A Residual Bootstrap for Conditional Value-at-Risk

Eric Beutner† Alexander Heinemann† Stephan Smeekes†

†Department of Quantitative Economics
Maastricht University
August 29, 2018

Abstract

This paper proposes a fixed-design residual bootstrap method for the two-step estimator of [Francq and Zakoïan (2015)] associated with the conditional Value-at-Risk. The bootstrap’s consistency is proven under mild assumptions for a general class of volatility models and bootstrap intervals are constructed for the conditional Value-at-Risk to quantify the uncertainty induced by estimation. A large-scale simulation study is conducted revealing that the equal-tailed percentile interval based on the fixed-design residual bootstrap tends to fall short of its nominal value. In contrast, the reversed-tails interval based on the fixed-design residual bootstrap yields accurate coverage. In the simulation study we also consider the recursive-design bootstrap. It turns out that the recursive-design and the fixed-design bootstrap perform equally well in terms of average coverage. Yet in smaller samples the fixed-design scheme leads on average to shorter intervals. An empirical application illustrates the interval estimation using the fixed-design residual bootstrap.

Key words: Residual bootstrap; Value-at-Risk; GARCH
JEL codes: C14; C15; C58

1 Introduction

Risk management has tremendously developed in past decades becoming an increasing practice. With minimum 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 distribution by consistent estimates (Gao and Song, 2008). 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 among researchers 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.

Unfortunately, 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 proves its consistency for a general class of volatility models. Moreover, an algorithm is provided for the construction of bootstrap intervals for the conditional VaR. A simulation study is conducted to evaluate the intervals’ finite sample performance.

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. 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

$$\epsilon_t = \sigma_t \eta_t$$  \hspace{1cm} (2.1)

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

$$\sigma_{t+1} = \sigma_{t+1}(\theta_0) = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots; \theta_0)$$  \hspace{1cm} (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 $\Theta$. Subsequently, we consider two examples for the functional form of (2.2): the well-known GARCH model (Engle 1982; Bollerslev 1986) and the threshold GARCH (T-GARCH) model of Zakoïan (1994).

Example 1. Suppose $\{\epsilon_t\}$ follows a GARCH(1,1) process given by (2.1) and

$$\sigma_{t+1}^2 = \omega_0 + \alpha_0 \epsilon_t^2 + \beta_0 \sigma_t^2,$$  \hspace{1cm} (2.3)
where \( \theta_0 = (\omega_0, \alpha_0, \beta_0)' \in (0, \infty) \times [0, \infty) \times [0, 1) \). The recursive structure implies

\[
\sigma_{t+1} = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots; \theta_0) = \sqrt{\sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0 \epsilon_{t-k}^2)}.
\]  

(2.4)

**Example 2.** Suppose \( \{\epsilon_t\} \) follows a T-GARCH(1, 1) process given by (2.1) and

\[
\sigma_{t+1} = \omega_0 + \alpha_0^+ \epsilon_t^+ + \alpha_0^- \epsilon_t^- + \beta_0 \sigma_t
\]

(2.5)

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

\[
\sigma_{t+1} = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots; \theta_0) = \sum_{k=0}^{\infty} \beta_0^k (\omega_0 + \alpha_0^+ \epsilon_{t-k}^+ + \alpha_0^- \epsilon_{t-k}^-).
\]

(2.6)

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 \( G(-) \) 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 \( \text{VaR}_\alpha(X) = -F_X^{-1}(\alpha) \)\(^1\). Let \( \mathcal{F}_{t-1} \) denote the \( \sigma \)-algebra generated by \( \{\epsilon_u, u < t\} \). It follows that the conditional VaR of \( \epsilon_t \) given \( \mathcal{F}_{t-1} \) at level \( \alpha \in (0, 1) \) is

\[
\text{VaR}_\alpha(\epsilon_t | \mathcal{F}_{t-1}) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \theta_0) \text{VaR}_\alpha(\eta_t).
\]

(2.7)

For given \( \alpha \), the quantile of \( \eta_t \) is constant and can be treated as a parameter. Thus, denoting the cdf of \( \eta_t \) by \( F \) and setting \( \xi_\alpha = F^{-1}(\alpha) \), equation (2.7) reduces to

\[
\text{VaR}_\alpha(\epsilon_t | \mathcal{F}_{t-1}) = -\xi_\alpha \sigma_t(\theta_0).
\]

(2.8)

\(^1\)The negative sign is included to conform to the convention of reporting VaR as a positive number.
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

$$
\sigma_{t+1}(\theta) = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_1, \epsilon_0, \epsilon_{-1}, \ldots; \theta).
$$

(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 $\epsilon_1, \ldots, \epsilon_n$, 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

$$
\tilde{\sigma}_{t+1}(\theta) = \sigma(\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_1, \epsilon_0, \tilde{\epsilon}_{-1}, \ldots; \theta),
$$

(3.2)

which serves as an approximation for (3.1). The QML estimator of $\theta_0$ is defined as a measurable solution $\hat{\theta}_n$ of

$$
\hat{\theta}_n = \arg \max_{\theta \in \Theta} \hat{L}_n(\theta).
$$

(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 $\hat{\eta}_1, \ldots, \hat{\eta}_n$ is given by

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

where $\rho_\alpha(u) = u(\alpha - 1_{\{u < 0\}})$ is the usual asymmetric absolute loss function (c.f. Koenker and Xiao, 2006). Equivalently, we can write $\hat{\xi}_{n,\alpha} = \hat{F}_{n}^{-1}(\alpha)$ with $\hat{F}_n(x) = \frac{1}{n} \sum_{t=1}^{n} 1_{\{\hat{\eta}_t \leq x\}}$ 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 $\text{VaR}_\alpha(\epsilon_{n+1}|F_n)$ stresses the object’s conditional nature, we henceforth proceed with the abbreviation $\text{VaR}_{n,\alpha}$ for notational convenience. Employing (3.2) – (3.4) we can estimate $\text{VaR}_{n,\alpha}$ by

\[
\hat{\text{VaR}}_{n,\alpha} = -\hat{\xi}_{n,\alpha} \tilde{\sigma}_{n+1}(\hat{\theta}_n).
\]

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) For any real sequence \( \{x_i\} \), the function \( \theta \rightarrow \sigma(x_1, x_2, \ldots; \theta) \) is continuous. Almost surely, \( \sigma_t(\theta) > \omega \) for any \( \theta \in \Theta \) and some \( \omega > 0 \) 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) \( \sup_{\theta \in \Theta} |\sigma_t(\theta) - \tilde{\sigma}_t(\theta)| \leq C_1 \rho \);

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

\[
\sup_{\theta \in \Theta} \left\| \frac{\partial \sigma_t(\theta)}{\partial \theta} - \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} \right\| \leq C_1 \rho, \quad \sup_{\theta \in \Theta} \left\| \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} \right\| \leq C_1 \rho,
\]

where \( \| \cdot \| \) denotes the Euclidean norm.

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

(i) \( \eta_t \overset{iid}{\sim} F \) 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) \( \mathbb{E}[|\eta_t|^{2d}] < \infty \) for some \( d \) (to be specified);

(iv) \( M = \sup_{x \in \mathbb{R}} |x| f(x) < \infty \).

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

Assumption 7. (Non-degeneracy) There does not exist a non-zero \( \lambda \in \mathbb{R}^r \) such that \( \lambda' \frac{\partial \sigma_t(\theta)}{\partial \theta} = 0 \) 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 \( \sigma(x_1, x_2, \ldots; \theta_1) \leq \sigma(x_1, x_2, \ldots; \theta_2) \).
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).

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

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

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). Regarding the innovation process we do not need to assume $\mathbb{E}[\eta_t] = 0$ (c.f. Francq and Zakoïan, 2004, Remark 2.5). 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 slightly more than the fourth moment to be finite, i.e. $d > 2$ in Assumption 5(iii). It is needed to verify the Lyapunov condition of the central limit theorem. Assumption 5(iv) is related to Francq and Zakoïan (2015, Assumption A11), however it can possibly be relaxed at the cost of more involved proofs. In Assumption 8 the function $\sigma(x_1, x_2, \ldots; \theta)$ is presumed to be monotonically increasing in $\theta$, which is a feature shared by various stochastic volatility models (c.f. Berkes and Horváth, 2003, Lemma 4.1). 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 (c.f. Francq and Zakoïan, 2011, p. 165; Hamadeh and Zakoïan, 2011, p. 501).
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. \( \hat{\theta}_n \overset{a.s.}{\rightarrow} \theta_0 \). If in addition Assumptions 6, 8 and 9(i) hold with \( a = -1 \), then the estimator in (3.4) satisfies \( \hat{\xi}_{n,\alpha} \overset{a.s.}{\rightarrow} \xi_\alpha \).

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

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 1–4, 5(i)–(iii), 6, 7, 9 and 10 hold with \( a = b = 4 \) and \( c = d = 2 \). Then, we have

\[
\left( \sqrt{n} (\hat{\theta}_n - \theta_0), \sqrt{n} (\hat{\xi}_\alpha - \hat{\xi}_{n,\alpha}) \right) \overset{d}{\rightarrow} N(0, \Sigma_\alpha) \quad \text{with} \quad \Sigma_\alpha = \begin{pmatrix} \kappa^{-1}J^{-1} & \lambda_\alpha J^{-1} \Omega \\ \lambda_\alpha \Omega'J^{-1} & \zeta_\alpha \end{pmatrix},
\]

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

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

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 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
\[
\hat{\Omega}_n = \frac{1}{n} \sum_{t=1}^{n} \hat{D}_t, \quad \hat{J}_n = \frac{1}{n} \sum_{t=1}^{n} \hat{D}_t \hat{D}_t', \quad \hat{\kappa}_n = \frac{1}{n} \sum_{t=1}^{n} \hat{\eta}_t^4, \quad \hat{p}_{n,\alpha} = \frac{1}{n} \sum_{t=1}^{n} \hat{\eta}_t^2 \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}} - \alpha, \tag{3.7}
\]
respectively, with $\hat{D}_t = D_t(\hat{\theta}_n)$ and $D_t(\theta) = \frac{1}{\hat{\sigma}_t(\theta)} \frac{\partial \hat{\sigma}_t(\theta)}{\partial \theta}$. The strong consistency of the estimators in (3.7) follows from Lemma 2 in Appendix A.1. Moreover, kernel smoothing is commonly employed to estimate the density $f$, i.e.
\[
\hat{f}_n^S(x) = \frac{1}{nh_n} \sum_{t=1}^{n} k\left( \frac{x - \hat{\eta}_t}{h_n} \right) \tag{3.8}
\]
with kernel function $k$ and bandwidth $h_n > 0$. Whereas Gao and Song (2008) consider Lipschitz-continuous kernels, an alternative estimator is based on the uniform kernel $k(x) = \frac{1}{2} \mathbb{1}_{\{|x| \leq 1\}}$ yielding $\hat{f}_n^S(\hat{\xi}_{n,\alpha}) \xrightarrow{p} f(\xi_\alpha)$. This follows from Lemma 3 in Appendix A.2 with $v = -2w = 2h_n \sqrt{n}$. 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. (2017, Section 4) advocate a merging concept generalizing the notion of weak convergence, 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). Presuming two independent samples, one for parameter estimation and one for conditioning, the delta method
suggests that the VaR estimator, centered at \( VaR_{n,\alpha} \) and inflated by \( \sqrt{n} \), and

\[
\begin{align*}
N \left( 0, \begin{pmatrix}
-\xi_{n,\alpha} \frac{\partial \sigma_{n+1}(\hat{\theta}_n)}{\partial \theta} \\
\sigma_{n+1}
\end{pmatrix}^\prime \Sigma_{\alpha} \left( -\xi_{n,\alpha} \frac{\partial \sigma_{n+1}(\hat{\theta}_n)}{\partial \theta} \\
\sigma_{n+1}
\right) \right)
\end{align*}
\] (3.9)

given \( F_n \) 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 \( n \to \infty \). Together with Theorem 1 and \( \Sigma_{\alpha} \stackrel{p}{\to} \hat{\Sigma}_{n,\alpha} \), it yields a 100(1\( -\gamma \)%
confidence interval for \( VaR_{n,\alpha} \) with bounds (c.f. Francq and Zakoïan, 2015, Eq. (23))

\[
\hat{VaR}_{n,\alpha} \pm \Phi^{-1}(\gamma/2) \frac{1}{\sqrt{n}} \left\{ \left( -\hat{\xi}_{n,\alpha} \frac{\partial \sigma_{n+1}(\hat{\theta}_n)}{\partial \theta} \right) \hat{\Sigma}_{n,\alpha} \left( -\hat{\xi}_{n,\alpha} \frac{\partial \sigma_{n+1}(\hat{\theta}_n)}{\partial \theta} \right) \right\}^{1/2},
\] (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. An asymptotic justification for the interval on the basis of a single
sample is given in Beutner et al. (2017). Nevertheless, the interval in (3.10) often
performs rather poorly since (3.8) appears sensitive regarding the choice of bandwidth.
Bootstrap methods offer an alternative way to quantify the uncertainty around the
estimators.

4 Bootstrap

Bootstrap approximations frequently provide better insight into the actual distribu-
tion 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 errors’ 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(∞) 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 small-scale simulation study indicates that the fixed-design bootstrap performs as good as the recursive 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 \( t = 1, \ldots, n \), generate \( \eta^*_t \overset{iid}{\sim} \hat{F}_n \) and the bootstrap observation \( \epsilon^*_t = \hat{\sigma}_t(\hat{\theta}_n) \eta^*_t \), where \( \hat{\sigma}_t(\theta) \) and \( \hat{\theta}_n \) are given in (3.2) and (3.3), respectively.

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

\[
L^*_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \ell^*_t(\theta) \quad \text{and} \quad \ell^*_t(\theta) = -\frac{1}{2} \left( \frac{\epsilon^*_t}{\hat{\sigma}_t(\theta)} \right)^2 - \log \hat{\sigma}_t(\theta).
\]

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

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

4. Obtain the bootstrap estimator of the conditional VaR

\[
\hat{\text{VaR}}^*_n,\alpha = -\hat{\xi}^*_{n,\alpha} \tilde{\sigma}_{n+1}(\hat{\theta}^*_n). \tag{4.3}
\]

Remark 1. 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 $\hat{\theta}_n \in \hat{\Theta}$ under Assumption 10.

