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} e'_i D_t(\check{\theta}_n) D'_t(\check{\theta}_n) e_j \eta_t^{*2} \leq \bar{U}_n^*(i, j)$$

with

$$\begin{aligned} \bar{L}_n^*(i, j) &= \frac{1}{n} \sum_{t=1}^n \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \eta_t^{*2} \\ \bar{U}_n^*(i, j) &= \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \eta_t^{*2}. \end{aligned}$$

Using the uniform ergodic theorem, the conditional mean of the upper bound satisfies

$$\begin{aligned} \mathbb{E}^*[\bar{U}_n^*(i, j)] &= \mathbb{E}^*[\eta_t^{*2}] \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \\ &\xrightarrow{a.s.} \mathbb{E} \left[ \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \right] < \mathbb{E}[e'_i D_t D'_t e_j] + \varepsilon \end{aligned}$$

whereas its conditional variance vanishes:

$$\begin{aligned}
\mathbb{V}\text{ar}^*[\bar{U}_n^*(i, j)] &= \mathbb{V}\text{ar}^*[\eta_t^{*2}] \frac{1}{n^2} \sum_{t=1}^n \left( \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \right)^2 \\
&\leq \mathbb{V}\text{ar}^*[\eta_t^{*2}] \frac{1}{n^2} \sum_{t=1}^n S_t^4 T_t^4 U_t^4 \\
&\leq \mathbb{V}\text{ar}^*[\eta_t^{*2}] \frac{1}{n} \underbrace{\left( \frac{1}{n} \sum_{t=1}^n S_t^{12} \right)}_{\xrightarrow{a.s.} \mathbb{E}[S_t^{12}] < \infty}^{\frac{1}{3}} \underbrace{\left( \frac{1}{n} \sum_{t=1}^n T_t^{12} \right)}_{\xrightarrow{a.s.} \mathbb{E}[T_t^{12}] < \infty}^{\frac{1}{3}} \underbrace{\left( \frac{1}{n} \sum_{t=1}^n U_t^{12} \right)}_{\xrightarrow{a.s.} \mathbb{E}[U_t^{12}] < \infty}^{\frac{1}{3}} \xrightarrow{a.s.} 0.
\end{aligned}$$

Similarly, we obtain for the lower bound

$$\mathbb{E}^*[\bar{L}_n^*(i, j)] \xrightarrow{a.s.} \mathbb{E} \left[ \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t^2(\theta_1)}{\sigma_t^2(\theta_2)} e'_i D_t(\theta_2) D'_t(\theta_2) e_j \right] > \mathbb{E}[e'_i D_t D'_t e_j] - \varepsilon$$

and  $\mathbb{V}\text{ar}^*[\bar{L}_n^*(i, j)] \xrightarrow{a.s.} 0$ . Next, we take  $\varepsilon \searrow 0$  and get  $\frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} e'_i D_t(\check{\theta}_n) D'_t(\check{\theta}_n) e'_j \eta_t^{*2} \xrightarrow{p^*} \mathbb{E}[e'_i D_t D'_t e_j]$  almost surely for all pairs  $(i, j)$ , which in turn yields  $II_1 \xrightarrow{p^*} \mathbb{E}[D_t D'_t] = J$  almost surely. Regarding  $II_2$ , we find

$$\begin{aligned}
\|II_2\| &\leq \frac{1}{n} \sum_{t=1}^n \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \tilde{D}_t(\check{\theta}_n) \tilde{D}'_t(\check{\theta}_n) - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right\| \eta_t^{*2} \\
&= \frac{1}{n} \sum_{t=1}^n \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \left( \tilde{D}_t(\check{\theta}_n) \tilde{D}'_t(\check{\theta}_n) - D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right) + \left( \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right\| \eta_t^{*2} \\
&\leq \frac{1}{n} \sum_{t=1}^n \left\{ \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} \left\| \tilde{D}_t(\check{\theta}_n) \tilde{D}'_t(\check{\theta}_n) - D_t(\check{\theta}_n) D'_t(\check{\theta}_n) \right\| + \left| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\check{\theta}_n)} - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right| \|D_t(\check{\theta}_n)\|^2 \right\} \eta_t^{*2} \\
&\leq \frac{1}{n} \sum_{t=1}^n \left\{ \left( \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} + \left( \frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left( 1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \right) \left( \frac{C_1^2}{\underline{\omega}^2} + \frac{2C_1}{\underline{\omega}} \right) \rho^t \left( 1 + \|D_t(\check{\theta}_n)\|^2 \right) \right. \\
&\quad \left. + \left( \frac{2C_1^2}{\underline{\omega}^2} + \frac{4C_1}{\underline{\omega}} \right) \rho^t \left( 1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \|D_t(\check{\theta}_n)\|^2 \right\} \eta_t^{*2} \\
&\leq \left( \frac{6C_1}{\underline{\omega}} + \frac{11C_1^2}{\underline{\omega}^2} + \frac{8C_1^3}{\underline{\omega}^3} + \frac{2C_1^4}{\underline{\omega}^4} \right) \frac{1}{n} \sum_{t=1}^n \rho^t \left( 1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \left( 1 + \|D_t(\check{\theta}_n)\|^2 \right) \eta_t^{*2},
\end{aligned}$$



where the third inequality follows from (A.6) and (A.54). In the case of  $\hat{\theta}_n \in \mathcal{V}(\theta_0)$  and  $\check{\theta}_n \in \mathcal{V}(\theta_0)$ , we get

$$\frac{1}{n} \sum_{t=1}^n \rho^t \left( 1 + \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\check{\theta}_n)} \right) \left( 1 + \|D_t(\check{\theta}_n)\|^2 \right) \eta_t^{*2} \leq \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \eta_t^{*2}.$$

For any  $\delta > 0$  we find

$$\mathbb{P}^* \left[ \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \eta_t^{*2} \geq \delta \right] = \frac{\mathbb{E}^*[\eta_t^{*2}]}{\delta} \frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2).$$

using Markov's inequality. Moreover, for  $\varepsilon > 0$  we have

$$\begin{aligned} \sum_{t=1}^{\infty} \mathbb{P} \left[ \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) > \varepsilon \right] &\leq \sum_{t=1}^{\infty} \rho^t \frac{\mathbb{E}[(1 + S_t^2 T_t^2) (1 + U_t^2)]}{\varepsilon} \\ &= \frac{\mathbb{E}[(1 + S_t^2 T_t^2) (1 + U_t^2)]}{\varepsilon(1 - \rho)} < \infty \end{aligned}$$

such that the Borel-Cantelli Lemma implies  $\rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \xrightarrow{a.s.} 0$  as  $t \rightarrow \infty$ . Therefore,  $\frac{1}{n^2} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \xrightarrow{a.s.} 0$  follows by Césaro's lemma and we get  $\frac{1}{n} \sum_{t=1}^n \rho^t (1 + S_t^2 T_t^2) (1 + U_t^2) \eta_t^{*2} \xrightarrow{p^*} 0$  almost surely. Combining results gives  $\|II_2\| \xrightarrow{p^*} 0$  almost surely. Similar to the proof of Lemma 2(ii), we establish  $II_3 \xrightarrow{p^*} \mathbb{E}[D_t D_t'] = J$  almost surely using  $\check{\theta}_n \xrightarrow{p^*} \theta_0$  almost surely. Combining results we find  $II = 3II_1 + 3II_2 - II_3 \xrightarrow{p^*} 3J + 0 - J = 2J$  almost surely. In conclusion, we have

$$\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t^*(\check{\theta}_n) = I - II \xrightarrow{p^*} -2J$$

almost surely, which completes the proof.  $\square$

**Lemma 7.** *Suppose Assumptions 1–4, 5(i), 5(iii), 6, 9 and 10 hold with  $a = -1, 4$ ,*

$b = 4$  and  $c = 2$ . Then, we have

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} \hat{D}_t(\eta_t^{*2} - 1) \\ \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha \end{pmatrix} \xrightarrow{d^*} N(0, \Upsilon_\alpha) \quad \text{with} \quad \Upsilon_\alpha = \begin{pmatrix} (\kappa - 1)J & p_\alpha \Omega \\ p_\alpha \Omega' & \alpha(1 - \alpha) \end{pmatrix}$$

almost surely.

*Proof.* Set  $\alpha_n = \mathbb{E}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}]$  and expand

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} \hat{D}_t(\eta_t^{*2} - 1) \\ \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} \hat{D}_t(\eta_t^{*2} - \mathbb{E}^*[\eta_t^{*2}]) \\ \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha_n \end{pmatrix} + \frac{1}{\sqrt{n}} \sum_{t=1}^n \begin{pmatrix} \hat{D}_t(\mathbb{E}^*[\eta_t^{*2}] - 1) \\ \alpha_n - \alpha \end{pmatrix}.$$

Consider the second term; with regard to Remark 2 we have  $\mathbb{E}^*[\eta_t^{*2}] = 1$  whenever  $\hat{\theta}_n \in \mathring{\Theta}$  under Assumption 10. Since  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0 \in \mathring{\Theta}$  by Theorem 1 and Assumption 6, we have  $\frac{1}{\sqrt{n}} \sum_{t=1}^n \hat{D}_t(\mathbb{E}^*[\eta_t^{*2}] - 1) = 0$  for sufficiently large  $n$  almost surely. Further,

$$\alpha_n = \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}} \stackrel{a.s.}{=} \frac{\lfloor n\alpha \rfloor + 1}{n} = \alpha + O(n^{-1})$$

and hence  $\frac{1}{\sqrt{n}} \sum_{t=1}^n (\alpha_n - \alpha) \xrightarrow{a.s.} 0$ . Using the Cramér-Wold device it remains to show that for each  $\lambda = (\lambda'_1, \lambda_2)' \in \mathbb{R}^{r+1}$  with  $\|\lambda\| \neq 0$

$$\underbrace{\sum_{t=1}^n \frac{1}{\sqrt{n}} \lambda' \begin{pmatrix} \hat{D}_t(\eta_t^{*2} - \mathbb{E}^*[\eta_t^{*2}]) \\ \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha_n \end{pmatrix}}_{Z_{n,t}^*} \xrightarrow{d^*} N(0, \lambda' \Upsilon_\alpha \lambda)$$

almost surely. By construction, we have  $\mathbb{E}[Z_{n,t}^*] = 0$ . Further, we obtain

$$s_n^2 = \sum_{t=1}^n \mathbb{V}\text{ar}^*[Z_{n,t}^*] = \lambda' \begin{pmatrix} \mathbb{V}\text{ar}^*[\eta_t^{*2}] \hat{J}_n & \mathbb{C}\text{ov}^*[\eta_t^{*2}, \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \hat{\Omega}_n \\ \mathbb{C}\text{ov}^*[\eta_t^{*2}, \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \hat{\Omega}_n' & \mathbb{V}\text{ar}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \end{pmatrix} \lambda.$$