Remark 2. The term ‘fixed-design’ refers to the fact that the bootstrap observations are generated using $\tilde{\sigma}_t(\hat{\theta}_n) = \sigma(\epsilon_{t-1}, \ldots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \ldots; \hat{\theta}_n)$. In contrast, a recursive-design scheme replicates the model’s dynamic structure, i.e. $\epsilon^*_t = \sigma^* \eta^* _t$ with $\sigma^* _t = \sigma(\epsilon^* _{t-1}, \ldots, \epsilon^*_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \ldots; \hat{\theta}_n)$ and $\eta^* _t \overset{iid}{\sim} \hat{F}_n$, 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 3. 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

$$\hat{\theta}^*_{NR} = \hat{\theta}_n - \hat{j}_n^{-1} \frac{1}{2n} \sum_{t=1}^{n} \hat{D}_t(\eta^* t^2 - 1),$$

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. "$p^* \to$" and "$d^* \to$", as well as the standard bootstrap stochastic order symbol “$o_p^*(1)$” (c.f. 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, i.e. \( \frac{\partial L^*_n(\hat{\theta}^*_n)}{\partial \theta} = 0 \). A Taylor expansion around \( \hat{\theta}_n \) yields

\[
0 = \sqrt{n} \frac{\partial L^*_n(\hat{\theta}^*_n)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{\partial}{\partial \theta} \ell^*_t(\hat{\theta}_n) + \left( \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta'} \ell^*_t(\hat{\theta}) \right) \sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n)
\]

with \( \hat{\theta} \) between \( \hat{\theta}^*_n \) and \( \hat{\theta}_n \). Lemma 6 establishes \( \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta'} \ell^*_t(\hat{\theta}) \overset{p}{\to} -2J \) almost surely. Since \( \frac{\partial}{\partial \theta} \ell^*_t(\theta) = D_t(\theta)(\frac{\sigma^2(t)}{\sigma^2(\theta)} - 1) \), the first term on the right hand side reduces to \( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} D_t(\eta_t^2 - 1) \). Hence, we obtain

\[
\sqrt{n}(\hat{\theta}^*_n - \hat{\theta}_n) = - \frac{1}{2} J^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} D_t(\eta_t^2 - 1) + o_p(1)
\]

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

**Proposition 1.** Suppose Assumptions 1–4, 5(i), 5(iii) and 6–10 hold with \( a = \pm 12 \), \( b = 12 \), \( c = 6 \) and \( d > 2 \). Then, we have

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

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 \( \hat{F}_n \). To overcome this issue we rely on arguments employed by 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

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

Employing the identity of Koenker and Xiao (2006, Eq. (A.3)) we obtain

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

with

$$X_n^* = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha \right),$$

$$Y_n^* = \sum_{t=1}^{n} (\eta_t^* - \hat{\eta}_t^*) \left( \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha \right),$$

$$I_n^*(z) = \sum_{t=1}^{n} \int_{0}^{z} \left( \mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) ds,$$

$$J_n^*(z) = \sum_{t=1}^{n} \int_{\frac{z - \eta_t^*}{\sqrt{n}}}^{\frac{\eta_t^*}{\sqrt{n}}} \left( \mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) ds.$$

Subsequently, we look at each term in turn while resorting to Lemmas 7 to 10 in Appendix A.2. Lemma 7 yields $X_n^* \xrightarrow{d} N(0, \alpha(1 - \alpha))$ in probability. 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.\footnote{Note that the identity holds not only for $u \neq 0$ but also for $u = 0$.}
into \( J_n^*(z) = J_{n,1}^*(z) + J_{n,2}^*(z) \) with

\[
J_{n,1}^*(z) = \sum_{t=1}^{n} \int_{0}^{\eta_t^* - \hat{\eta}_t^*} \left( \mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \}} - \mathbb{1}_{\{\eta_t^* - \hat{\xi}_{n,\alpha} - z/\sqrt{n} < 0 \}} \right) ds \\
J_{n,2}^*(z) = \sum_{t=1}^{n} \left( \eta_t^* - \hat{\eta}_t^* \right) \left( \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha} \}} \right).
\] (4.6) (4.7)

Deviating from the proof of \textcite{Francq2015}, 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 in Appendix A.2 for more details on the difference of the proofs. Further, the second term is equal to \( J_{n,2}^*(z) = z \xi_\alpha f(\xi_\alpha) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) + o_p(1) \) in probability by Lemma 10. By the preceding discussion we obtain

\[
Q_n^*(z) = \frac{z^2}{2} f(\xi_\alpha) + z \left( X_n^* + \xi_\alpha f(\xi_\alpha) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) \right) + J_{n,1}^*(z) + Y_n^* + o_p(1)
\]
in probability. Employing \textcite{Xiong2008} Theorem 3.3 and the basic corollary of \textcite{HjortPollard2011}, we obtain

\[
\sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*) = \xi_\alpha \Omega' \sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n) + \frac{1}{f(\xi_\alpha)} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} \}} - \alpha) + o_p(1)
\]
in probability. Together with (4.4) we have

\[
\left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \frac{1}{f(\xi_\alpha)} \right) \left( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} \}} - \alpha \right) + o_p(1).
\]

\[\text{Matching notation, we take } A_n(z) = Q_n^*(z), \text{ which is convex, and set } B_n(z) = \frac{z^2}{2} V + z U_n + C_n, \text{ where } V = f(\xi_\alpha), \text{ } U_n = X_n^* + \xi_\alpha f(\xi_\alpha) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) \text{ and } C_n + r_n(z) = J_{n,1}^*(z) + Y_n^* + o_p(1) \text{ with } r_n(z) \xrightarrow{p} 0 \text{ for each } z \in \mathbb{R}. \text{ The minimizers of } A_n(z) \text{ and } B_n(z) \text{ are } \alpha_n = \sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*) \text{ and } \beta_n = -V^{-1} U_n, \text{ respectively. The basic corollary of } \textcite{HjortPollard2011} \text{ states } \alpha_n - \beta_n = o_p(1), \text{ which implies } \alpha_n - \beta_n = o_p(1) \text{ in probability } \textcite{Xiong2008} \text{ Thm. 3.3}.\]
Employing Lemma 7 leads to the paper’s main result.

**Theorem 3. (Bootstrap consistency)** Suppose Assumptions 1–10 hold with \( a = \pm 12, b = 12, c = 6 \) and \( d > 2 \). Then, we have

\[
\left( \frac{\sqrt{n}(\hat{\theta}_n^* - \hat{\theta}_n)}{\sqrt{n}(\hat{\xi}_{n,\alpha} - \hat{\xi}_{n,\alpha}^*)} \right) \xrightarrow{d^*} N(0, \Sigma_{\alpha}) \tag{4.8}
\]

**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

\[
\sqrt{n}(\tilde{\text{VaR}}_{n,\alpha}^* - \tilde{\text{VaR}}_{n,\alpha}) \quad \text{given } F_n \text{ and } (3.9) \quad \text{given } F_n \text{ merge in probability.}
\]

Its proof is along the lines of Beutner et al. (2017, proof of Theorem 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. \( \tilde{\text{VaR}}_{n,\alpha}^{(b)} \) for \( b = 1, \ldots, B \), by repeating Algorithm 1.
2.1. Obtain the *equal-tailed percentile* (EP) interval

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

(4.9)

with \( \hat{G}^{*\text{E}}_{n,B} (x) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1} \left\{ \sqrt{n} \left( \hat{VaR}^{* (b)}_{n,\alpha} - \hat{VaR}_{n,\alpha} \right) \leq x \} \).

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

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

(4.10)

2.3. Compute the *symmetric* (SY) interval

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

(4.11)

with \( \hat{H}^{*\text{S}}_{n,B} (x) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1} \left\{ \sqrt{n} \left( \hat{VaR}^{* (b)}_{n,\alpha} - \hat{VaR}_{n,\alpha} \right) \leq x \} \).

The interval in \( (4.9) \) is obtained by the EP method, that is frequently encountered in the bootstrap literature. “Flipping around” its tails leads to the RT interval given in \( (4.10) \), which can be motivated by the results of Falk and Kaufmann (1991).\(^4\) Clearly, the RT and the EP have equal length. Whereas \( (4.10) \) 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 \( (4.10) \) simplify to the \( \gamma/2 \) and \( 1 - \gamma/2 \) quantiles of \( \frac{1}{B} \sum_{b=1}^{B} \mathbb{1} \left\{ \hat{VaR}^{* (b)}_{n,\alpha} \leq x \} \), respectively. A RT type bootstrap interval for the \( VaR \) is also constructed in reduced form by Christoffersen and Gonçalves (2005). Last, the interval in \( (4.11) \) presumes symmetry for rationalizing its construction and completes the trinity of intervals.

\(^4\)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).
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) high persistence: $\theta_0 = (0.05 \times 20^2/252, 0.15, 0.8)'$;

(ii) low persistence: $\theta_0 = (0.05 \times 20^2/252, 0.4, 0.55)'$,

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) high persistence: $\theta_0 = (0.05 \times 20/\sqrt{252}, 0.05, 0.10, 0.8)'$;

(iv) low persistence: $\theta_0 = (0.05 \times 20/\sqrt{252}, 0.1, 0.3, 0.55)'$.

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.\(^5\) We consider four estimation sample sizes, $n \in \{500, 1000, 5000, 10000\}$, whereas the number of bootstrap replicates is fixed and equal to $B = 2000$. For each model version we simulate $S = 2000$ independent Monte Carlo trajectories.

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

\(^5\)The Student-$t$ innovations are appropriately standardized to satisfy $E\eta_t^2 = 1$. 

21
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).

out employing the build-in function \texttt{fmincon} and running time is reduced by parallel computing using \texttt{parfor}.

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). In the GARCH(1, 1) high persistence case (Panel I, right), we see that average coverage varies around 90% across all sample sizes for the RT and the SY interval. In contrast, the EP interval falls short of the nominal 90% by 5.75 percentage points (pp) for small sample size ($n = 500$). Nevertheless, its average coverage approaches the nominal value as the sample size increases. Remarkably, for all three 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 small ($n = 500$). 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 (Panel I, 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. This is intuitive as the conditional volatility tends to vary less in the low persistent case. Regarding the T-GARCH(1, 1) in Panel II, the overall picture is similar as in the GARCH case, however the under-coverage of the EP interval in small samples appears to be more extreme.

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 $\eta_i$’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 interval exhibits under-coverage in samples of modest size. Although falling short of the nominal size, the EP intervals exhibit a tendency of improved average coverage, e.g. 83.50% in the high-persistent T-GARCH case with $n = 500$ compared to 80.45% in Table 1. 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 interval performs 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 $n \geq 5,000$ while a small 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 three 95%−bootstrap 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 interval falls short of the nominal coverage value, yet the discrepancy appears 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% = 4.75pp compared to 90% − 84.25% = 5.75pp (see Table 1).

The question arises why the EP interval performs worse than the other two interval types in small samples, which seems counter-intuitive at first. Howbeit the results are in line with the theoretical findings of [Falk and Kaufmann (1991, Corollary)]. 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\(^6\)

(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\(^7\).  

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 of the baseline with \(n = 500\), the average coverage gap amounts to 91.50% − 84.50% = 7.00pp (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, Corollary)] 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.

\(^6\)We neglect their \(o(n^{-1/2})\) term.

\(^7\)Clearly, the theoretical results of [Falk and Kaufmann (1991)] are not directly applicable in our setting due to GARCH-type effects. Nevertheless, a link seems to be apparent.
With regard to Remark 2 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 recursive-design bootstrap scheme operates slower, 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 counterpart takes about 2.9 seconds per simulation.

In summary, the simulations suggest that for both bootstrap designs the RT and the SY bootstrap interval work well in terms of average coverage even though their tails are unequally represented for smaller sample sizes. 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 (4.10).
Table 1: The table reports distinct features of the fixed-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 intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) Student-t innovations (6 degrees of freedom), whereas in Panel II the DGP is a Student-t T-GARCH(1,1).
| Sample size | Average coverage | Av. coverage below/above | Average length | Average coverage | Av. coverage below/above | Average length |
|-------------|------------------|-------------------------|---------------|------------------|-------------------------|---------------|
|             |                  | low persistence         |               | high persistence  |                         |               |
| 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         |
| 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         |
| 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         |
| 10,000 EP   | 88.85            | 5.65/5.50               | 0.086         | 89.25            | 5.30/5.45               | 0.105         |
| RT          | 90.50            | 4.55/4.95               | 0.086         | 90.30            | 4.70/5.00               | 0.105         |
| SY          | 90.40            | 4.85/4.75               | 0.087         | 90.10            | 4.95/4.95               | 0.106         |

Panel II: T-GARCH(1,1)

| Sample size | Average coverage | Av. coverage below/above | Average length | Average coverage | Av. coverage below/above | Average length |
|-------------|------------------|-------------------------|---------------|------------------|-------------------------|---------------|
|             |                  | low persistence         |               | high persistence  |                         |               |
| 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         |
| 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         |
| 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         |
| 10,000 EP   | 89.70            | 4.70/5.60               | 0.020         | 88.60            | 5.25/6.15               | 0.041         |
| RT          | 90.60            | 4.40/5.00               | 0.020         | 90.50            | 4.45/5.05               | 0.041         |
| SY          | 90.95            | 4.15/4.90               | 0.020         | 90.25            | 4.65/5.10               | 0.041         |

Table 2: The table reports distinct features of the fixed-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 intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with Gaussian innovations, whereas in Panel II the DGP is a Gaussian T-GARCH(1,1).
Table 3: The table reports distinct features of the fixed-design bootstrap confidence intervals 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 intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) Student-t innovations (6 degrees of freedom), whereas in Panel II the DGP is a Student-t T-GARCH(1,1).

| Sample size | Average coverage | Av. coverage below/above | Average length | Average coverage | Av. coverage below/above | Average length |
|-------------|-----------------|--------------------------|---------------|-----------------|--------------------------|---------------|
| low persistence | high persistence | low persistence | high persistence |
| 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 |
| 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 |
| 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 |
| 10,000 EP | 86.25 | 6.30/7.45 | 0.218 | 87.00 | 5.75/7.25 | 0.289 |
| RT | 90.45 | 4.80/4.75 | 0.218 | 90.20 | 4.75/5.05 | 0.289 |
| SY | 89.35 | 5.30/5.35 | 0.221 | 89.45 | 4.90/5.65 | 0.292 |

Panel II: T-GARCH(1,1)

| Sample size | Average coverage | Av. coverage below/above | Average length | Average coverage | Av. coverage below/above | Average length |
|-------------|-----------------|--------------------------|---------------|-----------------|--------------------------|---------------|
| low persistence | high persistence | low persistence | high persistence |
| 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 |
| 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 |
| 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 |
| 10,000 EP | 87.50 | 5.25/7.25 | 0.054 | 86.50 | 6.10/7.40 | 0.115 |
| RT | 91.25 | 4.00/4.75 | 0.054 | 90.90 | 4.50/4.60 | 0.112 |
| SY | 90.25 | 4.05/5.70 | 0.055 | 89.85 | 4.65/5.50 | 0.114 |
| Sample size |  | Average coverage | Av. coverage below/above | Average length | Average coverage | Av. coverage below/above | Average length |
|-------------|-----------------|-----------------|-----------------|----------------|-----------------|-----------------|----------------|
| 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 |
| 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 |
| 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 |
| 10,000 | EP | 93.05 | 3.00/3.95 | 0.122 | 93.20 | 2.90/3.90 | 0.161 |
| RT | 94.70 | 2.65/2.65 | 0.122 | 94.70 | 2.45/2.85 | 0.161 |
| SY | 94.45 | 2.65/2.90 | 0.123 | 94.10 | 2.35/3.55 | 0.162 |

Panel I: GARCH(1, 1)

| Sample size |  | Average coverage | Av. coverage below/above | Average length | Average coverage | Av. coverage below/above | Average length |
|-------------|-----------------|-----------------|-----------------|----------------|-----------------|-----------------|----------------|
| 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 |
| 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 |
| 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 |
| 10,000 | EP | 93.90 | 2.30/3.80 | 0.031 | 93.25 | 2.60/3.85 | 0.064 |
| RT | 95.10 | 2.50/2.40 | 0.031 | 94.95 | 2.35/2.70 | 0.064 |
| SY | 94.95 | 2.25/2.80 | 0.031 | 94.70 | 2.35/2.95 | 0.064 |

Panel II: T-GARCH(1, 1)

Table 4: The table reports distinct features of the fixed-design bootstrap confidence intervals 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 intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) Student-t innovations (6 degrees of freedom), whereas in Panel II the DGP is a Student-t T-GARCH(1,1).
| Sample size | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
|              | low persistence | high persistence |
| 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 |
| 10,000      | 1.80 | 1.65 | 4.20  | 1.65 | 1.40 | 1.05 | 3.20  | 1.50 |

Panel II: T-GARCH(1, 1)

| Sample size | (1) | (2) | (3) | (4) | (1) | (2) | (3) | (4) |
|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
|              | low persistence | high persistence |  
| 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 |
| 10,000      | 1.90 | 0.90 | 3.75  | 1.20 | 1.60 | 1.90 | 4.40  | 1.70 |

Table 5: The table 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
The table 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 intervals are based on $B = 2,000$ bootstrap replications and the averages are computed using $S = 2,000$ simulations. Panel I presents the results for the low and high persistence parametrization of a GARCH(1,1) with (normalized) **Student-t innovations (6 degrees of freedom)**, whereas in Panel II the DGP is a Student-t T-GARCH(1,1).

| Sample size | Average coverage | Av. coverage below/above | Average length | Average coverage | Av. coverage below/above | Average length |
|-------------|------------------|--------------------------|----------------|------------------|--------------------------|----------------|
|             | low persistence  |                          |                | high persistence  |                          |                |
| 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          |
| 10,000      | EP 87.95         | 5.45/6.60                | 0.103          | 89.00            | 4.80/6.20                | 0.135          |
|             | RT 89.70         | 5.00/5.30                | 0.103          | 89.30            | 5.35/5.35                | 0.135          |
|             | SY 89.50         | 4.95/5.55                | 0.103          | 89.15            | 5.10/5.75                | 0.136          |

**Table 6:**
5.2 Empirical Application

We analyze the French stock market index CAC 40 for the period January 1, 1998 – July 1, 2018. 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 of the year 2018, which we leave for the out-of-sample evaluation, yielding $n = 5,100$ remaining observations (i.e. Jan. 1, 1998 - Dec. 31, 2017). For the volatility process we consider the T-GARCH(1,1) model specified in Example 2. 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 small and insignificant. Further, we

---

Figure 2: The returns of the French stock market index CAC 40 are plotted in (a) for the period January 1, 1998 – July 1, 2018. The histogram of the residuals is plotted in (b) after fitting a T-GARCH(1,1) model to the subperiod January 1, 1998 – December 31, 2017. A scaled normal density is superimposed.

---

8 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 $\delta$ of 1.03 in favor for the T-GARCH specification.
observe that $\hat{\alpha}_n^- >> \hat{\alpha}_n^+$ 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)

| point estimate | $\hat{\omega}_n$ | $\hat{\alpha}_n^+$ | $\hat{\alpha}_n^-$ | $\beta_n$ |
|----------------|-------------------|---------------------|---------------------|-----------|
| std. error     | 0.0039            | 0.0099              | 0.0112              | 0.0084    |

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.

plots the histogram of the residuals with the normal distribution superimposed. We find that a (normalized) Student-t distribution with 10 degrees of freedom provides an improved fit, for which Assumption 3(iii) with $d > 2$ is met.

Next, we perform a rolling window analysis starting with subperiod January 1, 1998 – December 31, 2017 and ending with subperiod July 8, 1998 – June 30, 2018. We have 125 subperiods each consisting of $n = 5,100$ 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 (fixed-design RT and recursive-design RT) and asymptotic normality, where the latter is given in (3.10). For example, for the first window the T-GARCH(1, 1) estimates are reported in Table 7 and the conditional 5%-VaR of the one-period ahead (i.e. January 2, 2018) is estimated by 1.48. The corresponding intervals are [1.39, 1.58] (fixed design), [1.38, 1.57] (recursive design) and [1.41, 1.55] (asymp. normality). Whereas both bootstrap methods yield RT intervals of (approx.) equal length, it is striking that the interval based on the asymptotic approximation is considerably shorter. 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 fixed-design bootstrap interval. We observe that in more turbulent times (e.g. February, 2018), 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 on 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 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. These extensions are 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 $\hat{H}_t(\hat{\theta}_n)$ with $\hat{H}_t(\hat{\theta}_n) = \frac{1}{\hat{\sigma}_t(\hat{\theta}_n)} \frac{\partial^2 \hat{\sigma}_t(\hat{\theta}_n)}{\partial \hat{\theta} \partial \hat{\theta}}$. Further, we introduce

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

**Lemma 1.** Suppose Assumptions 1, 2, 3, 4(i), 5(i), 6, 8 and 9(i) hold with $a = -1$. Then, we have $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \overset{a.s.}{\to} 0$.

**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

$$\limsup_{n \to \infty} \sup_{||\theta - \theta_0|| \leq \tau} \left| \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\{\eta_t \leq x\hat{\sigma}_t(\theta)/\sigma_t(\theta_0)\} - F(x) \right| \leq 2 \left( F(x + \varepsilon|x|) - F(x - \varepsilon|x|) \right) \tag{A.2}$$

almost surely for any $x \in \mathbb{R}$. In the second step, we show $\hat{F}_n(x) \overset{a.s.}{\to} F(x)$ for any $x \in \mathbb{R}$ using (A.2) and thereafter prove $\sup_{x \in \mathbb{R}} |\hat{F}_n(x) - F(x)| \overset{a.s.}{\to} 0$.

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

$$\frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\{\eta_t \leq x\hat{\sigma}_t(\theta)/\sigma_t(\theta_0)\} \leq \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\{\eta_t \leq x\sigma_t(\theta)/\sigma_t(\theta_0) + |x|C_1 \rho^t / \sigma_t(\theta_0)\}$$

$$\leq \frac{n_0}{n} + \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}\{\eta_t \leq x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|x|\}$$

37
holds almost surely. Let $\tau > 0$ (to be specified later) and define $\theta_{\tau} = \theta_0 + \text{sign}(x)\tau t$ with $t = (1, \ldots, 1)' \in \mathbb{R}^r$. Since $\theta_0 \in \hat{\Theta}$ by Assumption 6, one can choose $\tau$ sufficiently small such that $\theta_{\tau} \in \Theta$. By Assumption 8, we get for any $\theta$ satisfying $||\theta - \theta_0|| \leq \tau$

\[
\frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta)/\sigma_t(\theta_0) + \varepsilon|\theta|\}} \leq \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta_{\tau})/\sigma_t(\theta_0) + \varepsilon|\theta|\}}
\]

almost surely. The ergodic theorem for strictly stationary sequences (c.f. [Francq and Zakoian, 2011] Theorem A.2), henceforth called the ergodic theorem, and Assumptions 2, 3 and 5(i) yield

\[
\frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta_{\tau})/\sigma_t(\theta_0) + \varepsilon|\theta|\}} \xrightarrow{a.s.} \mathbb{E}\mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta_{\tau})/\sigma_t(\theta_0) + \varepsilon|\theta|\}} = \mathbb{E}F\left(x\sigma_t(\theta_{\tau})/\sigma_t(\theta_0) + \varepsilon|\theta|\right).
\]

Further, Assumptions 3 and 9(i) with $a = -1$ imply $\lim_{\tau \to 0} \sigma_t(\theta_{\tau})/\sigma_t(\theta_0) = 1$ almost surely. Thus, the dominated convergence theorem entails

\[
\lim_{\tau \to 0} \mathbb{E}F\left(x\sigma_t(\theta_{\tau})/\sigma_t(\theta_0) + \varepsilon|\theta|\right) = F(x + \varepsilon|x|).
\]

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

\[
\limsup_{n \to \infty} \sup_{||\theta - \theta_0|| \leq \tau} \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta)/\sigma_t(\theta_0)\}} \leq F(x) + 2\left(F(x + \varepsilon|x|) - F(x)\right)
\]

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

\[
\liminf_{n \to \infty} \sup_{||\theta - \theta_0|| \leq \tau} \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{\eta_t \leq x\sigma_t(\theta)/\sigma_t(\theta_0)\}} \geq F(x) - 2\left(F(x) - F(x - \varepsilon|x|)\right).
\]
almost surely for any \(x \in \mathbb{R}\). Combining both results, we establish (A.2).

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

\[
\sup_{||\theta - \theta_0|| \leq \tau} \left| \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{n_t \leq x\hat{\sigma}_t(\theta)/\sigma_t(\theta_0)\}} - F(x) \right| < \zeta
\]

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

\[
\left| \hat{F}_n(x) - F(x) \right| \leq \sup_{||\theta - \theta_0|| \leq \tau} \left| \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{n_t \leq x\hat{\sigma}_t(\theta)/\sigma_t(\theta_0)\}} - F(x) \right| < \zeta
\]

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

**Lemma 2.** Suppose Assumptions \(\mathbb{1} - \mathbb{5}\) and \(\mathbb{9}\) hold.

(i) If in addition Assumptions \(\mathbb{9}(i)\) and \(\mathbb{9}(ii)\) hold with \(b = 1\), then \(\hat{\Omega}_n \xrightarrow{a.s.} \Omega\).

(ii) If in addition Assumptions \(\mathbb{9}(i)\) and \(\mathbb{9}(ii)\) hold with \(b = 2\), then \(\hat{J}_n \xrightarrow{a.s.} J\).

(iii) If in addition Assumptions \(\mathbb{9}(i)\) and \(\mathbb{9}(iii)\) 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 \(\mathbb{9}(iii)\) and \(\mathbb{9}(i)\) hold with \(a = 2d\) and \(d > 2\), then \(\frac{1}{n} \sum_{t=1}^{n} |\hat{\eta}_t|^m \xrightarrow{a.s.} \mathbb{E}[|\eta|^m]\) for \(1 \leq m \leq 2d\).

(v) If in addition Assumptions \(\mathbb{9}(iii)\), \(\mathbb{6}\), \(\mathbb{7}\) and \(\mathbb{9}(i)\) hold with \(a = -1, 4\) and \(d = 2\), then \(\frac{1}{n} \sum_{t=1}^{n} \hat{\eta}_t^2 \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n, \alpha}\}} \xrightarrow{a.s.} \mathbb{E}[\eta_2 \mathbb{1}_{\{\eta < \xi_{\alpha}\}}]\).
Proof. Consider the first statement and expand

\[ \frac{1}{n} \sum_{t=1}^{n} D_t = \frac{1}{n} \sum_{t=1}^{n} D_t(\hat{\theta}_n) + \frac{1}{n} \sum_{t=1}^{n} \left( \tilde{D}_t(\hat{\theta}_n) - D_t(\hat{\theta}_n) \right). \]

Focusing on I, we take \( \varepsilon > 0 \) and let \( e_1, \ldots, e_r \) 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 \( i = 1, \ldots, r \). 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

\[ \frac{1}{n} \sum_{t=1}^{n} e'_i D_t(\hat{\theta}_n) \leq \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \xrightarrow{\text{a.s.}} \mathbb{E}\left[ \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] < \mathbb{E}[e'_i D_t] + \varepsilon \]

\[ \frac{1}{n} \sum_{t=1}^{n} e'_i D_t(\hat{\theta}_n) \geq \frac{1}{n} \sum_{t=1}^{n} \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \xrightarrow{\text{a.s.}} \mathbb{E}\left[ \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} e'_i D_t(\theta) \right] > \mathbb{E}[e'_i D_t] - \varepsilon \]

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

Regarding II, we note that for each \( \theta \in \Theta \), Assumption 3 implies

\[ \left\| \tilde{D}_t(\theta) - D_t(\theta) \right\| = \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} - \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \]

\[ = \left\| \frac{1}{\tilde{\sigma}_t(\theta)} \left( \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} - \frac{\partial \sigma_t(\theta)}{\partial \theta} \right) + \frac{\sigma_t(\theta) - \tilde{\sigma}_t(\theta)}{\tilde{\sigma}_t(\theta)} \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \]

\[ \leq \frac{1}{\tilde{\sigma}_t(\theta)} \left\| \frac{\partial \tilde{\sigma}_t(\theta)}{\partial \theta} - \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| + \left\| \frac{\sigma_t(\theta) - \tilde{\sigma}_t(\theta)}{\tilde{\sigma}_t(\theta)} \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \right\| \]

\[ \leq C_1 \rho \left\| \frac{1}{\omega} \right\| ||D_t(\theta)|| = \frac{C_1 \rho}{\omega} \left( 1 + ||D_t(\theta)|| \right). \]

40
We obtain
\[
\|II\| \leq \frac{1}{n} \sum_{t=1}^{n} \| \tilde{D}_t(\hat{\theta}_n) - D_t(\hat{\theta}_n) \| \leq \frac{C_1}{\omega} \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + \| D_t(\hat{\theta}_n) \| \right) \overset{a.s.}{\leq} \frac{C_1}{\omega} \frac{1}{n} \sum_{t=1}^{n} \rho^t (1 + U_t).
\]

For each \( \varepsilon > 0 \), Markov’s inequality entails
\[
\sum_{t=1}^{\infty} \mathbb{P} \left[ \rho^t (1 + U_t) > \varepsilon \right] \leq \sum_{t=1}^{\infty} \rho^t \frac{1 + \mathbb{E}[U_t]}{\varepsilon} = \frac{1 + \mathbb{E}[U_t]}{\varepsilon (1 - \rho)} < \infty
\]

since \( \rho \in (0, 1) \) and \( \mathbb{E}[U_t] < \infty \) by Assumption 9(ii). The Borel-Cantelli lemma implies
\[
0 = \mathbb{P} \left[ \lim_{t \to \infty} \bigcup_{s=t}^{\infty} \{ \rho^s (1 + U_s) > \varepsilon \} \right] \geq \mathbb{P} \left[ \lim_{t \to \infty} \rho^t (1 + U_t) > \varepsilon \right] \quad (A.4)
\]

and hence \( \rho^t (1 + U_t) \to 0 \) almost surely. Cesáro’s lemma yields \( \frac{1}{n} \sum_{t=1}^{n} \rho^t (1 + U_t) \overset{a.s.}{\to} 0 \) and hence \( \|II\| \overset{a.s.}{\to} 0 \), which validates the first statement.