Lemma 2 states  $\hat{J}_n \xrightarrow{a.s.} J$  and  $\hat{\Omega}_n \xrightarrow{a.s.} \Omega$ . Employing Lemma 4 yields

$$\begin{aligned} \mathbb{V}\text{ar}^*[\eta_t^{*2}] &= \mathbb{E}^*[\eta_t^{*4}] - \left(\mathbb{E}[\eta_t^{*2}]\right)^2 \xrightarrow{a.s.} \kappa - 1, \\ \mathbb{V}\text{ar}^*[\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] &= \alpha_n(1 - \alpha_n) \xrightarrow{a.s.} \alpha(1 - \alpha), \\ \mathbb{C}\text{ov}^*[\eta_t^{*2}, \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] &= \mathbb{E}^*[\eta_t^{*2} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] - \mathbb{E}^*[\eta_t^{*2}] \alpha_n \xrightarrow{a.s.} p_\alpha \end{aligned}$$

and  $s_n^2 \xrightarrow{a.s.} \lambda' \Upsilon_\alpha \lambda$  follows. Next, we verify Lindeberg's condition. For arbitrary  $\varepsilon > 0$

$$\sum_{t=1}^n \mathbb{E}^*[Z_{n,t}^{*2} \mathbb{1}_{\{|Z_{n,t}^*| \geq s_n \varepsilon\}}] \leq \underbrace{\sum_{t=1}^n \mathbb{E}^*[Z_{n,t}^{*2} \mathbb{1}_{\{|\eta_t^*| > C\}}]}_I + \underbrace{\sum_{t=1}^n \mathbb{E}^*[Z_{n,t}^{*2} \mathbb{1}_{\{|Z_{n,t}^*| \geq s_n \varepsilon\}} \mathbb{1}_{\{|\eta_t^*| \leq C\}}]}_{II}$$

holds, where  $C > 0$ . Employing the elementary inequalities

$$(x + y)^z \leq 2^z(x^z + y^z) \tag{A.55}$$

and  $|x - y|^z \leq x^z + y^z$  for all  $x, y, z \geq 0$  we find that

$$\begin{aligned} Z_{n,t}^{*2} &\leq \frac{4}{n} \left( (\lambda_1' \hat{D}_t)^2 (\eta_t^{*2} - \mathbb{E}^*[\eta_t^{*2}])^2 + \lambda_2^2 (\mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha_n)^2 \right) \\ &\leq \frac{4}{n} \left( (\lambda_1' \hat{D}_t)^2 (\eta_t^{*4} + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right). \end{aligned}$$

Hence, we obtain

$$\begin{aligned}
I &\leq \frac{4}{n} \sum_{t=1}^n \mathbb{E}^* \left[ \left( (\lambda'_1 \hat{D}_t)^2 (\eta_t^{*4} + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right) \mathbb{1}_{\{|\eta_t^*| > C\}} \right] \\
&= 4 \left( \lambda'_1 \hat{J}_n \lambda_1 \mathbb{E}^* [\eta_t^{*4} \mathbb{1}_{\{|\eta_t^*| > C\}}] + (\lambda'_1 \hat{J}_n \lambda_1 \mathbb{E}^* [\eta_t^{*2}]^2 + \lambda_2^2) \mathbb{E}^* [\mathbb{1}_{\{|\eta_t^*| > C\}}] \right) \\
&\stackrel{a.s.}{\rightarrow} 4 \left( \lambda'_1 J \lambda_1 \mathbb{E} [\eta_t^4 \mathbb{1}_{\{|\eta_t| > C\}}] + (\lambda'_1 J \lambda_1 \mathbb{E} [\eta_t^2]^2 + \lambda_2^2) \mathbb{E} [\mathbb{1}_{\{|\eta_t| > C\}}] \right)
\end{aligned}$$

and choosing  $C$  sufficiently large yields  $I \stackrel{a.s.}{\rightarrow} 0$ . Given a value of  $C$ , we have

$$\begin{aligned}
II &\leq \frac{4}{n} \sum_{t=1}^n \mathbb{E}^* \left[ \left( (\lambda'_1 \hat{D}_t)^2 (\eta_t^{*4} + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right) \mathbb{1}_{\{|\lambda_1|((\eta_t^{*2} + \mathbb{E}^*[\eta_t^{*2}]) \max_t \|\hat{D}_t\| + |\lambda_2| \geq \sqrt{n} s_n \varepsilon)\} \mathbb{1}_{\{|\eta_t^*| \leq C\}}} \right] \\
&\leq \frac{4}{n} \sum_{t=1}^n \left( (\lambda'_1 \hat{D}_t)^2 (C^4 + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right) \mathbb{1}_{\{|\lambda_1|((C^2 + \mathbb{E}^*[\eta_t^{*2}]) \max_t \|\hat{D}_t\| + |\lambda_2| \geq \sqrt{n} s_n \varepsilon)\}} \\
&= 4 \left( \lambda'_1 \hat{J}_n \lambda_1 (C^4 + \mathbb{E}^*[\eta_t^{*2}]^2) + \lambda_2^2 \right) \mathbb{1}_{\{|\lambda_1|((C^2 + \mathbb{E}^*[\eta_t^{*2}]) \max_t \|\hat{D}_t\| + |\lambda_2| \geq \sqrt{n} s_n \varepsilon)\}} \\
&\stackrel{a.s.}{\rightarrow} 4 \left( \lambda'_1 J \lambda_1 (C^4 + \mathbb{E}[\eta_t^2]^2) + \lambda_2^2 \right) \times 0 = 0
\end{aligned}$$

To appreciate why the indicator function converges to 0 almost surely we employ

(A.3) as well as (A.55) and note  $\hat{\theta}_n \in \mathcal{V}(\theta_0)$  almost surely to get

$$\begin{aligned}
\frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\|^4 &\leq \frac{1}{n} \sum_{t=1}^n \left( \|D_t(\hat{\theta}_n)\| + \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|D_t(\hat{\theta}_n)\|) \right)^4 \\
&\stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n \left( U_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + U_t) \right)^4 \leq 2^4 \left( \frac{1}{n} \sum_{t=1}^n U_t^4 + \frac{C_1^4}{\underline{\omega}^4} \frac{1}{n} \sum_{t=1}^n \{\rho^t (1 + U_t)\}^4 \right). \tag{A.56}
\end{aligned}$$

The uniform ergodic theorem and Assumption 9(ii) imply  $\frac{1}{n} \sum_{t=1}^n U_t^4 \stackrel{a.s.}{\rightarrow} \mathbb{E}[U_t^4] < \infty$ .

Further, (A.4) leads to  $\rho^t (1 + U_t) \stackrel{a.s.}{\rightarrow} 0$  as  $t \rightarrow \infty$ , which in turn implies  $\{\rho^t (1 + U_t)\}^4 \stackrel{a.s.}{\rightarrow} 0$  as  $t \rightarrow \infty$ . Cesàro's lemma yields  $\frac{1}{n} \sum_{t=1}^n \{\rho^t (1 + U_t)\}^4 \stackrel{a.s.}{\rightarrow} 0$  and we have

$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \|\hat{D}_t\|^4 < \infty$  almost surely. Thus,  $\max_t \|\hat{D}_t\|/\sqrt{n} \xrightarrow{a.s.} 0$  as

$$\left( \frac{\max_t \|\hat{D}_t\|}{\sqrt{n}} \right)^4 \leq \frac{1}{n^2} \sum_{t=1}^n \|\hat{D}_t\|^4 \xrightarrow{a.s.} 0.$$

and  $\mathbb{1}_{\{|\lambda_1|(C^2 + \mathbb{E}^*[\eta_t^{*2}]) \max_t \|\hat{D}_t\| + |\lambda_2| \geq \sqrt{n} s_n \varepsilon\}} \xrightarrow{a.s.} 0$  follows. Combining results, establishes  $\frac{1}{s_n^2} \sum_{t=1}^n \mathbb{E}^*[Z_{n,t}^{*2} \mathbb{1}_{\{|Z_{n,t}^*| \geq s_n \varepsilon\}}] \xrightarrow{a.s.} 0$ . The Central Limit Theorem for triangular arrays (cf. Billingsley, 1986, Theorem 27.3) implies that  $\sum_{t=1}^n Z_{n,t}^*$  converges in conditional distribution to  $N(0, \lambda' \Upsilon_\alpha \lambda)$  almost surely.  $\square$

**Lemma 8.** *Suppose Assumptions 1–9 hold with  $a = \pm 6$ ,  $b = 6$  and  $c = 2$ . Then, we have  $I_n^*(z) \xrightarrow{p^*} \frac{z^2}{2} f(\xi_\alpha)$  in probability.*

*Proof.* Using Fubini's theorem, the conditional expectation is equal to

$$\begin{aligned} \mathbb{E}^*[I_n^*(z)] &= \sum_{t=1}^n \int_0^{z/\sqrt{n}} \mathbb{E}^*\left[\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}\right] ds \\ &= n \int_0^{z/\sqrt{n}} \left( \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} + s) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha}-) \right) ds \\ &= \int_0^z \sqrt{n} \left( \hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}}\right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha}-) \right) du \\ &= \underbrace{\int_0^z \sqrt{n} \left( \hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}}\right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha}-) - F\left(\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}}\right) + F(\hat{\xi}_{n,\alpha}) \right) du}_I \\ &\quad + \underbrace{\int_0^z \sqrt{n} \left( F\left(\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}}\right) - F(\hat{\xi}_{n,\alpha}) \right) du}_{II}. \end{aligned}$$

Regarding  $I$ , take  $\varrho \in (0, 1/2)$  and set  $\bar{\mathcal{I}}_n = [\xi_\alpha - 0.5n^{-\varrho}, \xi_\alpha + 0.5n^{-\varrho}]$ . Since  $\sqrt{n}(\hat{\xi}_{n,\alpha} - \xi_\alpha) = O_p(1)$ , the probabilities of the events  $\{\hat{\xi}_{n,\alpha} + \frac{|z|}{\sqrt{n}} \notin \bar{\mathcal{I}}_n\}$  and  $\{\hat{\xi}_{n,\alpha} - \frac{|z|}{\sqrt{n}} \notin \bar{\mathcal{I}}_n\}$  can be made arbitrarily small for large  $n$ . If  $\hat{\xi}_{n,\alpha} + \frac{|z|}{\sqrt{n}} \in \bar{\mathcal{I}}_n$  and  $\hat{\xi}_{n,\alpha} - \frac{|z|}{\sqrt{n}} \in \bar{\mathcal{I}}_n$ , then

$\hat{\xi}_{n,\alpha} \in \bar{\mathcal{I}}_n$  and  $\hat{\xi}_{n,\alpha} + \frac{u}{\sqrt{n}} \in \bar{\mathcal{I}}_n$  belong to  $\bar{\mathcal{I}}_n$  for all  $u$  between 0 and  $z$ . In that case

$$|I| \leq |z| \sup_{x,y \in \bar{\mathcal{I}}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y-)) - \sqrt{n}(F(x) - F(y)) \right| \xrightarrow{p} 0$$

by Lemma 3. Focusing on  $II$ , the mean value theorem implies that

$$II = \int_0^z u f(\hat{\xi}_{n,\alpha} + \varepsilon_n) du = \underbrace{\int_0^z u \left( f(\hat{\xi}_{n,\alpha} + \varepsilon_n) - f(\xi_\alpha) \right) du}_{II_1} + \underbrace{\int_0^z u f(\xi_\alpha) du}_{II_2}$$

with  $\varepsilon_n$  lying between 0 and  $u/\sqrt{n}$ . Since  $|\varepsilon_n| \leq |z|/\sqrt{n}$  and  $\hat{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha$  we have

$$|II_1| \leq \frac{z^2}{2} \sup_{|v| \leq |z|} \left| f\left(\hat{\xi}_{n,\alpha} + \frac{v}{n}\right) - f(\xi_\alpha) \right| \xrightarrow{a.s.} 0.$$

Further,  $II_2$  simplifies to  $II_2 = \frac{z^2}{2} f(\xi_\alpha)$  and combining results establishes

$$\mathbb{E}^*[I_n^*(z)] \xrightarrow{p} \frac{z^2}{2} f(\xi_\alpha).$$

The conditional variance vanishes in probability as

$$\begin{aligned} \mathbb{V}\text{ar}^*[I_n^*(z)] &= \sum_{t=1}^n \mathbb{V}\text{ar}^* \left[ \int_0^{z/\sqrt{n}} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) ds \right] \\ &\leq \sum_{t=1}^n \frac{|z|}{\sqrt{n}} \mathbb{E}^* \left[ \int_0^{z/\sqrt{n}} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}) ds \right] \\ &= \frac{|z|}{\sqrt{n}} \mathbb{E}^*[I_n^*(z)] \xrightarrow{p} 0, \end{aligned}$$

where the inequality follows from the fact that

$$\mathbb{V}\text{ar}(Y) \leq |c| \mathbb{E}[Y] \tag{A.57}$$

with  $Y = \int_0^c (\mathbb{1}_{\{X \leq s\}} - \mathbb{1}_{\{X < 0\}}) ds$ ,  $X$  being a real-valued integrable random variable and  $c \in \mathbb{R}$  (cf. Francq and Zakoian, 2015, p. 171).  $\square$

**Lemma 9.** *Suppose Assumptions 1–10 hold with  $a = \pm 12$ ,  $b = 12$  and  $c = 6$ . Then,  $J_{n,1}^*(z)$  given in (4.6) satisfies  $J_{n,1}^*(z) \xrightarrow{d^*} \Gamma(\frac{r}{2}, \frac{\kappa-1}{4} \xi_\alpha^2 f(\xi_\alpha))$  in probability, i.e. a Gamma distribution with shape parameter  $\frac{r}{2}$  and scale parameter  $\frac{\kappa-1}{4} \xi_\alpha^2 f(\xi_\alpha)$ .*

*Proof.* We set  $\bar{\xi}_{n,\alpha} = \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}$  and define for  $z \in \mathbb{R}$  and  $u \in \mathbb{R}^r$

$$\begin{aligned} T_n^* &= T_n^*(z, u) = \sum_{t=1}^n \tau_t^* \\ \tau_t^* &= \tau_t^*(z, u) = \int_0^{(1-\tilde{\lambda}_t^{-1}(u))\eta_t^*} (\mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha} \leq s\}} - \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha} < 0\}}) ds \\ \tilde{\lambda}_t &= \tilde{\lambda}_t(u) = \frac{\tilde{\sigma}_t(\hat{\theta}_n + n^{-1/2}u)}{\tilde{\sigma}_t(\hat{\theta}_n)}, \end{aligned}$$

where we suppress the dependence of  $\tau_t^*$  and  $\tilde{\lambda}_t$  on  $n$  and drop the arguments  $z$  and  $u$  at times for notational simplicity. Further, we split  $T_n^*$  into  $T_{n,1}^* = \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} \tau_t^*$  and  $T_{n,2}^* = \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t < 1\}} \tau_t^*$ . Let  $A > 0$ ; We establish the lemma's claim in three steps:

$$\text{Step 1: } T_{n,k}^*(z, u) \xrightarrow{p^*} \begin{cases} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u] & \text{if } k = 1 \\ \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u < 0\}} u' D_t D'_t u] & \text{if } k = 2 \end{cases}$$

in probability for all  $z \in \mathbb{R}$  and for all  $u \in \{u \in \mathbb{R}^r : \|u\| \leq A\}$ ;

*Step 2:*  $\sup_{\|u\| \leq A} |T_n^*(z, u) - \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) u' J u| \xrightarrow{p^*} 0$  in probability for all  $z \in \mathbb{R}$ ;

*Step 3:*  $J_{n,1}^*(z) \xrightarrow{d^*} \Gamma(\frac{r}{2}, \frac{\kappa-1}{4} \xi_\alpha^2 f(\xi_\alpha))$  in probability.

Consider *Step 1*; using the identity  $\int_0^c (\mathbb{1}_{\{x \leq s\}} - \mathbb{1}_{\{x < 0\}}) ds = (x-c)(\mathbb{1}_{\{c \leq x < 0\}} - \mathbb{1}_{\{0 \leq x < c\}})$

for  $c, s, x \in \mathbb{R}$  we rewrite  $\tau_t^*$  yielding

$$T_{n,1}^* = \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} \underbrace{\tilde{\lambda}_t^{-1} \int_0^{(1-\tilde{\lambda}_t)\bar{\xi}_{n,\alpha}} \left( \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha} \leq s\}} - \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha} < 0\}} \right) ds}_{=\tau_t^*}.$$

Using Fubini's theorem and expanding, the bootstrap mean of  $T_{n,1}^*$  is equal to

$$\begin{aligned} \mathbb{E}^*[T_{n,1}^*] &= \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} \tilde{\lambda}_t^{-1} \int_0^{(1-\tilde{\lambda}_t)\bar{\xi}_{n,\alpha}} \left( \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha} + s) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha} -) \right) ds \\ &= \underbrace{\frac{1}{2} \bar{\xi}_{n,\alpha}^2 f(\xi_\alpha) \frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} \tilde{\lambda}_t^{-1} n (\tilde{\lambda}_t - 1)^2}_I \\ &\quad + \underbrace{\sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} \tilde{\lambda}_t^{-1} \int_0^{(1-\tilde{\lambda}_t)\bar{\xi}_{n,\alpha}} \left( F(\bar{\xi}_{n,\alpha} + s) - F(\bar{\xi}_{n,\alpha}) - s f(\xi_\alpha) \right) ds}_{II} \\ &\quad + \underbrace{\sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} \tilde{\lambda}_t^{-1} \int_0^{(1-\tilde{\lambda}_t)\bar{\xi}_{n,\alpha}} \left( \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha} + s) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha} -) - F(\bar{\xi}_{n,\alpha} + s) + F(\bar{\xi}_{n,\alpha}) \right) ds}_{III}. \end{aligned} \tag{A.58}$$

We consider each term in turn. Expanding  $I$  we obtain

$$I = \underbrace{\frac{1}{2} \bar{\xi}_{n,\alpha}^2 f(\xi_\alpha)}_{I_1} \left( \underbrace{\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} n (\tilde{\lambda}_t - 1)^2}_{I_2} + \underbrace{\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} (\tilde{\lambda}_t^{-1} - 1) n (\tilde{\lambda}_t - 1)^2}_{I_3} \right).$$