Consider the second statement and expand
\[
\frac{1}{n} \sum_{t=1}^{n} \dot{D}_t \dot{D}'_t = \frac{1}{n} \sum_{t=1}^{n} D_t(\hat{\theta}_n)D'_t(\hat{\theta}_n) + \frac{1}{n} \sum_{t=1}^{n} \left( \tilde{D}_t(\hat{\theta}_n)\tilde{D}'_t(\hat{\theta}_n) - D_t(\hat{\theta}_n)D'_t(\hat{\theta}_n) \right).
\]

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

for all \( i, j = 1, \ldots, r \). Since \( \hat{\theta}_n \overset{a.s.}{\to} \theta_0 \) by Theorem 1, we have \( \hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0) \) almost surely.
Together with the uniform ergodic theorem we obtain

\[
\frac{1}{n} \sum_{t=1}^{n} e_i' D_t(\hat{\theta}_n) D_t'(\hat{\theta}_n) e_j \quad \text{a.s.} \quad \leq \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \mathcal{Y}_n(\theta_0)} e_i' D_t(\theta) D_t'(\theta) e_j
\]

\[
\Rightarrow \quad \mathbb{E}\left[ \sup_{\theta \in \mathcal{Y}_n(\theta_0)} e_i' D_t(\theta) D_t'(\theta) e_j \right] \leq \mathbb{E}[e_i' D_t D_t' e_j] + \varepsilon
\]

\[
\frac{1}{n} \sum_{t=1}^{n} e_i' D_t(\hat{\theta}_n) D_t'(\hat{\theta}_n) e_j \quad \text{a.s.} \quad \geq \frac{1}{n} \sum_{t=1}^{n} \inf_{\theta \in \mathcal{Y}_n(\theta_0)} e_i' D_t(\theta) D_t'(\theta) e_j
\]

\[
\Rightarrow \quad \mathbb{E}\left[ \inf_{\theta \in \mathcal{Y}_n(\theta_0)} e_i' D_t(\theta) D_t'(\theta) e_j \right] \geq \mathbb{E}[e_i' D_t D_t' e_j] - \varepsilon
\]

Taking \( \varepsilon \downarrow 0 \) 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 \( \text{III} \xrightarrow{a.s.} \mathbb{E}[D_t D_t'] = J \). Consider \( \text{IV} \); using (A.3) and the elementary inequality

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

for all \( x, y \in \mathbb{R}^m \), we obtain for \( \theta \in \Theta \)

\[
\left\| \tilde{D}_t(\theta) \tilde{D}_t'(\theta) - D_t(\theta) D_t'(\theta) \right\|
\]

\[
\leq \left\| \tilde{D}_t(\theta) - D_t(\theta) \right\|^2 + 2\left\| \tilde{D}_t(\theta) - D_t(\theta) \right\| \left\| D_t(\theta) \right\|
\]

\[
\leq \frac{C_2^2}{\omega^2} \rho(1 + \|D_t(\theta)\|)^2 + 2\frac{C_1}{\omega} \rho(1 + \|D_t(\theta)\|) \left\| D_t(\theta) \right\|
\]

\[
\leq \frac{C_2^2}{\omega^2} \rho(1 + \|D_t(\theta)\|)^2 + 2\frac{C_1}{\omega} \rho(1 + \|D_t(\theta)\|)^2
\]

\[
= \left( \frac{C_2^2}{\omega^2} + \frac{2C_1}{\omega} \right) \rho(1 + \|D_t(\theta)\|)^2.
\]

Hence, we get

\[
\|I V\| \leq \frac{1}{n} \sum_{t=1}^{n} \left\| \tilde{D}_t(\hat{\theta}_n) \tilde{D}_t'(\hat{\theta}_n) - D_t(\hat{\theta}_n) D_t'(\hat{\theta}_n) \right\| \leq \left( \frac{C_2^2}{\omega^2} + \frac{2C_1}{\omega} \right) \frac{1}{n} \sum_{t=1}^{n} \rho(1 + \|D_t(\hat{\theta}_n)\|)^2
\]

\[
\xRightarrow{\text{a.s.}} \left( \frac{C_2^2}{\omega^2} + \frac{2C_1}{\omega} \right) \frac{1}{n} \sum_{t=1}^{n} \rho(1 + U_t)^2.
\]
For each $\varepsilon > 0$, Markov’s inequality yields
\[
\sum_{t=1}^{\infty} \mathbb{P}\left[ \rho^t (1 + U_t)^2 > \varepsilon \right] \leq \sum_{t=1}^{\infty} \frac{\rho^{t/2} (1 + \mathbb{E}[U_t])}{\sqrt{\varepsilon}} = \frac{1 + \mathbb{E}[U_t]}{\sqrt{\varepsilon}(1 - \sqrt{\rho})} < \infty
\]
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 Cesáro’s lemma. Hence, $\|IV\| \xrightarrow{a.s.} 0$, which validates the second statement.

Consider the third statement and expand
\[
\frac{1}{n} \sum_{t=1}^{n} \hat{H}_t = \frac{1}{n} \sum_{t=1}^{n} H_t(\hat{\theta}_n) + \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\hat{H}_t(\hat{\theta}_n) - H_t(\hat{\theta}_n)}{V} \right) \xrightarrow{a.s.} E\left[ e_i H_t e_j \right] + \varepsilon
\]
for all $i,j \in \{1, \ldots, r\}$. Since $\hat{\theta}_n \xrightarrow{a.s.} \theta_0$ by Theorem [1] we have $\hat{\theta}_n \in V(\theta_0)$ almost surely. Together with the uniform ergodic theorem we obtain
\[
\frac{1}{n} \sum_{t=1}^{n} e_i H_t(\hat{\theta}_n) e_j \xrightarrow{a.s.} \mathbb{E}\left[ e_i H_t(\theta) e_j \right] \xrightarrow{a.s.} \mathbb{E}\left[ \sup_{\theta \in V(\theta_0)} e_i H_t(\theta) e_j \right] < \mathbb{E}\left[ e_i H_t e_j \right] + \varepsilon
\]
for all $i,j \in \{1, \ldots, r\}$. 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

\[
\| \tilde{H}_t(\theta) - H_t(\theta) \| = \left\| \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \tilde{\sigma}_t(\theta)}{\partial \theta \partial \theta'} - \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\|
\leq \frac{1}{\sigma_t(\theta)} \left\| \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} - \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'} \right\| + \frac{\sigma_t(\theta) - \tilde{\sigma}_t(\theta)}{\sigma_t(\theta)} \frac{1}{\sigma_t(\theta)} \frac{\partial^2 \sigma_t(\theta)}{\partial \theta \partial \theta'}
\leq \frac{C_1 \rho^t}{\omega} + \frac{C_1 \rho^t}{\omega} \| H_t(\theta) \| = \frac{C_1 \rho^t}{\omega} \left( 1 + \| H_t(\theta) \| \right)
\]

for each \( \theta \in \Theta \). We obtain

\[
\| VI \| \leq \frac{1}{n} \sum_{t=1}^{\infty} \| \tilde{H}_t(\hat{\theta}_n) - H_t(\hat{\theta}_n) \| \leq \frac{C_1}{\omega} \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + \| H_t(\hat{\theta}_n) \| \right) \xrightarrow{a.s.} \frac{C_1}{\omega} \frac{1}{n} \sum_{t=1}^{n} \rho^t (1 + V_t).
\]

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 Cesáro’s lemma. Hence, \( \| VI \| \xrightarrow{a.s.} 0 \), which validates the third statement.

Consider the fourth statement and let \( 1 \leq m \leq 2d \). We expand and obtain

\[
\frac{1}{n} \sum_{t=1}^{n} |\hat{\eta}_t|^m = \frac{1}{n} \sum_{t=1}^{n} \left( \frac{|\epsilon_t|}{\sigma_t(\hat{\theta}_n)} \right)^m + \frac{1}{n} \sum_{t=1}^{n} \left( \left( \frac{|\epsilon_t|}{\hat{\sigma}_t(\hat{\theta}_n)} \right)^m - \left( \frac{|\epsilon_t|}{\sigma_t(\hat{\theta}_n)} \right)^m \right).
\]

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

\[
1 - \varepsilon \leq \mathbb{E} \left[ \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \right] \leq \mathbb{E} \left[ \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} \left| \frac{\sigma_t(\theta_0)}{\sigma_t(\theta)} \right|^m \right] < 1 + \varepsilon.
\]
Since $\hat{\theta}_n \overset{a.s.}{\rightarrow} \theta_0$ (Theorem 1), we have $\hat{\theta}_n \in \mathcal{Y}_\varepsilon(\theta_0)$ almost surely. Together with the uniform ergodic theorem we obtain

$$VII \overset{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} \left( \frac{|\varepsilon_t|}{\sigma_t(\theta)} \right)^m \overset{a.s.}{\rightarrow} \mathbb{E} \left[ \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} \left( \frac{|\varepsilon_t|}{\sigma_t(\theta)} \right)^m \right]$$

$$= \mathbb{E}[|\eta|^m] \mathbb{E} \left[ \sup_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} \left( \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \right)^m \right] < \mathbb{E}[|\eta|^m] (1 + \varepsilon)$$

$$VII \overset{a.s.}{\geq} \frac{1}{n} \sum_{t=1}^{n} \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} \left( \frac{|\varepsilon_t|}{\sigma_t(\theta)} \right)^m \overset{a.s.}{\rightarrow} \mathbb{E} \left[ \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} \left( \frac{|\varepsilon_t|}{\sigma_t(\theta)} \right)^m \right]$$

$$= \mathbb{E}[|\eta|^m] \mathbb{E} \left[ \inf_{\theta \in \mathcal{V}_\varepsilon(\theta_0)} \left( \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \right)^m \right] > \mathbb{E}[|\eta|^m] (1 - \varepsilon)$$

Taking $\varepsilon \searrow 0$ establishes $VII \overset{a.s.}{\rightarrow} \mathbb{E}[|\eta|^m]$. Employing the elementary inequality

$$| (x + y)^z - x^z | \leq z2^{z-1}|y|(|x|^{z-1} + |y|^{z-1})$$

for $x, y \in \mathbb{R}$ with $x \geq 0$ and $x + y \geq 0$ and $z \geq 1$, the term $VIII$ can be bounded by

$$|VIII| = \left| \frac{1}{n} \sum_{t=1}^{n} \left( \frac{|\varepsilon_t|}{\sigma_t(\hat{\theta}_n)} \right)^m \left( \left( \frac{\sigma_t(\hat{\theta}_n)}{\sigma_t(\hat{\theta}_n)} \right)^m - 1 \right) \right|$$

$$= \left| \frac{1}{n} \sum_{t=1}^{n} \left( \frac{|\varepsilon_t|}{\sigma_t(\hat{\theta}_n)} \right)^m \left( \left( 1 + \frac{\sigma_t(\hat{\theta}_n) - \sigma_t(\hat{\theta}_n)}{\sigma_t(\hat{\theta}_n)} \right)^m - 1 \right) \right|$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \left( \frac{|\varepsilon_t|}{\sigma_t(\hat{\theta}_n)} \right)^m \left( \left( 1 + \frac{C_1\rho^t}{\omega} \right)^m - 1 \right)$$

$$\leq \frac{1}{n} \sum_{t=1}^{n} \left( \frac{|\varepsilon_t|}{\sigma_t(\hat{\theta}_n)} \right)^m 2^{m-1} \frac{C_1\rho^t}{\omega} \left( 1 + \left( \frac{C_1\rho^t}{\omega} \right)^{m-1} \right)$$

$$\leq 2^{m-1} \frac{C_1}{\omega} + \frac{C_1^m}{\omega^m} \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( \frac{|\varepsilon_t|}{\sigma_t(\hat{\theta}_n)} \right)^m$$

$$\overset{a.s.}{\leq} 2^{m-1} \frac{C_1}{\omega} + \frac{C_1^m}{\omega^m} \frac{1}{n} \sum_{t=1}^{n} \rho^t |\eta|^m S_t^m.$$
where $S_t$ is defined in (A.1). For each $\varepsilon > 0$, Markov’s inequality yields

$$\sum_{t=1}^{\infty} P \left[ \rho^t |\eta_t|^m S_t^m > \varepsilon \right] \leq \sum_{t=1}^{\infty} \rho^t \frac{E[|\eta_t|^m S_t^m]}{\varepsilon} = \frac{E[|\eta_t|^m] E[S_t^m]}{\varepsilon (1 - \rho)} < \infty$$

and $\frac{1}{n} \sum_{t=1}^{n} \rho^t |\eta_t|^m S_t^m \xrightarrow{a.s.} 0$ follows from combining the Borel-Cantelli lemma with Cesáro’s lemma. Thus, $VIII \xrightarrow{a.s.} 0$, which validates the fourth statement.

Consider the fifth statement; for every $C \in \mathbb{R}$, we have

$$\left| \frac{1}{n} \sum_{t=1}^{n} \hat{\eta}_t^2 \mathbb{1}_{\{\hat{\eta}_t < \hat{\xi}_{n,\alpha}\}} - E[\eta_t^2 \mathbb{1}_{\{\eta_t < \xi_{n,\alpha}\}}] \right| = \left| \int x^2 \mathbb{1}_{\{x < \hat{\xi}_{n,\alpha}\}} d\hat{F}_n(x) - \int x^2 \mathbb{1}_{\{x < \xi_{n,\alpha}\}} dF(x) \right|$$

$$\leq \left\{ \begin{array}{ll} \int x^2 \mathbb{1}_{\{x < C\}} dF(x) + \int x^2 \mathbb{1}_{\{x < C\}} d\hat{F}_n(x) \\
X \end{array} \right\} \leq \left\{ \begin{array}{ll} \int x^2 (\mathbb{1}_{\{x < \hat{\xi}_{n,\alpha}\}} - \mathbb{1}_{\{x < C\}}) d\hat{F}_n(x) - \int x^2 (\mathbb{1}_{\{x < \xi_{n,\alpha}\}} - \mathbb{1}_{\{x < C\}}) dF(x) \\
X I \end{array} \right\}.$$

Hölder’s inequality implies

$$IX \leq \left( \int x^4 dF(x) \int \mathbb{1}_{\{x < C\}} dF(x) \right)^{\frac{1}{2}} = \sqrt{\kappa F(C)}$$

and

$$X \leq \left( \int x^4 d\hat{F}_n(x) \int \mathbb{1}_{\{x < C\}} d\hat{F}_n(x) \right)^{\frac{1}{2}} \xrightarrow{a.s.} \sqrt{\kappa F(C)}$$

by the fourth statement and Lemma \[. Choosing $C$ sufficiently small with $C < \xi_{n,\alpha}$, we can make $IX$ arbitrarily small and $X$ arbitrarily small almost surely. Regarding

46
XI we employ the partial integration formula

\[ G(b-)H(b-) - G(a-)H(a-) = \int_{(a,b)} G(t-) \, dH(t) + \int_{(a,b)} H(s) \, dG(s) \quad (A.8) \]

with \( G \) and \( H \) both right-continuous functions being locally of bounded variation and \( a \leq t < b \). Given \( C \), we obtain

\[
|XI| = \left| \int_{[C, \hat{\xi}_{n,\alpha})} x^2 \, d\hat{F}_n(x) - \int_{[C, \xi_\alpha)} x^2 \, dF(x) \right| \\
= \left| \xi_{n,\alpha}^2 \hat{F}_n(\hat{\xi}_{n,\alpha} - C) - C^2 \hat{F}_n(C) - \int_{[C, \hat{\xi}_{n,\alpha})} \hat{F}_n(x) \, dx^2 \\
- \xi_\alpha^2 F(\xi_\alpha) + C^2 F(C) + \int_{[C, \xi_\alpha)} F(x) \, dx^2 \right|\]

\[
\leq \left[ \xi_{n,\alpha}^2 \hat{F}_n(\hat{\xi}_{n,\alpha} - C) - \xi_\alpha^2 F(\xi_\alpha) \right] + C^2 \left| \hat{F}_n(C) - F(C) \right| \\
+ \left| \int_{[C, \hat{\xi}_{n,\alpha})} (\hat{F}_n(x) - F(x)) \, dx^2 \right| + \left| \int_{[\xi_\alpha, \hat{\xi}_{n,\alpha})} \hat{F}_n(x) \, dx^2 \right| \\
= \left| XI_1 \right| + \left| XI_2 \right| + \left| XI_3 \right| + \left| XI_4 \right|.
\]

We have \( XI_1 \xrightarrow{a.s.} 0 \) as \( \hat{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_\alpha \) by Theorem 1 and \( \hat{F}_n(\hat{\xi}_{n,\alpha} - C) \xrightarrow{a.s.} \frac{|na| + 1}{n} \rightarrow \alpha = F(\xi_\alpha) \). Further, we obtain \( XI_2 \leq C^2 \sup_{x \in \mathbb{R}} | \hat{F}_n(x) - F(x) | \xrightarrow{a.s.} 0 \) and \( XI_3 \leq \sup_{x \in \mathbb{R}} | \hat{F}_n(x) - F(x) | (C^2 - \xi_\alpha^2) \xrightarrow{a.s.} 0 \) by Lemma 1. Together with \( XI_4 \leq | \hat{\xi}_{n,\alpha}^2 - \xi_\alpha^2 | \xrightarrow{a.s.} 0 \) by Theorem 1, we conclude that \( XI \xrightarrow{a.s.} 0 \), which completes the proof.

**Lemma 3.** Suppose Assumptions 1–9 hold with \( a = \pm 6, b = 6 \) and \( c = d = 2 \). Then, for \( v, w \in \mathbb{R} \), we have

\[
\sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + (v + w)/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n}) \right) \xrightarrow{p} v f(\xi_\alpha).
\]
Proof. We closely follow the proofs of Berkes and Horváth (2003) and define

\[
\tilde{\gamma}_t(u) = \tilde{\sigma}_t(\theta_0 + n^{-1/2} u)/\sigma_t(\theta_0)
\]

\[
\gamma_t(u) = \sigma_t(\theta_0 + n^{-1/2} u)/\sigma_t(\theta_0)
\]

\[
\zeta_t(x, u) = \mathbb{1}_{\{\eta_t \leq x \tilde{\gamma}_t(u)\}} - F(x \tilde{\gamma}_t(u)) - (\mathbb{1}_{\{\eta_t \leq x\}} - F(x))
\]

\[
S_n(x, u) = \sum_{t=1}^{n} \zeta_t(x, u)
\]

\[
F_n(x) = \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{\eta_t \leq x\}}.
\]

Let \( A > 0 \) and take \( X = [x, \bar{x}] \) to be a compact neighborhood around \( \xi_\alpha \) with \( \bar{x} < 0 \).

We establish the result in six steps:

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

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

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

**Step 4:** \( \sup_{||u|| \leq A} \sup_{x \in X} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( F(x \tilde{\gamma}_t(u)) - F(x) - x f(x) \Omega^\top u \right) \right| = o_p(1) \);

**Step 5:** \( \sup_{x \in X} \left| \sqrt{n} \left( \hat{F}_n(x) - F_n(x) \right) - x f(x) \Omega^\top \sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \right| = o_p(1) \);

**Step 6:** \( \sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + (v + w)/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n}) \right) \xrightarrow{p} v f(\xi_\alpha) \) for all \( v, w \in \mathbb{R} \).

Let \( \mathcal{F}_t \) be the \( \sigma \)-algebra generated by \( \zeta_t, \zeta_{t-1}, \ldots \) 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

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

48
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 $\mathbb{E} \left[ \max_{1 \leq t \leq n} \zeta_t^4(x, u) \right] \leq 16$, it suffices to show that

$$
\mathbb{E} \left[ \left( \sum_{t=1}^{n} \mathbb{E}_{t-1} \left[ \zeta_t^2(x, u) \right] \right)^2 \right] = O(n). \quad (A.9)
$$

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

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

with

$$
\zeta_{t,1}(x, u) = \mathbb{I}_{\{n \leq x_{\bar{\gamma}_t(u)}\}} - F(x_{\bar{\gamma}_t(u)}) - \mathbb{I}_{\{n \leq x_{\gamma_t(u)}\}} + F(x_{\gamma_t(u)})
$$

$$
\zeta_{t,2}(x, u) = \mathbb{I}_{\{n \leq x_{\gamma_t(u)}\}} - F(x_{\gamma_t(u)}) - \mathbb{I}_{\{n \leq x\}} + F(x).
$$

The elementary inequality

$$
\left( \sum_{i=1}^{m} x_i \right)^2 \leq m \sum_{i=1}^{m} x_i^2 \quad (A.10)
$$

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

$$
\mathbb{E}_{t-1} \left[ \zeta_t^4(x, u) \right] \leq 2 \left( \mathbb{E}_{t-1} \left[ \zeta_{t,1}^2(x, u) \right] + \mathbb{E}_{t-1} \left[ \zeta_{t,2}^2(x, u) \right] \right).
$$

Moreover, the inequality $\text{Var}_{t-1} [\mathbb{I}_{\{n \leq a\}} - \mathbb{I}_{\{n \leq b\}}] \leq |F(a) - F(b)|$ for $a, b \in \mathbb{R}$ gives

$$
\mathbb{E}_{t-1} \left[ \zeta_{t,1}^2(x, u) \right] = \text{Var}_{t-1} \left[ \mathbb{I}_{\{n \leq x_{\bar{\gamma}_t(u)}\}} - \mathbb{I}_{\{n \leq x_{\gamma_t(u)}\}} \right] \leq |F(x_{\bar{\gamma}_t(u)}) - F(x_{\gamma_t(u)})|
$$

$$
\mathbb{E}_{t-1} \left[ \zeta_{t,2}^2(x, u) \right] = \text{Var}_{t-1} \left[ \mathbb{I}_{\{n \leq x_{\gamma_t(u)}\}} - \mathbb{I}_{\{n \leq x\}} \right] \leq |F(x_{\gamma_t(u)}) - F(x)|.
$$
Combining results, it follows that

\[ E_{t-1} [\zeta_t^2(x, u)] \leq 2 \left( |F(x_{\gamma_t}(u)) - F(x)| + |F(x_{\gamma_t}(u)) - F(x_{\gamma_t}(u))| \right). \quad (A.11) \]

Employing (A.11), we obtain that the left-hand side in (A.9) is bounded by

\[
E \left[ \left( \sum_{t=1}^{n} E_{t-1} [\zeta_t^2(x, u)] \right)^2 \right] \\
\leq 4E \left[ \left( \sum_{t=1}^{n} |F(x_{\gamma_t}(u)) - F(x)| + \sum_{t=1}^{n} |F(x_{\gamma_t}(u)) - F(x_{\gamma_t}(u))| \right)^2 \right] \\
\leq 8 \left( \sum_{t=1}^{n} \left[ E \left( \left| F(x_{\gamma_t}(u)) - F(x) \right|^2 \right) \right] + \sum_{t=1}^{n} \left[ E \left( \left| F(x_{\gamma_t}(u)) - F(x_{\gamma_t}(u)) \right|^2 \right) \right] \right),
\]

where the last inequality follows from applying (A.10) once more. It suffices to show that both terms are \( O(n) \). Consider \( I \); The Cauchy-Schwarz inequality yields

\[
I = \sum_{t=1}^{n} \sum_{\tau=1}^{n} E \left[ \left| F(x_{\gamma_t}(u)) - F(x) \right| \left| F(x_{\gamma_{\tau}}(u)) - F(x) \right| \right] \\
\leq \sum_{t=1}^{n} \sum_{\tau=1}^{n} \left( E \left[ \left( F(x_{\gamma_t}(u)) - F(x) \right)^2 \right] \right)^{\frac{1}{2}} \left( E \left[ \left( F(x_{\gamma_{\tau}}(u)) - F(x) \right)^2 \right] \right)^{\frac{1}{2}}. \quad (A.12)
\]

We define the event

\[ \mathcal{A}_{n,t} = \left\{ \frac{T_t U_t A}{\sqrt{n}} \leq \frac{1 - \sqrt{p}}{2} \right\}, \quad (A.13) \]
where $T_t$ and $U_t$ are defined in (A.1). The inner term of (A.12) can be bounded by

$$
\mathbb{E} \left[ \left( F(x\gamma_t(u)) - F(x) \right)^2 \right] = \mathbb{E} \left[ \left( F(x\gamma_t(u)) - F(x) \right)^2 (\mathbb{1}_{\varphi_{n,t}^c} + \mathbb{1}_{\varphi_{n,t}}) \right] 
\leq \mathbb{P}[\varphi_{n,t}^c] + \mathbb{E} \left[ (F(x\gamma_t(u)) - F(x))^2 \mathbb{1}_{\varphi_{n,t}} \right].
$$

where the superscript $c$ denotes the event’s complement. Markov’s inequality yields

$$
I_1 = \mathbb{P} \left[ T_t U_t \frac{A}{\sqrt{n}} > \frac{1 - \sqrt{\rho}}{2} \right] \leq \frac{1 (2A)^2 \mathbb{E}[T_t^2 U_t^2]}{n (1 - \sqrt{\rho})^2}
$$

and $\mathbb{E}[T_t^2 U_t^2] \leq \mathbb{E}[T_t^4]^\frac{1}{2} \mathbb{E}[U_t^4]^\frac{1}{2} < \infty$ by the Cauchy-Schwarz inequality and Assumption 9. Hence, $I_1 = O(n^{-1})$. Regarding $I_2$, the mean value theorem implies

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

with $\gamma_t$ being between $\gamma_t(u)$ and 1. We have $(x\gamma_t f(x\gamma_t))^2 \leq M^2$ by Assumption 5(iv).