Theorem 1 yields  $\bar{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha$  such that  $I_1 \xrightarrow{a.s.} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha)$ . Lemma 2 implies  $I_2 \xrightarrow{a.s.} \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$ . Further, the lemma entails  $n^{1/8} \max_{t=1, \dots, n} |\tilde{\lambda}_t^{-1} - 1| \xrightarrow{a.s.} 0$  as

$$\left( n^{1/8} \max_{t=1, \dots, n} |\tilde{\lambda}_t^{-1} - 1| \right)^3 \leq \underbrace{\frac{1}{n^{1/8}} \frac{1}{n} \sum_{t=1}^n (\sqrt{n} |\tilde{\lambda}_t^{-1} - 1|)^3}_{\xrightarrow{a.s.} \mathbb{E}[|D'_t u|^3]} \xrightarrow{a.s.} 0. \tag{A.59}$$



It follows that

$$|I_3| \leq \max_{t=1,\dots,n} |\tilde{\lambda}_t^{-1} - 1| \underbrace{\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} n(\tilde{\lambda}_t - 1)^2}_{=I_2} \xrightarrow{a.s.} 0,$$

which establishes  $I \xrightarrow{a.s.} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$ . Consider  $II$  in (A.58); we define

$$\begin{aligned} \bar{\xi}_{n,\alpha}^+ &= \bar{\xi}_{n,\alpha} + \max_{t=1,\dots,n} |\tilde{\lambda}_t - 1| |\bar{\xi}_{n,\alpha}| \\ \bar{\xi}_{n,\alpha}^- &= \bar{\xi}_{n,\alpha} - \max_{t=1,\dots,n} |\tilde{\lambda}_t - 1| |\bar{\xi}_{n,\alpha}| \end{aligned}$$

and set  $\mathcal{I}_n = [\xi_\alpha - a_n, \xi_\alpha + a_n]$  with  $a_n \sim n^{-1/8} \log n$ . Similar to (A.59) we obtain

$$n^{1/8} \max_{t=1,\dots,n} |\tilde{\lambda}_t - 1| \xrightarrow{a.s.} 0 \quad (\text{A.60})$$

and together with  $\sqrt{n}(\hat{\xi}_{n,\alpha} - \xi_\alpha) = O_p(1)$  we find that  $n^{1/8}(\bar{\xi}_{n,\alpha}^+ - \xi_\alpha) \xrightarrow{p} 0$  and  $n^{1/8}(\bar{\xi}_{n,\alpha}^- - \xi_\alpha) \xrightarrow{p} 0$ . Hence, the probabilities of the events  $\{\bar{\xi}_{n,\alpha}^+ \notin \mathcal{I}_n\}$  and  $\{\bar{\xi}_{n,\alpha}^- \notin \mathcal{I}_n\}$  can be made arbitrarily small for large  $n$ . If  $\bar{\xi}_{n,\alpha}^+$  and  $\bar{\xi}_{n,\alpha}^-$  belong to  $\mathcal{I}_n$ , then

$$\begin{aligned} |II| &= \left| \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} \tilde{\lambda}_t^{-1} \int_0^{(1-\tilde{\lambda}_t)\bar{\xi}_{n,\alpha}^-} s(f(\bar{\xi}_{n,\alpha} + \varepsilon_{t,n}) - f(\xi_\alpha)) ds \right| \\ &\leq \frac{1}{2} \bar{\xi}_{n,\alpha}^2 \sup_{x \in \mathcal{I}_n} |f(x) - f(\xi_\alpha)| \underbrace{\frac{1}{n} \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} \tilde{\lambda}_t^{-1} n(\tilde{\lambda}_t - 1)^2}_{=I_2+I_3}. \end{aligned}$$

with  $\varepsilon_{t,n}$  between 0 and  $(1 - \tilde{\lambda}_t)\bar{\xi}_{n,\alpha}^-$ . As  $\mathcal{I}_n$  shrinks to  $\xi_\alpha$  and  $f$  is continuous in a neighborhood of  $\xi_\alpha$  (see Assumption 4(ii)) we have  $\sup_{x \in \mathcal{I}_n} |f(x) - f(\xi_\alpha)| \rightarrow 0$ . Together with  $\bar{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha$  and  $I_2 + I_3 \xrightarrow{a.s.} \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$  we establish  $II \xrightarrow{p} 0$ . Focusing on  $III$  in (A.58), we only consider the case of  $\hat{\xi}_{n,\alpha}^+, \hat{\xi}_{n,\alpha}^- \in \mathcal{I}_n$ . In this case

$\bar{\xi}_{n,\alpha}$  and  $\bar{\xi}_{n,\alpha} + s$  belong to  $\mathcal{I}_n$  for all  $s$  between 0 and  $(1 - \tilde{\lambda}_t)\bar{\xi}_{n,\alpha}$  for all  $t$ . We obtain

$$|III| \leq |\bar{\xi}_{n,\alpha}| \sup_{x,y \in \mathcal{I}_n} \left| \sqrt{n}(\hat{\mathbb{F}}_n(x) - \hat{\mathbb{F}}_n(y-)) - \sqrt{n}(F(x) - F(y)) \right| \frac{1}{n} \sum_{t=1}^n \sqrt{n} |\tilde{\lambda}_t^{-1} - 1| \xrightarrow{a.s.} 0$$

by  $\bar{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha$  and Lemmas 2 and 3. We conclude  $III \xrightarrow{p} 0$  and establish

$$\mathbb{E}[T_{n,1}^*] \xrightarrow{p} \frac{1}{2} \xi_\alpha f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]. \quad (\text{A.61})$$

Employing (A.57), the bootstrap variance of  $T_{n,1}^*$  is bounded by

$$\begin{aligned} \mathbb{V}\text{ar}^*[T_{n,1}^*] &= \sum_{t=1}^n \mathbb{1}_{\{\tilde{\lambda}_t > 1\}} \tilde{\lambda}_t^{-2} \mathbb{V}\text{ar}^* \left[ \int_0^{(1-\tilde{\lambda}_t)\bar{\xi}_{n,\alpha}} (\mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha} \leq s\}} - \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha} < 0\}}) ds \right] \\ &\leq \sum_{t=1}^n \tilde{\lambda}_t^{-2} |\tilde{\lambda}_t - 1| |\bar{\xi}_{n,\alpha}| \mathbb{E}^* \left[ \int_0^{(1-\tilde{\lambda}_t)\bar{\xi}_{n,\alpha}} (\mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha} \leq s\}} - \mathbb{1}_{\{\eta_t^* - \bar{\xi}_{n,\alpha} < 0\}}) ds \right] \\ &= |\bar{\xi}_{n,\alpha}| \sum_{t=1}^n \tilde{\lambda}_t^{-2} |\tilde{\lambda}_t - 1| \int_0^{(1-\tilde{\lambda}_t)\bar{\xi}_{n,\alpha}} (\hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha} + s) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha} -)) ds \\ &\leq \bar{\xi}_{n,\alpha}^2 \frac{1}{n} \sum_{t=1}^n n |\tilde{\lambda}_t^{-1} - 1|^2 (\hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^+) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^-)). \end{aligned}$$

We have  $\bar{\xi}_{n,\alpha}^2 \xrightarrow{a.s.} \xi_\alpha^2$  and  $\frac{1}{n} \sum_{t=1}^n n |\tilde{\lambda}_t^{-1} - 1|^2 \xrightarrow{a.s.} \mathbb{E}[u' D_t D'_t u]$  by Lemma 2. Moreover,  $\hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^+) - \hat{\mathbb{F}}_n(\bar{\xi}_{n,\alpha}^-) \xrightarrow{p} 0$  since  $\bar{\xi}_{n,\alpha}^+ \xrightarrow{p} \xi_\alpha$ ,  $\bar{\xi}_{n,\alpha}^- \xrightarrow{p} \xi_\alpha$  and  $\sup_{x \in \mathbb{R}} |\hat{\mathbb{F}}_n(x) - F(x)| \xrightarrow{a.s.} 0$  (Lemma 1) and  $\mathbb{V}\text{ar}^*[T_{n,1}^*] \xrightarrow{p} 0$  follows. Together with (A.61) we establish  $T_{n,1}^* \xrightarrow{p^*} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u]$  in probability. The proof of  $T_{n,2}^* \xrightarrow{p^*} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u < 0\}} u' D_t D'_t u]$  in probability is analogous and hence omitted.

Regarding *Step 2* the triangle inequality yields

$$\begin{aligned} \sup_{\|u\| \leq A} \left| T_n^*(z, u) - \text{plim}_{n \rightarrow \infty} T_n^*(z, u) \right| &\leq \sup_{\|u\| \leq A} \left| T_{n,1}^*(z, u) - \text{plim}_{n \rightarrow \infty} T_{n,1}^*(z, u) \right| \\ &\quad + \sup_{\|u\| \leq A} \left| T_{n,2}^*(z, u) - \text{plim}_{n \rightarrow \infty} T_{n,2}^*(z, u) \right|. \end{aligned} \quad (\text{A.62})$$

Let  $N \geq 1$  be an integer. We divide the (hyper-)cube  $[-A, A]^r$  into  $L = (2N)^r$  cubes with side length  $A/N$ . Let  $u_\bullet(\ell)$  and  $u^\bullet(\ell)$  denote the lower left and upper right vertex of cube  $\ell$ . For  $u$  satisfying  $u_\bullet(\ell) \leq u \leq u^\bullet(\ell)$  (element-by-element comparison) Assumption 8 implies  $\tilde{\lambda}_t(u_\bullet(\ell)) \leq \tilde{\lambda}_t(u) \leq \tilde{\lambda}_t(u^\bullet(\ell))$ . Further, Theorem 1 results in  $\bar{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha < 0$ . Thus, we have for  $n$  sufficiently large

$$\begin{aligned} T_{n,1}^*(z, u_\bullet(\ell)) &\leq T_{n,1}^*(z, u) \leq T_{n,1}^*(z, u^\bullet(\ell)) \\ T_{n,2}^*(z, u^\bullet(\ell)) &\leq T_{n,2}^*(z, u) \leq T_{n,2}^*(z, u_\bullet(\ell)). \end{aligned}$$

Let  $k \in \{1, 2\}$ ; we obtain

$$\begin{aligned} &\sup_{\|u\| \leq A} \left| T_{n,k}^*(z, u) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u) \right| \\ &\leq \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^\bullet(\ell)) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) \right| + \underbrace{\max_{1 \leq \ell \leq L} \sup_{u_\bullet(\ell) \leq u \leq u^\bullet(\ell)} \left| T_{n,k}^*(z, u^\bullet(\ell)) - T_{n,k}^*(z, u) \right|}_{A_n} \\ &\quad + \underbrace{\max_{1 \leq \ell \leq L} \sup_{u_\bullet(\ell) \leq u \leq u^\bullet(\ell)} \left| \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u) \right|}_{B_n} \end{aligned}$$

with

$$\begin{aligned}
A_n &\leq \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^\bullet(\ell)) - T_{n,k}^*(z, u_\bullet(\ell)) \right| \\
&\leq \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^\bullet(\ell)) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) \right| \\
&\quad + \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u_\bullet(\ell)) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right| \\
&\quad + \max_{1 \leq \ell \leq L} \left| \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right|
\end{aligned}$$

$$B_n \leq \max_{1 \leq \ell \leq L} \left| \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right|.$$

Hence, we establish the following bound

$$\sup_{\|u\| \leq A} \left| T_{n,k}^*(z, u) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u) \right| \leq 2IV + V + 2VI$$

with

$$\begin{aligned}
IV &= \max_{1 \leq \ell \leq L} \left| \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right| \\
V &= \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u_\bullet(\ell)) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right| \\
VI &= \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^\bullet(\ell)) - \text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u^\bullet(\ell)) \right|.
\end{aligned}$$

Regarding  $IV$ , we have for every  $u$  satisfying  $\|u\| \leq A$  that

$$\text{plim}_{n \rightarrow \infty} T_{n,k}^*(z, u) = \begin{cases} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u > 0\}} u' D_t D'_t u] & \text{if } k = 1 \\ \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E}[\mathbb{1}_{\{D'_t u < 0\}} u' D_t D'_t u] & \text{if } k = 2 \end{cases}$$

is continuous in  $u$ . Together with  $\|u^\bullet(\ell) - u_\bullet(\ell)\| \leq \frac{A}{N}$  for every  $\ell$ , it follows that

$IV$  can be made arbitrarily small by choosing  $N$  sufficiently large. Given  $N$  (and  $L$ ),  $V \xrightarrow{p^*} 0$  in probability and  $VI \xrightarrow{p^*} 0$  in probability by *Step 1*, which completes *Step 2*.

Consider *Step 3*; for each  $\varepsilon > 0$  we obtain

$$\begin{aligned} & \mathbb{P}^* \left[ \left| J_{n,1}^*(z) - \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n)' J \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) \right| \geq \varepsilon \right] \\ & \leq \mathbb{P}^* \left[ \sup_{\|u\| \leq A} \left| T_n^*(u) - \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) u' J u \right| \geq \varepsilon \right] + \mathbb{P}^* \left[ \sqrt{n} \|\hat{\theta}_n^* - \hat{\theta}_n\| > A \right]. \end{aligned}$$

With regard to Proposition 1, the second term can be made arbitrarily small for large  $n$  by choosing  $A$  sufficiently large. Given  $A$ , the first term vanishes in probability by *Step 2*. Expanding  $\frac{1}{2} = \frac{\kappa-1}{8} \frac{4}{\kappa-1}$ , we establish

$$J_{n,1}^*(z) = \frac{\kappa-1}{8} \xi_\alpha^2 f(\xi_\alpha) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n)' \frac{4}{\kappa-1} J \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) + o_{p^*}(1)$$

in probability. Proposition 1 implies that  $\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)' \frac{4}{\kappa-1} J \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) \xrightarrow{d^*} \chi_r^2$  almost surely, where  $\chi_r^2$  denotes the Chi Square distribution with  $r$  degrees of freedom. Further, note that  $Y = cQ$  with  $c > 0$  and  $Q \sim \chi_r^2$  implies  $Y \sim \Gamma(r/2, 2c)$ . It follows that  $J_{n,1}^*(z) \xrightarrow{d^*} \Gamma\left(\frac{r}{2}, \frac{\kappa-1}{4} \xi_\alpha^2 f(\xi_\alpha)\right)$  in probability, which establishes the lemma's claim.  $\square$

*Remark 6.* In the preceding proof of Lemma 9 a compactness/supremum argument is employed, in which the monotonicity condition of Assumption 8 plays a central role. In contrast, the proof of Francq and Zakoïan (2015, p.172) rests on a conditional argument involving the density of  $\eta_t$  given  $\{\hat{\theta}_n - \theta_0, \eta_u : u < t\}$ . This argument does not carry over to the residual bootstrap since the probability mass function of  $\eta_t^*$  given  $\{\hat{\theta}_n^* - \hat{\theta}_n, \eta_u^* : u < t\}$  and  $\mathcal{F}_n$  has, almost surely, a single point mass.

**Lemma 10.** *Suppose Assumptions 1–10 with  $a = \pm 12$ ,  $b = 12$  and  $c = 6$ . Then,  $J_{n,2}^*(z)$  given in (4.7) satisfies  $J_{n,2}^*(z) = z \xi_\alpha f(\xi_\alpha) \Omega' \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) + o_{p^*}(1)$  in probability.*

*Proof.* Inserting  $\hat{\eta}_t^* = \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n^*)} \eta_t^*$  into (4.7) leads to

$$J_{n,2}^*(z) = \sum_{t=1}^n \left( 1 - \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n^*)} \right) \underbrace{\eta_t^* (\mathbb{1}_{\{\eta_t^* < \xi_{n,\alpha} + \frac{z}{\sqrt{n}}\}} - \mathbb{1}_{\{\eta_t^* < \xi_{n,\alpha}\}})}_{j_{n,t}^{*(2)}(z)}. \quad (\text{A.63})$$

A Taylor expansion around  $\hat{\theta}_n$  yields

$$\begin{aligned} 1 - \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta}_n^*)} &= \frac{1}{\tilde{\sigma}_t(\hat{\theta}_n)} \frac{\partial \tilde{\sigma}_t(\hat{\theta}_n)}{\partial \theta} (\hat{\theta}_n^* - \hat{\theta}_n) \\ &\quad + \frac{1}{2} (\hat{\theta}_n^* - \hat{\theta}_n)' \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\check{\theta}_n)} \left( \frac{1}{\tilde{\sigma}_t(\check{\theta}_n)} \frac{\partial^2 \tilde{\sigma}_t(\check{\theta}_n)}{\partial \theta \partial \theta'} - \frac{2}{\tilde{\sigma}_t^2(\check{\theta}_n)} \frac{\partial \tilde{\sigma}_t(\check{\theta}_n)}{\partial \theta} \frac{\partial \tilde{\sigma}_t(\check{\theta}_n)}{\partial \theta'} \right) (\hat{\theta}_n^* - \hat{\theta}_n) \\ &= \hat{D}_t' (\hat{\theta}_n^* - \hat{\theta}_n) + \frac{1}{2} (\hat{\theta}_n^* - \hat{\theta}_n)' \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\check{\theta}_n)} \left( \tilde{H}_t(\check{\theta}_n) - 2\tilde{D}_t(\check{\theta}_n) \tilde{D}_t'(\check{\theta}_n) \right) (\hat{\theta}_n^* - \hat{\theta}_n), \end{aligned} \quad (\text{A.64})$$

where  $\check{\theta}_n$  lies between  $\hat{\theta}_n^*$  and  $\hat{\theta}_n$ . Plugging this result into (A.63) gives

$$\begin{aligned} J_{n,2}^*(z) &= \underbrace{\frac{1}{\sqrt{n}} \sum_{t=1}^n j_{n,t}^{*(2)}(z) \hat{D}_t'}_I \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) \\ &\quad + \frac{1}{2} \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n)' \underbrace{\frac{1}{n} \sum_{t=1}^n \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\check{\theta}_n)} \left( \tilde{H}_t(\check{\theta}_n) - 2\tilde{D}_t(\check{\theta}_n) \tilde{D}_t'(\check{\theta}_n) \right) j_{n,t}^{*(2)}(z)}_{II} \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n). \end{aligned}$$

With regard to Proposition 1, it suffices to show that  $I \xrightarrow{p^*} \xi_\alpha z f(\xi_\alpha) \Omega'$  in probability and  $II \xrightarrow{p^*} 0$  in probability. The conditional mean and variance of the first term are

$$\begin{aligned} \mathbb{E}^*[I] &= \sqrt{n} \mathbb{E}^*[j_{n,t}^{*(2)}] \frac{1}{n} \sum_{t=1}^n \hat{D}_t' = \sqrt{n} \mathbb{E}^*[j_{n,t}^{*(2)}(z)] \hat{\Omega}_n' \\ \mathbb{V}\text{ar}^*[I] &= \mathbb{V}\text{ar}^*[j_{n,t}^{*(2)}] \frac{1}{n} \sum_{t=1}^n \hat{D}_t \hat{D}_t' = \mathbb{V}\text{ar}^*[j_{n,t}^{*(2)}(z)] \hat{J}_n. \end{aligned} \quad (\text{A.65})$$