Further, for $n$ sufficiently large such that $\{ \theta : ||\theta - \theta_0|| \leq A/\sqrt{n} \} \subseteq \mathcal{V}(\theta_0)$ the mean value theorem implies

$$
\sup_{||u|| \leq A} |\gamma_t(u) - 1| = \sup_{||u|| \leq A} \left| \frac{\sigma_t(\theta_0 + u/\sqrt{n}) - \sigma_t(\theta_0)}{\sigma_t(\theta_0)} \right|
= \sup_{||u|| \leq A} \left| \frac{1}{\sigma_t(\theta_0)} \partial_t(\theta) \frac{1}{\sqrt{n}} u \right| = \frac{1}{\sqrt{n}} \sup_{||u|| \leq A} \left| \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} D_t(\theta) u \right|
\leq \frac{1}{\sqrt{n}} \sup_{||\theta - \theta_0|| \leq A^{n^{-1/2}}} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \sup_{||\theta - \theta_0|| \leq A^{n^{-1/2}}} ||D_t(\theta)|| \sup_{||u|| \leq A} ||u|| \leq A \sqrt{n} T_t U_t,
$$

51
where \( \bar{\theta} \) between \( \theta_0 \) and \( \theta_0 + u/\sqrt{n} \). Hence, if

\[
\mathcal{A}_{n,t} \text{ holds } \Rightarrow \sup_{||u|| \leq A} |\gamma_t(u) - 1| \leq \frac{1 - \sqrt{\rho}}{2}.
\]

(A.18)

and we obtain

\[
\bar{\gamma}_t \geq 1 - |\gamma_t(u) - 1| \geq 1 - \frac{1 - \sqrt{\rho}}{2} = \frac{1 + \sqrt{\rho}}{2},
\]

where the first inequality follows from \( \bar{\gamma}_t \) being between \( \gamma_t(u) \) and 1. Thus, in the event of \( \mathcal{A}_{n,t} \), we have \( \bar{\gamma}_t^{-2} \leq \frac{4}{(1 + \sqrt{\rho})^2} \). With regard to (A.16), we establish

\[
I_2 \leq E \left[ M^2 \frac{A^2 T^2 U_t^2}{(1 + \sqrt{\rho})^2 n \mathbb{1}\{\mathcal{A}_{n,t}\}} \right] \leq \frac{4M^2 A^2 E[T^2 U_t^2]}{(1 + \sqrt{\rho})^2 n} = O(n^{-1}).
\]

(A.19)

Combining equations (A.14), (A.15) and (A.19) yields

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

and, together with (A.12), 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.12) by

\[
II \leq \sum_{t=1}^{n} \sum_{r=1}^{n} \left( E \left[ \left( F(x\bar{\gamma}_t(u)) - F(x\gamma_t(u)) \right)^2 \right] \right)^{1/2}
\]

\[
\times \left( E \left[ \left( F(x\bar{\gamma}_t(u)) - F(x\gamma_s(u)) \right)^2 \right] \right)^{1/2}.
\]

(A.20)
We define the events

\[ B_t = \left\{ \frac{\rho C_1}{\omega} \leq \rho^{t/2} \right\} \quad \text{and} \quad C_{n,t} = A_{n,t} \cap B_t. \quad (A.21) \]

In analogy to (A.14), the inner part of (A.20) can be bounded by

\[
\mathbb{E}\left[ \left( F(x_{\tilde{\gamma}_t}(u)) - F(x_{\gamma_t}(u)) \right)^2 \right] \leq \mathbb{P}\left[ C_{n,t}^c \right] + \mathbb{E}\left[ \left( F(x_{\tilde{\gamma}_t}(u)) - F(x_{\gamma_t}(u)) \right)^2 \mathbbm{1}_{C_{n,t}} \right].
\]

Employing (A.15) and Markov’s inequality yields

\[
II_1 = \mathbb{P}\left[ A_{n,t}^c \cup B_t^c \right] \leq \mathbb{P}\left[ A_{n,t}^c \right] + \mathbb{P}\left[ B_t^c \right] = \mathbb{P}\left[ A_{n,t}^c \right] + \mathbb{P}\left[ \rho^{t/2} \frac{C_1}{\omega} > 1 \right] \leq \frac{1}{n} \left( 2A \right)^2 E[T_{t}^2 U_{t}^2] + \left( \rho^{s/2}/\omega^{s} \right) \mathbb{E}[C_{n,t}^s] = O\left(n^{-1}\right) + O\left(\left(\rho^{s/2}\right)^t\right).
\]

Regarding \( II_2 \), the mean value theorem implies

\[
II_2 = \mathbb{E}\left[ \left( x_{\tilde{\gamma}_t} f(x_{\tilde{\gamma}_t}) \right)^2 \tilde{\gamma}_t^{-2} \left( \tilde{\gamma}_t(u) - \gamma_t(u) \right)^2 \mathbbm{1}_{C_{n,t}} \right],
\]

with \( \tilde{\gamma}_t \) between \( \gamma_t(u) \) and \( \gamma_t(u) \). Whereas \( \left( x_{\tilde{\gamma}_t} f(x_{\tilde{\gamma}_t}) \right)^2 \leq M^2 \) by Assumption 5(iv), Assumption 4(i) gives

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

Hence, if

\[ B_t \text{ holds } \Rightarrow \sup_{||u|| \leq A} |\tilde{\gamma}_t(u) - \gamma_t(u)| \leq \rho^{t/2} \leq \sqrt{\rho}. \quad (A.25) \]
In the event $\mathcal{C}_{n,t} = \mathcal{A}_{n,t} \cap \mathcal{B}_t$, equations (A.18) and (A.25) yield

$$\gamma_t \geq \gamma_t(u) - |\tilde{\gamma}_t(u) - \gamma_t(u)| \geq 1 - |\gamma_t(u) - 1| - |\tilde{\gamma}_t(u) - \gamma_t(u)|$$

$$\geq 1 - \frac{1 - \sqrt{\rho}}{2} - \sqrt{\rho} = \frac{1 - \sqrt{\rho}}{2},$$

where the first inequality follows from $\tilde{\gamma}_t$ being between $\tilde{\gamma}_t(u)$ and $\gamma_t(u)$. Therefore, we have $\tilde{\gamma}_t^{-2} \leq \frac{4}{(1 - \sqrt{\rho})^2}$ in the event of $\mathcal{C}_{n,t}$. With regard to (A.23), we establish

$$II_2 \leq \mathbb{E}\left[ \frac{M^2}{(1 - \sqrt{\rho})^2} \left( \frac{\rho^t C_1}{\omega} \right)^2 1_{\{\mathcal{C}_{n,t}\}} \right] \leq \frac{4M^2}{(1 - \sqrt{\rho})^2} \rho^t = O(\rho^t). \quad (A.26)$$

Equations (A.22) and (A.26) imply

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

for some constant $C > 0$. Inserting this result into (A.20), 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 $\bar{x} = x_1 < x_2 < \cdots < x_N < x_{N+1} = \bar{x}$ satisfying $M(x_{j+1} - x_j)/|\bar{x}| \leq n^{-3/4}$ for all $j = 1, \ldots, N$ and $N \in \mathbb{N}$. It
follows that \(N = O(n^{3/4})\). We obtain

\[
\sup_{x \in X} |S_n(x, u)| = \max_{1 \leq j \leq N} \sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u)|
\]

\[
\leq \max_{1 \leq j \leq N} \sup_{x_j \leq x \leq x_{j+1}} \left( |S_n(x_{j+1}, u)| + |S_n(x, u) - S_n(x_{j+1}, u)| \right)
\]

\[
\leq \max_{1 \leq j \leq N} |S_n(x_{j+1}, u)| + \max_{1 \leq j \leq N} \sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u) - S_n(x_{j+1}, u)|.
\]

We bound the second term using the elementary inequality

\[
|x - y| \leq \max\{x, y\}
\]

for all \(x, y \geq 0\). For \(j = 1, \ldots, N\), we have

\[
\sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u) - S_n(x_{j+1}, u)|
\]

\[
= \sup_{x_j \leq x \leq x_{j+1}} \left| \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_j} - \mathbb{1}_{\eta \leq x} + F(x_{j+1} \tilde{\gamma}_t(u)) - F(x \tilde{\gamma}_t(u)) \right) \right|
\]

\[
- \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_j} - \mathbb{1}_{\eta \leq x_{j+1}} + F(x_{j+1}) - F(x) \right)
\]

\[
\leq \sup_{x_j \leq x \leq x_{j+1}} \max \left\{ \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_j} - \mathbb{1}_{\eta \leq x} + F(x_{j+1} \tilde{\gamma}_t(u)) - F(x \tilde{\gamma}_t(u)) \right), \right. \]

\[
\sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_j} - \mathbb{1}_{\eta \leq x_{j+1}} + F(x_{j+1}) - F(x) \right) \right\}
\]

\[
\leq \max \left\{ \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_j} - \mathbb{1}_{\eta \leq x} + F(x_{j+1} \tilde{\gamma}_t(u)) - F(x_{j} \tilde{\gamma}_t(u)) \right), \right. \]

\[
\left. \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_j} - \mathbb{1}_{\eta \leq x} + F(x_{j+1}) - F(x) \right) \right\}
\]

\[
\leq \mathcal{A}_n \quad \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_j} - \mathbb{1}_{\eta \leq x} + F(x_{j+1}) - F(x) \right) \quad \mathcal{B}_n
\]

55
Note that $A_n$ and $B_n$ are positive, where the later can be rewritten as

\[
B_n = \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_{j+1} \tilde{\gamma}_t(u)} - F(x_{j+1} \tilde{\gamma}_t(u)) - \mathbb{1}_{\eta \leq x_j} + F(x_j) \right)
- \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_j \tilde{\gamma}_t(u)} - F(x_j \tilde{\gamma}_t(u)) - \mathbb{1}_{\eta \leq x_j} + F(x_j) \right)
+ \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_{j+1}} - \mathbb{1}_{\eta \leq x_j} + F(x_{j+1} \tilde{\gamma}_t(u)) - F(x_j \tilde{\gamma}_t(u)) \right)
= S_n(x_{j+1}, u) - S_n(x_j, u) + A_n.
\] (A.30)

It follows from (A.29) and (A.30) that

\[
\sup_{x_j \leq x \leq x_{j+1}} |S_n(x, u) - S_n(x_{j+1}, u)| \leq |S_n(x_{j+1}, u)| + |S_n(x_j, u)| + A_n.
\] (A.31)

Moreover, $A_n$ can be bounded as follows:

\[
A_n = \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_{j+1}} - F(x_{j+1}) - \mathbb{1}_{\eta \leq x_j} + F(x_j) \right) + n(F(x_{j+1}) - F(x_j))
+ \sum_{t=1}^{n} \left( F(x_{j+1} \tilde{\gamma}_t(u)) - F(x_j \tilde{\gamma}_t(u)) \right)
\leq \sum_{t=1}^{n} \left( \mathbb{1}_{\eta \leq x_{j+1}} - F(x_{j+1}) - \mathbb{1}_{\eta \leq x_j} + F(x_j) \right) + n(F(x_{j+1}) - F(x_j))
+ \sum_{t=1}^{n} \left( F(x_{j+1} \tilde{\gamma}_t(u)) - F(x_j \tilde{\gamma}_t(u)) \right).
\] (A.32)

Using equations (A.27), (A.31) and (A.32), we establish

\[
\sup_{x \in X} |S_n(x, u)| \leq 3III + IV + 2V + VI + VII,
\] (A.33)
where

\[ III = \max_{1 \leq j \leq N+1} |S_n(x_j, u)| \]

\[ IV = \max_{0 \leq j \leq N} \left| \sum_{t=1}^{n} (\mathbb{1}_{\eta_t \leq x_{j+1}} - F(x_{j+1})) - \sum_{t=1}^{n} (\mathbb{1}_{\eta_t \leq x_j} - F(x_j)) \right| \]

\[ V = \max_{2 \leq j \leq N} \sum_{t=1}^{n} F(x_j \tilde{\gamma}_t(u)) - F(x_j \gamma_t(u)) \]

\[ VI = \max_{1 \leq j \leq N} \sum_{t=1}^{n} (F(x_{j+1} \gamma_t(u)) - F(x_j \gamma_t(u))) \]

\[ VII = n \max_{1 \leq j \leq N} (F(x_{j+1}) - F(x_j)). \]

We look at each term in turn. For each \( \varepsilon > 0 \), Markov’s inequality implies

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

\[ \leq \sum_{j=1}^{N+1} \frac{1}{n^2\varepsilon^4} \mathbb{E} \left[ |S_n(x_j, u)|^4 \right] = \sum_{j=1}^{N+1} O(n^{-1}) = O(n^{-1/4}) \]

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

Regarding \( IV \), 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

\[
IV/\sqrt{n} = \max_{0 \leq j \leq N} \left| \sqrt{n} (F_n(x_{j+1}) - F(x_{j+1})) - \sqrt{n} (F_n(x_j) - F(x_j)) \right|
\]

\[
\leq \max_{0 \leq j \leq N} \left| B_n(F(x_{j+1})) - B_n(F(x_j)) \right| + \max_{0 \leq j \leq N} \left| \sqrt{n} (F_n(x_j) - F(x_j)) - B_n(F(x_j)) \right|
\]

\[
+ \max_{0 \leq j \leq N} \left| \sqrt{n} (F_n(x_{j+1}) - F(x_{j+1})) - B_n(F(x_{j+1})) \right|
\]

\[
\leq \max_{0 \leq j \leq N} \left| B_n(F(x_{j+1})) - B_n(F(x_j)) \right| + 2 \sup_{x \in \mathbb{R}} \left| \sqrt{n} (F_n(x) - F(x)) - B_n(F(x)) \right|
\]

\[
\xrightarrow{a.s.} \max_{0 \leq j \leq N} \left| B_n(F(x_{j+1})) - B_n(F(x_j)) \right| + o(1)
\]

Next, we show that \( \max_{0 \leq j \leq N} |Z_{n,j}| = o_p(1) \). By the definition of a Brownian bridge (c.f. Csörgő and Révész 1981, p. 41), \( Z_{n,j} \) is Gaussian with mean 0 and variance

\[
\text{Var}[Z_{n,j}] = (F(x_{j+1}) - F(x_j)) \left( 1 - (F(x_{j+1}) - F(x_j)) \right).
\]