Lemma 2 states  $\hat{\Omega}_n \xrightarrow{a.s.} \Omega$  and  $\hat{J}_n \xrightarrow{a.s.} J$ . Further, we have  $\sqrt{n} \mathbb{E}^*[j_{n,t}^{*(2)}(z)] \xrightarrow{p} z \xi_\alpha f(\xi_\alpha)$

and  $\sqrt{n}\mathbb{E}^* \left[ \left( j_{n,t}^{*(2)}(z) \right)^2 \right] \xrightarrow{p} |z| \xi_\alpha^2 f(\xi_\alpha)$ , which implies  $\mathbb{V}\text{ar}^* \left[ j_{n,t}^{*(2)}(z) \right] \xrightarrow{p} 0$ . To appreciate why, we obtain for  $z \geq 0$

$$\begin{aligned}
\sqrt{n}\mathbb{E}^* \left[ j_{n,t}^{*(2)}(z) \right] &= \sqrt{n} \int_{\left[ \hat{\xi}_{n,\alpha}, \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \right)} x \, d\hat{\mathbb{F}}_n(x) \\
&= \left( \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \right) \sqrt{n} \hat{\mathbb{F}}_n \left( \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) - \hat{\xi}_{n,\alpha} \sqrt{n} \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) - \sqrt{n} \int_{\left[ \hat{\xi}_{n,\alpha}, \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \right)} \hat{\mathbb{F}}_n(x) \, dx \\
&= \underbrace{\hat{\xi}_{n,\alpha} \sqrt{n} \left( \hat{\mathbb{F}}_n \left( \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) \right)}_{I_1} + \underbrace{z \hat{\mathbb{F}}_n \left( \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right)}_{I_2} \\
&\quad - \underbrace{\int_{[0,z)} \hat{\mathbb{F}}_n \left( \hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}} \right) \, dy}_{I_3}.
\end{aligned}$$

Using Lemma 3 and the mean value theorem, we find

$$I_1 = \hat{\xi}_{n,\alpha} \sqrt{n} \left( F \left( \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} - \right) - F(\hat{\xi}_{n,\alpha}) \right) + o_p(1) = z \hat{\xi}_{n,\alpha} f(\hat{\xi}_{n,\alpha} + \varepsilon_n) + o_p(1),$$

where  $0 \leq \varepsilon_n \leq z/\sqrt{n}$ , and together with Theorem 1 we establish  $I_1 \xrightarrow{p} z \xi_\alpha f(\xi_\alpha)$ . Moreover, Theorem 1 and Lemma 1 imply  $I_2 \xrightarrow{p} z F(\xi_\alpha)$  and using additionally the dominated convergence theorem, we obtain  $I_3 \xrightarrow{p} z F(\xi_\alpha)$ . Hence,  $\sqrt{n}\mathbb{E}^* \left[ j_{n,t}^{*(2)}(z) \right] \xrightarrow{p} z \xi_\alpha f(\xi_\alpha)$  for  $z \geq 0$  and analogously one can show it to hold for  $z < 0$ . Similarly, we

find for  $z \geq 0$

$$\begin{aligned}
\sqrt{n}\mathbb{E}^* \left[ (j_{n,t}^{*(2)}(z))^2 \right] &= \sqrt{n} \int_{\left[\hat{\xi}_{n,\alpha}, \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}\right)} x^2 d\hat{\mathbb{F}}_n(x) \\
&= \left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}\right)^2 \sqrt{n}\hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} -\right) - \hat{\xi}_{n,\alpha}^2 \sqrt{n}\hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -) - \sqrt{n} \int_{\left[\hat{\xi}_{n,\alpha}, \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}\right)} \hat{\mathbb{F}}_n(x) dx^2 \\
&= \left(\left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}\right)^2 - \hat{\xi}_{n,\alpha}^2\right) \sqrt{n}\hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} -\right) + \hat{\xi}_{n,\alpha}^2 \sqrt{n} \left(\hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} -\right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -)\right) \\
&\quad - 2 \int_{[0,z)} \left(\hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}}\right) \hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}}\right) dy \\
&= \left(2z\hat{\xi}_{n,\alpha} + \frac{z^2}{\sqrt{n}}\right) \hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} -\right) + \hat{\xi}_{n,\alpha}^2 \sqrt{n} \left(\hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} -\right) - \hat{\mathbb{F}}_n(\hat{\xi}_{n,\alpha} -)\right) \\
&\quad - 2 \left(\hat{\xi}_{n,\alpha} \int_{[0,z)} \hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}}\right) dy + \int_{[0,z)} \frac{y}{\sqrt{n}} \hat{\mathbb{F}}_n\left(\hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}}\right) dy\right) \\
&\xrightarrow{p} 2z\xi_\alpha F(\xi_\alpha) + z\xi_\alpha^2 f(\xi_\alpha) - 2z\xi_\alpha F(\xi_\alpha) = z\xi_\alpha^2 f(\xi_\alpha)
\end{aligned}$$

and analogously for  $z < 0$ . Combining results we establish  $I \xrightarrow{p^*} \xi_\alpha z f(\xi_\alpha) \Omega'$  in probability. Consider the second term; since  $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$  (Theorem 1) and  $\hat{\theta}_n^* \xrightarrow{p^*} \theta_0$  almost surely (Lemma 5), we have  $\mathbb{P}^*[\check{\theta}_n \notin \mathcal{V}(\theta_0)] \xrightarrow{a.s.} 0$ . Thus, for every  $\varepsilon > 0$  we obtain

$$\begin{aligned}
&\mathbb{P}^* [||II|| \geq \varepsilon] \\
&\leq \mathbb{P}^* \left[ \left| \frac{1}{n} \sum_{t=1}^n \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\check{\theta}_n)} \left( \tilde{H}_t(\check{\theta}_n) - 2\tilde{D}_t(\check{\theta}_n) \tilde{D}'_t(\check{\theta}_n) \right) j_{n,t}^{*(2)} \right| \geq \varepsilon \cap \check{\theta}_n \in \mathcal{V}(\theta_0) \right] + \mathbb{P}^* [\check{\theta}_n \notin \mathcal{V}(\theta_0)] \\
&\leq \mathbb{P}^* \left[ \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left( \sup_{\theta \in \mathcal{V}(\theta_0)} ||\tilde{H}_t(\theta)|| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} ||\tilde{D}_t(\theta)||^2 \right) |j_{n,t}^{*(2)}| \geq \varepsilon \right] + o(1) \\
&\leq \frac{1}{\varepsilon} \mathbb{E}^* \left[ \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left( \sup_{\theta \in \mathcal{V}(\theta_0)} ||\tilde{H}_t(\theta)|| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} ||\tilde{D}_t(\theta)||^2 \right) |j_{n,t}^{*(2)}| \right] + o(1) \\
&= \frac{1}{\varepsilon} \mathbb{E}^* \left[ |j_{n,t}^{*(2)}| \right] \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left( \sup_{\theta \in \mathcal{V}(\theta_0)} ||\tilde{H}_t(\theta)|| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} ||\tilde{D}_t(\theta)||^2 \right) + o(1)
\end{aligned}$$

almost surely, where the third inequality follows from Markov's inequality. Because



$\mathbb{E}^* \left[ |j_{n,t}^{*(2)}| \right] \leq \mathbb{E}^* \left[ (j_{n,t}^{*(2)})^2 \right]^{\frac{1}{2}} \xrightarrow{p} 0$ , it remains to show that

$$\frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left( \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{H}_t(\theta)\| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{D}_t(\theta)\|^2 \right) \quad (\text{A.66})$$

is stochastically bounded. Using (A.8) we find

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{H}_t(\theta)\| \leq \sup_{\theta \in \mathcal{V}(\theta_0)} \left( \|H_t(\theta)\| + \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|H_t(\theta)\|) \right) \leq V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t).$$

Employing (A.11) we further have

$$\sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \leq \sup_{\theta \in \mathcal{V}(\theta_0)} \left( \frac{\sigma_t(\hat{\theta}_n)}{\sigma_t(\theta)} + \frac{C_1 \rho^t}{\underline{\omega}} \left( 1 + \frac{\sigma_t(\hat{\theta}_n)}{\sigma_t(\theta)} \right) \right) \stackrel{a.s.}{\leq} S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t).$$

In addition, (A.3) and (A.15) imply

$$\begin{aligned} \sup_{\theta \in \mathcal{V}(\theta_0)} \|\tilde{D}_t(\theta)\|^2 &\leq \sup_{\theta \in \mathcal{V}(\theta_0)} \left( \|D_t(\theta)\| + \frac{C_1 \rho^t}{\underline{\omega}} (1 + \|D_t(\theta)\|) \right)^2 \\ &\leq \sup_{\theta \in \mathcal{V}(\theta_0)} 3 \left( \|D_t(\theta)\|^2 + \frac{C_1^2 \rho^{2t}}{\underline{\omega}^2} (1 + \|D_t(\theta)\|^2) \right) \\ &\leq 3U_t^2 + \frac{3C_1^2 \rho^{2t}}{\underline{\omega}^2} (1 + U_t^2). \end{aligned}$$

Hence,

$$\begin{aligned}
& \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{Y}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \left( \sup_{\theta \in \mathcal{Y}(\theta_0)} \|\tilde{H}_t(\theta)\| + 2 \sup_{\theta \in \mathcal{Y}(\theta_0)} \|\tilde{D}_t(\theta)\|^2 \right) \\
& \stackrel{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^n \left( S_t T_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + S_t T_t) \right) \left( V_t + \frac{C_1 \rho^t}{\underline{\omega}} (1 + V_t) + 6U_t^2 + \frac{6C_1^2 \rho^{2t}}{\underline{\omega}^2} (1 + U_t^2) \right) \\
& = \underbrace{\frac{1}{n} \sum_{t=1}^n S_t T_t V_t}_{II_1} + \underbrace{\frac{6}{n} \sum_{t=1}^n S_t T_t U_t^2}_{II_2} + \underbrace{\frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t S_t T_t}_{II_3} + \underbrace{\frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t S_t T_t V_t}_{II_4} \\
& + \underbrace{\frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t V_t}_{II_5} + \underbrace{\frac{C_1}{\underline{\omega}} \frac{6}{n} \sum_{t=1}^n \rho^t U_t^2}_{II_6} + \underbrace{\frac{C_1}{\underline{\omega}} \frac{6}{n} \sum_{t=1}^n \rho^t S_t T_t U_t^2}_{II_7} + \underbrace{\frac{C_1}{\underline{\omega}} \frac{1}{n} \sum_{t=1}^n \rho^t S_t T_t V_t}_{II_8} \\
& + \underbrace{\frac{C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} V_t}_{II_9} + \underbrace{\frac{C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} S_t T_t}_{II_{10}} + \underbrace{\frac{C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} S_t T_t V_t}_{II_{11}} + \underbrace{\frac{6C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} S_t T_t}_{II_{12}} \\
& + \underbrace{\frac{C_1^3}{\underline{\omega}^2} \frac{6}{n} \sum_{t=1}^n \rho^{3t} U_t^2}_{II_{13}} + \underbrace{\frac{C_1^3}{\underline{\omega}^2} \frac{6}{n} \sum_{t=1}^n \rho^{3t} S_t T_t}_{II_{14}} + \underbrace{\frac{6C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t} S_t T_t U_t^2}_{II_{15}} + \underbrace{\frac{C_1^3}{\underline{\omega}^2} \frac{6}{n} \sum_{t=1}^n \rho^{3t} S_t T_t U_t^2}_{II_{16}} \\
& + \underbrace{\frac{C_1^2}{\underline{\omega}^2} \frac{1}{n} \sum_{t=1}^n \rho^{2t}}_{II_{17}} + \underbrace{\frac{C_1^3}{\underline{\omega}^2} \frac{6}{n} \sum_{t=1}^n \rho^{3t}}_{II_{18}}
\end{aligned}$$

From Assumption 9, the uniform ergodic theorem and Hölder's inequality, we obtain

$$II_1 \leq \left( \frac{1}{n} \sum_{t=1}^n S_t^3 \right)^{\frac{1}{3}} \left( \frac{1}{n} \sum_{t=1}^n T_t^3 \right)^{\frac{1}{3}} \left( \frac{1}{n} \sum_{t=1}^n V_t^3 \right)^{\frac{1}{3}} \stackrel{a.s.}{\rightarrow} \left( \mathbb{E}[S_t^3] \right)^{\frac{1}{3}} \left( \mathbb{E}[T_t^3] \right)^{\frac{1}{3}} \left( \mathbb{E}[V_t^3] \right)^{\frac{1}{3}} < \infty$$

and similarly we can show that  $\lim_{n \rightarrow \infty} II_2 < \infty$  almost surely. Consider  $II_3$ ; for each  $\varepsilon > 0$ , Markov's inequality and the Cauchy-Schwarz inequality yield

$$\sum_{t=1}^{\infty} \mathbb{P} \left[ \rho^t S_t T_t > \varepsilon \right] \leq \sum_{t=1}^{\infty} \rho^t \frac{1 + \mathbb{E}[S_t T_t]}{\varepsilon} = \frac{1 + (\mathbb{E}[S_t^2])^{\frac{1}{2}} (\mathbb{E}[T_t^2])^{\frac{1}{2}}}{\varepsilon(1 - \rho)} < \infty$$

and  $\frac{1}{n} \sum_{t=1}^n \rho^t S_t T_t \xrightarrow{a.s.} 0$  follows from combining the Borel-Cantelli lemma with Cesáro's lemma. Hence,  $II_3 \xrightarrow{a.s.} 0$ . Similarly we can show that the terms  $II_4, \dots, II_{16}$  vanish almost surely. Further,  $II_{17} \leq \frac{1}{n} \frac{C_1^2}{\omega^2(1-\rho^2)} \xrightarrow{a.s.} 0$  and similarly, we can prove that  $II_{18}$  vanishes almost surely, which completes the proof.  $\square$

*Proof of Corollary 1.* The proof is similar to Beutner et al. (2019, proof of Theorem 2) and given for completeness. A Taylor expansion yields

$$\sqrt{n}(\widehat{VaR}_{n,\alpha}^* - \widehat{VaR}_{n,\alpha}) = \underbrace{\begin{pmatrix} -\xi_\alpha \frac{\partial \sigma_{n+1}(\theta_0)}{\partial \theta} \\ \sigma_{n+1} \end{pmatrix}'}_{w_n} \underbrace{\begin{pmatrix} \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \\ \sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*) \end{pmatrix}}_{Z_n^*} + R_n^* \quad (\text{A.67})$$

with

$$\begin{aligned} R_n^* = & \left( \xi_\alpha \frac{\partial \sigma_{n+1}(\theta_0)}{\partial \theta'} - \hat{\xi}_{n,\alpha} \frac{\partial \tilde{\sigma}_{n+1}(\hat{\theta}_n)}{\partial \theta'} - \frac{1}{2} \bar{\xi}_{n,\alpha} (\hat{\theta}_n^* - \hat{\theta}_n)' \frac{\partial^2 \tilde{\sigma}_{n+1}(\bar{\theta}_n)}{\partial \theta \partial \theta'} \right) \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) \\ & + \left( \tilde{\sigma}_{n+1}(\hat{\theta}_n) - \sigma_{n+1}(\theta_0) + \frac{\partial \tilde{\sigma}_{n+1}(\bar{\theta}_n)}{\partial \theta'} (\hat{\theta}_n^* - \hat{\theta}_n) \right) \sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*), \end{aligned}$$

where  $\bar{\theta}_n$  lies between  $\hat{\theta}_n^*$  and  $\hat{\theta}_n$  while  $\bar{\xi}_{n,\alpha}$  lies between  $\hat{\xi}_{n,\alpha}^*$  and  $\hat{\xi}_{n,\alpha}$ . Note that  $R_n^* = o_p^*(1)$  in probability, which can easily be shown using Theorems 1 and 3 together with Assumptions 4 and 9. Further, let  $Z \sim N(0, \Sigma_\alpha)$  be generated independently of  $\{\epsilon_t, -\infty < t < \infty\}$  such that  $w_n Z$  given  $\mathcal{F}_n$  follows the conditional distribution in (3.9). Take  $\varepsilon > 0$  arbitrarily small and  $K \geq 1$  sufficiently large such that with

probability close to one  $\|w_n\| \leq K$ . In that case

$$\begin{aligned}
& \sup_{\|g\|_{BL} \leq 1} \left| \mathbb{E}^* [g(w_n Z_n^* + R_n^*)] - \mathbb{E}_Z [g(w_n Z) | \mathcal{F}_n] \right| \\
& \leq \sup_{\|g\|_{BL} \leq 1} \mathbb{E}^* \left[ |g(w_n Z_n^* + R_n^*) - g(w_n Z_n^*)| \right] \\
& \quad + \sup_{\|g\|_{BL} \leq 1} K \left| \mathbb{E}^* [g(w_n Z_n^*)/K] - \mathbb{E}_Z [g(w_n Z)/K | \mathcal{F}_n] \right| \\
& \leq \sup_{\|g\|_{BL} \leq 1} \mathbb{E}^* \left[ |g(w_n Z_n^* + R_n^*) - g(w_n Z_n^*)| (\mathbb{1}_{\{|R_n^*| \leq \varepsilon\}} + \mathbb{1}_{\{|R_n^*| > \varepsilon\}}) \right] \\
& \quad + \sup_{\|h\|_{BL} \leq 1} \left| \mathbb{E}^* [h(Z_n^*)] - \mathbb{E}_Z [h(Z) | \mathcal{F}_n] \right| \\
& \leq \varepsilon + 2 \mathbb{E}^* [\mathbb{1}_{\{|R_n^*| > \varepsilon\}}] + \sup_{\|h\|_{BL} \leq 1} \left| \mathbb{E}^* [h(Z_n^*)] - \mathbb{E}_Z [h(Z)] \right|,
\end{aligned}$$

with  $\|g\|_{BL} = \sup_x |g(x)| + \sup_{x \neq y} \frac{|g(x) - g(y)|}{\|x - y\|}$  being the bounded Lipschitz norm and  $\mathbb{E}_Z$  denoting the expectation operator corresponding to  $Z$ . Together with Theorem 3 and  $R_n^* = o_{p^*}(1)$  in probability, we obtain

$$\sup_{\|g\|_{BL} \leq 1} \left| \mathbb{E}^* [g(w_n Z_n^* + R_n^*)] - \mathbb{E}_Z [g(w_n Z) | \mathcal{F}_n] \right| \xrightarrow{p} 0,$$

which completes the proof. □

## B Recursive-design Residual Bootstrap

This appendix devotes attention to the recursive-design residual bootstrap. The bootstrap schemes described in Algorithms 4 and 5 are the recursive-design counterparts of Algorithms 1 and 2, respectively. Note that the bootstrap observation  $\epsilon_t^*$  is generated recursively on the basis of its past realizations  $\epsilon_{t-1}^*, \dots, \epsilon_1^*$ .

**Algorithm 4.** (*Recursive-design residual bootstrap*)

1. For  $t = 1, \dots, n$  generate  $\eta_t^* \stackrel{iid}{\sim} \hat{\mathbb{F}}_n$  and the bootstrap observation  $\epsilon_t^* = \sigma_t^* \eta_t^*$  with  $\sigma_t^* = \sigma_t^*(\hat{\theta}_n)$  and  $\sigma_t^*(\theta) = \sigma(\epsilon_{t-1}^*, \dots, \epsilon_1^*, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \dots; \theta)$
2. Calculate the bootstrap estimator

$$\hat{\theta}_n^* = \arg \max_{\theta \in \Theta} L_n^*(\theta)$$

with the bootstrap criterion function given by

$$L_n^*(\theta) = \frac{1}{n} \sum_{t=1}^n \ell_t^*(\theta) \quad \text{and} \quad \ell_t^*(\theta) = -\frac{1}{2} \left( \frac{\epsilon_t^*}{\sigma_t^*(\theta)} \right)^2 - \log \tilde{\sigma}_t(\theta).$$

3. For  $t = 1, \dots, n$  compute the bootstrap residual  $\hat{\eta}_t^* = \epsilon_t^* / \sigma_t^*(\hat{\theta}_n^*)$  and obtain

$$\hat{\xi}_{n,\alpha}^* = \arg \min_{z \in \mathbb{R}} \frac{1}{n} \sum_{t=1}^n \rho_\alpha(\hat{\eta}_t^* - z).$$

4. Obtain the bootstrap estimator of the conditional VaR

$$\widehat{VaR}_{n,\alpha}^* = -\hat{\xi}_{n,\alpha}^* \tilde{\sigma}_{n+1}(\hat{\theta}_n^*).$$

**Algorithm 5.** (*Recursive-design Bootstrap Confidence Intervals for VaR*)

1. Acquire a set of  $B$  bootstrap replicates, i.e.  $\widehat{VaR}_{n,\alpha}^{*(b)}$  for  $b = 1, \dots, B$ , by repeating Algorithm 4.