Whereas the second factor of the variance is less than one, the mean value theorem and Assumption 5(iv) yield that the first factor is bounded by

\[
F(x_{j+1}) - F(x_j) = f(\bar{x}_j)(x_{j+1} - x_j) = |\bar{x}_j| f(\bar{x}_j)(x_{j+1} - x_j)/|\bar{x}_j|
\]

\[
\leq M(x_{j+1} - x_j)/|\bar{x}| \leq n^{-3/4},
\]

where \( \bar{x}_j \in (x_j, x_{j+1}) \). Hence, we have \( \text{Var}(Z_{n,j}) \leq n^{-3/4} \). Together with \( Z_{n,j} \) being Gaussian, we obtain \( \mathbb{E}[Z_{n,j}^4] = 3(\text{Var}[Z_{n,j}])^2 \leq 3n^{-3/2} \). Thus, for each \( \varepsilon > 0 \),
Markov’s inequality implies
\[
\mathbb{P} \left[ \max_{0 \leq j \leq N} |Z_{n,j}| \geq \varepsilon \right] = \mathbb{P} \left[ \max_{0 \leq j \leq N} Z_{n,j}^4 \geq \varepsilon^4 \right] \leq \frac{1}{\varepsilon^4} \mathbb{E} \left[ \max_{0 \leq j \leq N} Z_{n,j}^4 \right]
\]
\[
\leq \frac{1}{\varepsilon^4} \mathbb{E} \left[ \sum_{j=0}^{N} Z_{n,j}^4 \right] \leq \frac{1}{\varepsilon^4} \sum_{j=0}^{N} 3n^{-3/2} = \frac{3}{\varepsilon^4} n^{-3/2} (N + 1) = O(n^{-3/4})
\]

and we conclude \(\max_{0 \leq j \leq N} |Z_{n,j}| = o_p(1)\). Thus, \(IV = o_p(\sqrt{n})\). Next, we show

\[
V^\circ = \sup_{||u|| \leq A} \sup_{x \in X} \sum_{t=1}^{n} \left| F(x_{t-1}^\circ(u)) - F(x_{t-1}(u)) \right| = O_p(1), \quad (A.35)
\]

which implies \(V = O_p(1)\). We have

\[
V^\circ = \sup_{||u|| \leq A} \sup_{x \in X} \sum_{t=1}^{n} \left| F(x_{t-1}^\circ(u)) - F(x_{t-1}(u)) \right| \leq \sum_{t=1}^{n} 1_{\{c_{n,t}^c\}} + \sup_{||u|| \leq A} \sup_{x \in X} \sum_{t=1}^{n} \left| F(x_{t-1}^\circ(u)) - F(x_{t-1}(u)) \right| 1_{\{c_{n,t}\}}
\]

where the event \(c_{n,t} = A_{n,t} \cap B_t\) is defined in (A.21). We show that both terms are \(O_p(1)\). Employing Markov’s inequality and equation (A.22), we have for each \(C > 0\)

\[
\mathbb{P}[V_1^\circ \geq C] \leq \frac{1}{C} \mathbb{E}[V_1^\circ] = \frac{1}{C} \sum_{t=1}^{n} \mathbb{P}[c_{n,t}^c] \leq \frac{1}{C} \sum_{t=1}^{n} \left( \mathbb{P}[c_{n,t}^c] + \mathbb{P}[B_t^c] \right)
\]

\[
\leq \frac{1}{C} \sum_{t=1}^{n} \left( \frac{(2A)^2 \mathbb{E}[T_t^2 U_t^2]}{n} + \left( \rho^{s/2} \right) \frac{\mathbb{E}[C_{t}^s]}{\omega^s} \right) \leq \frac{1}{C} \left( \frac{(2A)^2 \mathbb{E}[T_t^2 U_t^2]}{(1 - \sqrt{\rho})^2} + \frac{\mathbb{E}[C_{t}^s]}{\omega^s(1 - \rho^{s/2})} \right).
\]

Choosing \(C\) sufficiently large, \(\mathbb{P}[V_1^\circ \geq C]\) can be made sufficiently small and we
conclude $V_1^o = O_p(1)$. Similarly to equations (A.23) to (A.26) we obtain

$$V^o_2 = \sup_{\|u\| \leq A} \sup x \in X \sum_{t=1}^{n} |x_{\gamma_t} f(x_{\gamma_t})| |\gamma_t^{-1}(u) - \gamma_t(u)| \mathbb{1}_{\{\epsilon_{n,t}\}}$$

(A.37)

and we conclude $V^o = O_p(1)$. Regarding $VI$, the mean value theorem implies

$$VI = \max_{1 \leq j \leq N} \sum_{t=1}^{n} |\hat{x}_j \gamma_t(u) f(\hat{x}_j \gamma_t(u))| |(x_{j+1} - x_j)|/|\hat{x}|$$

(A.38)

$$\leq \sum_{t=1}^{n} M \frac{2}{1 - \sqrt{p}} \frac{C_1 \rho^t}{\omega} \mathbb{1}_{\{\epsilon_{n,t}\}} \leq \sum_{t=1}^{n} \frac{2M}{1 - \sqrt{p}} \rho^{t/2} \leq \frac{2M}{(1 - \sqrt{p})^2} = O(1)$$

where $\hat{x}_j$ lies between $x_j$ and $x_{j+1}$. Thus, $VI = O(n^{1/4})$. Last, $VII = O(n^{1/4})$ as

$$VII = n \max_{1 \leq j \leq N} (F(x_{j+1}) - F(x_j)) \leq n \ n_{-3/4} = n^{1/4},$$

(A.39)

where the inequality follows from (A.34). Step 2 is completed.

In Step 3 we divide the (hyper-)cube $[-A, A]^r$ into $L = (2N)^r$ (hyper-)cubes with side length $A/N$ and $N \in \mathbb{N}$. In case of a (hyper-)cube $\ell$, $u_{\bullet}(\ell)$ and $u^{*}(\ell)$ denote the lower left and upper right vertex of $\ell$. Similar to (A.27), we obtain

$$\sup_{\|u\| \leq A} \sup x \in X \left| S_n(x, u) \right| \leq \max_{1 \leq \ell \leq L} \sup_{x \in X} \left| S_n(x, u^{*}(\ell)) \right|$$

(A.40)

$$+ \max_{1 \leq \ell \leq L} \sup_{u_{\bullet}(\ell) \leq u \leq u^{*}(\ell)} \sup_{x \in X} \left| S_n(x, u) - S_n(x, u^{*}(\ell)) \right|.$$ 

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

9Lower left (right) vertex means that all coordinates of $u_{\bullet}(\ell)$ ($u^{*}(\ell)$) are less (larger) than or equal to the corresponding coordinates of any elements of $\ell$. 

60
\[ \tilde{\gamma}_\ell(u^\bullet(\ell)). \] Since \( x < 0 \) for all \( x \in X \), the elementary inequality \([A.28]\) implies

\[
\begin{align*}
|S_n(x, u) - S_n(x, u^\bullet(\ell))| &= \sum_{t=1}^n \left( \mathbb{1}_{\{y \leq x\tilde{\gamma}_t(u)\}} - F(x\tilde{\gamma}_t(u)) - \mathbb{1}_{\{y \leq x\}} - F(x) \right) \\
&\quad - \sum_{t=1}^n \left( \mathbb{1}_{\{y \leq x\tilde{\gamma}_t(u^\bullet(\ell))\}} - F(x\tilde{\gamma}_t(u^\bullet(\ell))) - \mathbb{1}_{\{y \leq x\}} - F(x) \right) \\
&= \sum_{t=1}^n \left( \mathbb{1}_{\{y \leq x\tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{y \leq x\tilde{\gamma}_t(u^\bullet(\ell))\}} - \sum_{t=1}^n \left( F(x\tilde{\gamma}_t(u)) - F(x\tilde{\gamma}_t(u^\bullet(\ell))) \right) \right) \\
&\leq \max \left\{ \sum_{t=1}^n \left( \mathbb{1}_{\{y \leq x\tilde{\gamma}_t(u)\}} - \mathbb{1}_{\{y \leq x\tilde{\gamma}_t(u^\bullet(\ell))\}} \right) , \sum_{t=1}^n \left( F(x\tilde{\gamma}_t(u)) - F(x\tilde{\gamma}_t(u^\bullet(\ell))) \right) \right\} \\
&\quad \text{[C_n]} \\
&\quad \text{[D_n]}
\end{align*}
\]

Note that \( C_n \) can be written as

\[
C_n = \sum_{t=1}^n \left( \mathbb{1}_{\{y \leq x\tilde{\gamma}_t(u^\bullet(\ell))\}} - F(x\tilde{\gamma}_t(u^\bullet(\ell))) - \mathbb{1}_{\{y \leq x\}} - F(x) \right) \\
&\quad - \sum_{t=1}^n \left( \mathbb{1}_{\{y \leq x\tilde{\gamma}_t(u^\bullet(\ell))\}} - F(x\tilde{\gamma}_t(u^\bullet(\ell))) - \mathbb{1}_{\{y \leq x\}} - F(x) \right) \\
&\quad + \sum_{t=1}^n \left( F(x\tilde{\gamma}_t(u^\bullet(\ell))) - F(x\tilde{\gamma}_t(u^\bullet(\ell))) \right) \\
&= S_n(x, u^\bullet(\ell)) - S_n(x, u^\bullet(\ell)) + D_n. 
\]

Combining equations \([A.41]\) and \([A.42]\), we find

\[
|S_n(x, u) - S_n(x, u^\bullet(\ell))| \leq |S_n(x, u^\bullet(\ell))| + |S_n(x, u^\bullet(\ell))| + |D_n|. \quad \text{[A.43]}
\]
Equations (A.40) and (A.43) lead to

\[
\sup_{||u|| \leq A} \sup_{x \in X} |S_n(x, u)| \leq 2VIII + IX + X \tag{A.44}
\]

with

\[
VIII = \max_{1 \leq \ell \leq L} \sup_{x \in X} |S_n(x, u^\star(\ell))|
\]

\[
IX = \max_{1 \leq \ell \leq L} \sup_{x \in X} |S_n(x, u_\star(\ell))|
\]

\[
X = \max_{1 \leq \ell \leq L} \sup_{x \in X} \sum_{t=1}^n \left| F(x^\gamma_t(u_\star(\ell))) - F(x^\gamma_t(u^\star(\ell))) \right|
\]

VIII and IX are \(o_p(\sqrt{n})\) for fixed \(L\) by Step 2. Moreover, we have

\[
X \leq 2 \sup_{||u|| \leq A} \sup_{x \in X} \sum_{t=1}^n \left| F(x^\gamma_t(u^\star)) - F(x^\gamma_t(u)) \right|
\]
\[
\underbrace{X_1 + \max_{1 \leq \ell \leq L} \sup_{x \in X} \sum_{t=1}^n \left| F(x^\gamma_t(u_\star(\ell))) - F(x^\gamma_t(u^\star(\ell))) \right|}_{X_2},
\]

where \(X_1 = O_p(1)\) by (A.35). For \(n\) sufficiently large such that \(\{ \theta : ||\theta - \theta_0|| \leq \)
Hence, \( A/\sqrt{n} \subseteq \nu(\theta_0) \), the mean value theorem implies

\[
X_2 = \max_{1 \leq \ell \leq L} \sup_{x \in X} \sum_{t=1}^{n} \left| x \gamma_t f(x \gamma_t) \right| \gamma_t^{-1}(\gamma_t(u^*(\ell)) - \gamma_t(u^*(\ell)))
\]

\[
\leq M \max_{1 \leq \ell \leq L} \sum_{t=1}^{n} \frac{\gamma_t(u^*(\ell)) - \gamma_t(u^*(\ell))}{\gamma_t(u^*(\ell))}
\]

\[
= M \max_{1 \leq \ell \leq L} \sum_{t=1}^{n} \frac{\sigma_t(\theta_0 + n^{-1/2}u^*(\ell)) - \sigma_t(\theta_0 + n^{-1/2}u^*(\ell))}{\sigma_t(\theta_0 + n^{-1/2}u^*(\ell))}
\]

\[
= M \max_{1 \leq \ell \leq L} \sum_{t=1}^{n} \frac{1}{\sigma_t(\theta_0 + n^{-1/2}u^*(\ell))} \frac{\partial \sigma_t(\theta)}{\partial \theta} \frac{1}{\sqrt{n}} (u^*(\ell) - u^*(\ell))
\]

\[
\leq M \max_{1 \leq \ell \leq L} \sum_{t=1}^{n} \frac{\sigma_t(\theta_0)}{\sigma_t(\theta_0 + n^{-1/2}u^*(\ell))} \frac{\sigma_t(\theta)}{\sigma_t(\theta_0)} \frac{1}{\sigma_t(\theta)} \frac{\partial \sigma_t(\theta)}{\partial \theta} \| u^*(\ell) - u^*(\ell) \|
\]

\[
\leq rAM \frac{1}{\bar{x} \sqrt{n} \sum_{1 \leq t \leq L}^n} \sum_{1 \leq t \leq L} S_t T_t U_t \leq \left( \frac{1}{n} \sum_{t=1}^{n} S_t^3 \right)^{1/4} \left( \frac{1}{n} \sum_{t=1}^{n} T_t^3 \right)^{1/4} \left( \frac{1}{n} \sum_{t=1}^{n} U_t^3 \right)^{1/4}.
\]

where \( \theta_0 + n^{-1/2}u^*(\ell) \leq \bar{\theta} \leq \theta_0 + n^{-1/2}u^*(\ell) \) (componentwise) and \( \gamma_t \) lies between \( \gamma_t(u^*(\ell)) \) and \( \gamma_t(u^*(\ell)) \). Hölder’s inequality, the uniform ergodic theorem and Assumption 4 yield

\[
\frac{1}{n} \sum_{t=1}^{n} S_t^3 T_t U_t \leq \left( \frac{1}{n} \sum_{t=1}^{n} S_t^3 \right)^{1/4} \left( \frac{1}{n} \sum_{t=1}^{n} T_t^3 \right)^{1/4} \left( \frac{1}{n} \sum_{t=1}^{n} U_t^3 \right)^{1/4}.
\]

Hence, \( X_2 = O(\sqrt{n})/N \) almost surely, where the \( O(\sqrt{n}) \) term does not depend on \( N \). Choosing \( N \) large, we obtain \( X_2 = o(\sqrt{n}) \) almost surely and we conclude that \( X = o_p(\sqrt{n}) \). Step 3 is completed.
Regarding Step 4 we establish the following bound:

\[
\sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( F(\gamma_t(u)x) - F(x) \right) - xf(x)\Omega' \right|
\leq \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( F(x\gamma_t(u)) - F(x\gamma_t(u)) \right) \right|
\leq \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( F(x\gamma_t(u)) - F(x) \right) - xf(x)\frac{1}{n} \sum_{t=1}^{n} D'_t u \right|
\]

\[
= X_{I}
\]

\[
+ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( F(x) - xf(x)\frac{1}{n} \sum_{t=1}^{n} D'_t u \right) \right|
\]

\[
= X_{II}
\]

\[
+ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( x f(x) \gamma_t(u) - F(x) \right) - xf(x)\frac{1}{n} \sum_{t=1}^{n} D'_t u \right|
\]

\[
= X_{III}
\]

where \( X_I = O_p(n^{-1/2}) \) follows from (A.35). Further, the ergodic theorem implies

\[
X_{II} \leq \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} D_t - \Omega \right| \|u\| \leq A \sup_{n} \left| \frac{1}{n} \sum_{t=1}^{n} D_t - \Omega \right| \xrightarrow{a.s.} 0.
\]