- 2.1. Obtain the EP interval

$$\left[ \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(1 - \gamma/2), \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{*-1}(\gamma/2) \right]$$

with  $\hat{G}_{n,B}^{\star-1}(\cdot)$  being the quantile function (generalized inverse) of  $\hat{G}_{n,B}^{\star}(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\sqrt{n}(\widehat{VaR}_{n,\alpha}^{\star(b)} - \widehat{VaR}_{n,\alpha}) \leq x\}}$ .

2.2. Calculate the RT interval

$$\left[ \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{\star-1}(\gamma/2), \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{G}_{n,B}^{\star-1}(1 - \gamma/2) \right].$$

2.3. Compute the SY interval

$$\left[ \widehat{VaR}_{n,\alpha} - \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{\star-1}(1 - \gamma), \widehat{VaR}_{n,\alpha} + \frac{1}{\sqrt{n}} \hat{H}_{n,B}^{\star-1}(1 - \gamma) \right]$$

with  $\hat{H}_{n,B}^{\star-1}(\cdot)$  being the quantile function (generalized inverse) of  $\hat{H}_{n,B}^{\star}(x) = \frac{1}{B} \sum_{b=1}^B \mathbb{1}_{\{\sqrt{n}|\widehat{VaR}_{n,\alpha}^{\star(b)} - \widehat{VaR}_{n,\alpha}| \leq x\}}$ .

In contrast to the fixed-design residual bootstrap, the bootstrap sample  $\epsilon_1^*, \dots, \epsilon_n^*$ , conditional on the original sample, is a dependent sequence. Therefore one likely needs a stronger set of conditions to show the validity of the recursive-design residual bootstrap. Moreover, whether the recursive bootstrap scheme is valid is contingent on the specific conditional volatility model, e.g. GARCH(1, 1), and as such needs to be investigated on a case-by-case basis. This is therefore outside the scope of the current paper.

## References

- Bahadur, R.R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics* 37(3), 577–580.
- Bardet, J.M., K. Kamila, and W. Kengne (2020). Consistent model selection criteria and goodness-of-fit test for common time series models. *Electronic Journal of Statistics* 14(1), 2009–2052.

- Berkes, I. and L. Horváth (2003). Limit results for the empirical process of squared residuals in GARCH models. *Stochastic Processes and their Applications* 105(2), 271–298.
- Beutner, E., A. Heinemann, and S. Smeeke (2019). A justification of conditional confidence intervals, *in revision*. <https://arxiv.org/pdf/1710.00643.pdf>.
- Billingsley, P. (1986). *Probability and Measure* (2nd ed.). New York: John Wiley & Sons.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3), 307–327.
- Cavaliere, G., R.S. Pedersen, and A. Rahbek (2018). The fixed volatility bootstrap for a class of ARCH( $q$ ) models. *Journal of Time Series Analysis* 39, 920–941.
- Chang, Y. and J.Y. Park (2003). A sieve bootstrap for the test of a unit root. *Journal of Time Series Analysis* 24(4), 379–400.
- Cho, J.S. and H. White (2011). Generalized runs tests for the iid hypothesis. *Journal of Econometrics* 162(2), 326–344.
- Christoffersen, P. and S. Gonçalves (2005). Estimation risk in financial risk management. *The Journal of Risk* 7(3), 1–28.
- Corradi, V. and E.M. Iglesias (2008). Bootstrap refinements for QML estimators of the GARCH(1,1) parameters. *Journal of Econometrics* 144(2), 500–510.
- Csörgő, M. and P. Révész (1981). *Strong Approximations in Probability and Statistics*. Budapest: Akadémiai Kiadó.
- D’Aristotile, A., P. Diaconis, and D. Freedman (1988). On merging of probabilities. *Sankhyā: The Indian Journal of Statistics, Series A* 50(3), 363–380.
- Davidson, R. and E. Flachaire (2008). The wild bootstrap, tamed at last. *Journal of Econometrics* 146, 162–169.
- Ding, Z., C.W. Granger, and R.F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1(1), 83–106.
- Dudley, R.M. (2002). *Real Analysis and Probability*. Cambridge: Cambridge University Press.
- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.
- Escanciano, J.C. (2009). Quasi-maximum likelihood estimation of semi-strong GARCH model. *Econometric Theory* 25(2), 561–570.

- Falk, M. and E. Kaufmann (1991). Coverage probabilities of bootstrap-confidence intervals for quantiles. *The Annals of Statistics* 19(1), 485–495.
- Francq, C., L. Horváth, and J.M. Zakoïan (2016). Variance targeting estimation of multivariate GARCH models. *Journal of Financial Econometrics* 14(2), 353–382.
- Francq, C. and J.M. Zakoïan (2004). Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10(4), 605–637.
- Francq, C. and J.M. Zakoïan (2011). *GARCH Models: Structure, Statistical Inference and Financial Applications*. Chichester: John Wiley & Sons.
- Francq, C. and J.M. Zakoïan (2015). Risk-parameter estimation in volatility models. *Journal of Econometrics* 184(1), 158–173.
- Francq, C. and J.M. Zakoïan (2016). Estimating multivariate volatility models equation by equation. *Journal of the Royal Statistical Society, Series B* 76(3), 613–635.
- Friedrich, M., S. Smeeke, and J.P. Urbain (2020). Autoregressive wild bootstrap inference for nonparametric trends. *Journal of Econometrics* 214, 81–109.
- Gao, F. and F. Song (2008). Estimation risk in GARCH VaR and ES estimates. *Econometric Theory* 24(5), 1404–1424.
- Geweke, J. (1986). Comment on: modelling the persistence of conditional variances. *Econometric Reviews* 5, 57–61.
- Glosten, L.R., R. Jagannathan, and D.E. Runkle (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48(5), 1779–1801.
- Hall, P., T.J. DiCiccio, and J.P. Romano (1989). On smoothing and the bootstrap. *The Annals of Statistics* 17(2), 692–704.
- Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and its Application*. New York: Academic Press.
- Hall, P. and M.A. Martin (1988). On bootstrap resampling and iteration. *Biometrika* 75(4), 661–671.
- Hall, P. and Q. Yao (2003). Inference in ARCH and GARCH models with heavy-tailed errors. *Econometrica* 71(1), 285–317.
- Hamadeh, T. and J.M. Zakoïan (2011). Asymptotic properties of LS and QML estimators for a class of nonlinear GARCH processes. *Journal of Statistical Planning and Inference* 141(1), 488–507.



- Hartz, C., S. Mittnik, and M. Paolella (2006). Accurate value-at-risk forecasting based on the normal-GARCH model. *Computational Statistics & Data Analysis* 51(4), 2295–2312.
- Heinemann, A. and S. Telg (2018). A residual bootstrap for conditional expected shortfall. *Preprint arXiv:1811.11557*.
- Hidalgo, J. and P. Zaffaroni (2007). A goodness-of-fit test for ARCH( $\infty$ ) models. *Journal of Econometrics* 141(2), 835–875.
- Hjort, N.L. and D. Pollard (2011). Asymptotics for minimisers of convex processes. *Preprint arXiv:1107.3806v1*.
- Jeong, M. (2017). Residual-based GARCH bootstrap and second order asymptotic refinement. *Econometric Theory* 33(3), 779–790.
- Jiménez-Gamero, M.D., S. Lee, and S.G. Meintanis (2019). Goodness-of-fit tests for parametric specifications of conditionally heteroscedastic models. *TEST*, to appear.
- Koenker, R. and Z. Xiao (2006). Quantile autoregression. *Journal of the American Statistical Association* 101(475), 980–990.
- Lahiri, S.N. (2003). *Resampling Methods for Dependent Data*. New York: Springer-Verlag.
- Linton, O., J. Pan, and H. Wang (2010). Estimation for a nonstationary semi-strong GARCH(1,1) model with heavy-tailed errors. *Econometric Theory* 26(1), 1–28.
- Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *Annals of Statistics* 21, 255–285.
- Nelson, D.B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society*, 347–370.
- Pantula, S.G. (1986). Modeling the persistence of conditional variances: a comment. *Econometric Reviews* 5, 79–97.
- Pascual, L., J. Romo, and E. Ruiz (2006). Bootstrap prediction for returns and volatilities in GARCH models. *Computational Statistics & Data Analysis* 50(9), 2293–2312.
- Roussas, G.G. (1997). *A Course in Mathematical Statistics* (2nd ed.). San Diego: Academic Press.
- Shao, X. (2010). The dependent wild bootstrap. *Journal of the American Statistical Association* 105, 218–235.

- Shimizu, K. (2009). *Bootstrapping Stationary ARMA–GARCH Models*. Springer.
- Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis Estimation Density*. London: Chapman and Hall.
- Spierdijk, L. (2016). Confidence intervals for ARMA–GARCH value-at-risk: the case of heavy tails and skewness. *Computational Statistics & Data Analysis* 100, 545–559.
- van der Vaart, A.W. (2000). *Asymptotic Statistics* (1st ed.). Cambridge: Cambridge University Press.
- Xiong, S. and G. Li (2008). Some results on the convergence of conditional distributions. *Statistics & Probability Letters* 78(18), 3249–3253.
- Zakoïan, J.M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control* 18(5), 931–955.