Regarding the last term, we use the mean value theorem to obtain

\[
X_{III} = \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( x f(x)\gamma_t(u) - 1 \right) - xf(x)\frac{1}{\sqrt{n}} D'_t u \right|
\]

\[
\leq \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( f(x)\gamma_t(u) - 1 \right) - xf(x)\frac{1}{\sqrt{n}} D'_t u \right|
\]

\[
+ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( x f(x)\gamma_t(u) - 1 \right) - xf(x)(\gamma_t(u) - 1) \right|
\]

\[
\leq \frac{M}{\sqrt{n}} \sum_{t=1}^{n} \sup_{\|u\| \leq A} \left| \gamma_t(u) - 1 - \frac{1}{\sqrt{n}} D'_t u \right|
\]

\[
X_{III_1}
\]

\[
+ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( x f(x)\gamma_t(u) - f(x) \right) \left( \gamma_t(u) - 1 \right) \right|
\]

\[
X_{III_2}
\]
with \( \hat{\gamma}_t \) 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

\[
XIII_1 = \frac{M}{\sqrt{n}} \sum_{t=1}^{n} \sup_{||u|| \leq A} \left( \frac{1}{\sigma_t(\theta_0)} \left| \sigma_t(\theta_0 + n^{-1/2}u) - \sigma_t(\theta_0) - \frac{1}{\sqrt{n}} \partial \sigma_t(\theta_0) u \right| \right)
\]

\[
= \frac{M}{\sqrt{n}} \sum_{t=1}^{n} \sup_{||u|| \leq A} \left( \frac{1}{\sigma_t(\theta_0)} \left| 1 + \frac{2}{2n^2} \frac{\partial^2 \sigma_t(\tilde{\theta})}{\partial \theta \partial \theta'} u \right| \leq \frac{A^2 M}{2n^{3/2}} \sum_{t=1}^{n} \sup_{||u|| \leq A/\sqrt{n}} \left( \frac{1}{\sigma_t(\theta_0)} \left| 1 + \frac{2}{2n^2} \frac{\partial^2 \sigma_t(\tilde{\theta})}{\partial \theta \partial \theta'} u \right| \right)
\]

\[
\leq \frac{A^2 M}{2n^{3/2}} \sum_{t=1}^{n} \sup_{||\theta - \theta_0|| \leq A/\sqrt{n}} \sup_{||\theta - \theta_0|| \leq A/\sqrt{n}} ||H_t(\theta)|| \leq \frac{A^2 M}{2n^{3/2}} \sum_{t=1}^{n} T_t V_t
\]

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

\[
\frac{1}{n} \sum_{t=1}^{n} T_t V_t \leq \left( \frac{1}{n} \sum_{t=1}^{n} T_t^2 \right)^{1/2} \left( \frac{1}{n} \sum_{t=1}^{n} V_t^2 \right)^{1/2}.
\]

and we conclude that \( XIII_1 = O(n^{-1/2}) \) almost surely. Before turning to \( XIII_2 \), 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) \( \sup_{||u|| \leq A} \sup_{x \in \mathcal{X}} \max_{1 \leq t \leq n} \left( |f(x\gamma_t) - f(x)| = o_p(1). \right) \)

The first statement follows from (A.17), the Cauchy-Schwarz inequality, the uniform ergodic theorem and Assumption 9

\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sup_{||u|| \leq A} |\gamma_t(u) - 1| \leq \frac{A}{n} \sum_{t=1}^{n} T_t U_t \leq A \left( \frac{1}{n} \sum_{t=1}^{n} T_t^2 \right)^{1/2} \left( \frac{1}{n} \sum_{t=1}^{n} U_t^2 \right)^{1/2}.
\]

Note that the Cauchy-Schwarz inequality and Assumption 9 yield \( \mathbb{E}\left[ (T_t U_t)^3 \right] \leq \mathbb{E}[T_t^6] \mathbb{E}[U_t^6]^{1/2} < \infty \). 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

\[
P \left[ \sup_{\|u\| \leq A} \max_{1 \leq t \leq n} |\gamma_t(u) - 1| \geq \varepsilon \right] \leq P \left[ A \max_{1 \leq t \leq n} U_{t,c} \geq \varepsilon \sqrt{n} \right]
\]

\[ \leq P \left[ A^3 \max_{1 \leq t \leq n} (T_t U_t)^3 \geq \varepsilon^3 n^{3/2} \right] \leq \frac{A^3}{n^{3/2}} \mathbb{E} \left[ \max_{1 \leq t \leq n} (T_t U_t)^3 \right] \leq \frac{A^3}{\sqrt{n} \varepsilon^3} \mathbb{E} \left[ (T_t U_t)^3 \right], \]

which converges to 0, and thus we obtain \( \sup_{\|u\| \leq A} \max_{1 \leq t \leq n} |\gamma_t(u) - 1| = o_p(1) \).

Together with the fact that \( f \) is uniformly\(^{10}\) continuous on \( \mathcal{X} \) the second auxiliary result follows. Employing both auxiliary results, we obtain

\[ XIII_2 \leq \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \frac{1}{\sqrt{n}} \sum_{t=1}^{n} |x| \left| f(x \hat{\gamma}_t) - f(x) \right| |\gamma_t(u) - 1|
\]

\[ \leq |x| \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \max_{1 \leq t \leq n} \left| f(x \hat{\gamma}_t) - f(x) \right| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \sup_{\|u\| \leq A} |\gamma_t(u) - 1| = o_p(1). \]

Thus \( XIII \) is \( o_p(1) \), which completes Step 4.

Concerning Step 5 we obtain for each \( \varepsilon > 0 \)

\[ \mathbb{P} \left[ \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1_{\{\hat{\gamma}_t \leq x\}} - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1_{\{\gamma_t \leq x\}} - x f(x) \Omega' \sqrt{n} (\hat{\theta}_n - \theta_0) \right| \geq \varepsilon \right]
\]

\[ \leq P \left[ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1_{\{\gamma_t \leq \hat{\gamma}(u)\}} - \frac{1}{\sqrt{n}} \sum_{t=1}^{n} 1_{\{\gamma_t \leq x\}} - x f(x) \Omega' u \right| \geq \varepsilon \right]
\]

\[ + \mathbb{P} \left[ \sqrt{n} \|\hat{\theta}_n - \theta_0\| > A \right]
\]

\[ \leq \mathbb{P} \left[ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} \left| \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( F(\hat{\gamma}_t(u) x) - F(x) \right) - x f(x) \Omega' u \right| \geq \frac{\varepsilon}{2} \right]
\]

\[ + \mathbb{P} \left[ \sup_{\|u\| \leq A} \sup_{x \in \mathcal{X}} |S_n(x, u)/\sqrt{n}| \geq \frac{\varepsilon}{2} \right] + \mathbb{P} \left[ \sqrt{n} \|\hat{\theta}_n - \theta_0\| > A \right]. \]

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

\(^{10}\)It follows from compactness of \( \mathcal{X} \) and \( f \) being continuous in the neighborhood of \( \xi_\alpha \).
to zero by Step 3 and Step 4, which completes Step 5.

Regarding Step 6 we first prove the statement for $w = 0$ and consider the case $w \neq 0$ thereafter. Since $\hat{\xi}_{n,\alpha} \xrightarrow{a.s.} \xi_{\alpha}$ (Theorem [1]) and thus $\hat{\xi}_{n,\alpha} + v/\sqrt{n} \xrightarrow{a.s.} \xi_{\alpha}$, we note that $\hat{\xi}_{n,\alpha}$ and $\hat{\xi}_{n,\alpha} + v/\sqrt{n}$ belong to $\mathcal{X}$ almost surely. Hence, almost surely, we have

$$\sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + v/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha}) \right)$$

$$= \sqrt{n} \left( F_n(\xi_{\alpha}) + \sqrt{n} \left( F(\hat{\xi}_{n,\alpha} + v/\sqrt{n}) - F(\xi_{\alpha}) \right) \right)$$

$$- \sqrt{n} F_n(\xi_{\alpha}) - \sqrt{n} \left( F(\hat{\xi}_{n,\alpha} - \xi_{\alpha}) \right) + o_p(1)$$

$$= \sqrt{n} \left( F(\hat{\xi}_{n,\alpha} + v/\sqrt{n}) - F(\hat{\xi}_{n,\alpha}) \right) + o_p(1)$$

where the first equality follows from Step 5 and the second equality from

$$(\hat{\xi}_{n,\alpha} + v/\sqrt{n}) f(\hat{\xi}_{n,\alpha} + v/\sqrt{n}) - \hat{\xi}_{n,\alpha} f(\hat{\xi}_{n,\alpha}) \xrightarrow{p} \xi_{\alpha} f(\xi_{\alpha}) - \xi_{\alpha} f(\xi_{\alpha}) = 0$$

and $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_p(1)$ (Theorem [2]). Let $\mathcal{I}_n = (\xi_{\alpha} - a_n, \xi_{\alpha} + a_n)$ with $a_n = \log n/\sqrt{n}$. Since $\sqrt{n}(\hat{\xi}_{n,\alpha} - \xi_{\alpha}) = O_p(1)$, the probabilities of the events $\{\hat{\xi}_{n,\alpha} \notin \mathcal{I}_n\}$ and $\{\hat{\xi}_{n,\alpha} + v/\sqrt{n} \notin \mathcal{I}_n\}$ can be made arbitrarily small for large $n$. If $\hat{\xi}_{n,\alpha}$ and $\hat{\xi}_{n,\alpha} + v/\sqrt{n}$ belong to $\mathcal{I}_n$, we employ [Bahadur (1966), Lemma 1] to obtain

$$\sqrt{n} \left( F_n(\hat{\xi}_{n,\alpha} + v/\sqrt{n}) - F_n(\hat{\xi}_{n,\alpha}) \right)$$

$$= \sqrt{n} F_n(\xi_{\alpha}) + \sqrt{n} \left( F(\hat{\xi}_{n,\alpha} + v/\sqrt{n}) - F(\xi_{\alpha}) \right)$$

$$- \sqrt{n} F_n(\xi_{\alpha}) - \sqrt{n} \left( F(\hat{\xi}_{n,\alpha} - \xi_{\alpha}) \right) + o_p(1)$$

$$= \sqrt{n} \left( F(\hat{\xi}_{n,\alpha} + v/\sqrt{n}) - F(\hat{\xi}_{n,\alpha}) \right) + o_p(1)$$

$$= \sqrt{n} f(\hat{\xi}_{n,\alpha} + b_n) + o_p(1) = v f(\hat{\xi}_{n,\alpha} + b_n) + o_p(1) \xrightarrow{p} v f(\xi_{\alpha}),$$
where \( b_n \) lies between 0 and \( v/\sqrt{n} \) and is hence \( o(1) \). Combining results, we establish

\[
\sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + v/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha}) \right) \xrightarrow{p} v f(\xi_{\alpha})
\] (A.46)

for each \( v \in \mathbb{R} \), which completes the case \( w = 0 \). If \( w \neq 0 \), equation (A.46) yields

\[
\sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + (v + w)/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n}) \right)
= \sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + (v + w)/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n}) \right) - \sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha}) \right)
\xrightarrow{p} (v + w) f(\xi_{\alpha}) - w f(\xi_{\alpha}) = v f(\xi_{\alpha}),
\]

which completes Step 6 and establishes the lemma’s claim.

Remark 4. Step 5 is closely related to Lemma 3.2 of Gao and Song (2008) with \( \Omega \) corresponding to their \( e/2 \). 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 \( \sup_{x \in \mathbb{R}} x^2 |f'(x)| < \infty \).

Remark 5. Lemma \( \exists \) entails that for each \( w \in \mathbb{R} \)

\[
\sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n} - ) \right) \xrightarrow{p} 0.
\]

To appreciate why we find for each \( v < 0 \)

\[
0 \leq \sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n} - ) \right)
\leq \sqrt{n} \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + w/\sqrt{n}) - \hat{F}_n(\hat{\xi}_{n,\alpha} + (v + w)/\sqrt{n}) \right) \xrightarrow{p} -v f(\xi_{\alpha})
\]

and the claim follows from choosing \( |v| \) sufficiently small.

68
A.2 Bootstrap Lemmas

Henceforth we use $P^*$, $E^*$, $\text{Var}^*$ 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.

1. If in addition Assumption 9(i) holds with $a = 2d$ and $d > 2$, then $E^*\left[|\eta_t^*|^m\right] \overset{a.s.}{\to} E\left[|\eta_t|^m\right]$ for $1 \leq m \leq 2d$.

2. If in addition Assumptions 6, 7 and 9(i) hold with $a = -1/4$ and $d = 2$, then $E^*\left[\eta_t^2 \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,a}\}}\right] \overset{p}{\to} E\left[\eta_t^2 \mathbb{1}_{\{\eta_t < \hat{\xi}_a\}}\right]$.

Proof. Both claims follow from Lemma 2 since $E^*\left[\eta_t^2 \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,a}\}}\right] = \frac{1}{n} \sum_{t=1}^n \hat{\eta}_t^2 \mathbb{1}_{\{\eta_t < \hat{\xi}_a\}}$ and $E^*\left[|\eta_t|^m\right] = \frac{1}{n} \sum_{t=1}^n |\hat{\eta}_t|^m$. 

Lemma 5. Suppose Assumptions 1–3, 4(i), 5(i), 6, 8 and 9(i)–(ii) hold with $a = \pm 8$ and $b = 8$. Then, we have $\hat{\theta}_n^* \overset{p}{\to} \theta_0$ almost surely.

Proof. Let $\nu > 0$ and set $\mathcal{B} = \{\theta \in \Theta : ||\theta - \theta_0|| \geq \nu\}$; We establish the result in the following three steps:

\[ \text{Step 1:} \text{ we obtain } L_n^*(\theta) - L_n^*(\hat{\theta}_n) = \frac{1}{2n} \sum_{t=1}^n \left(1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \frac{\eta_t^*}{\hat{\sigma}_t(\hat{\theta}_n)} + \log \frac{\hat{\sigma}_t(\hat{\theta}_n)}{\sigma_t^2(\theta)}\right) + R_n^*(\theta) \text{ with sup}_{\theta \in \mathcal{B}} |R_n^*(\theta)| \overset{p}{\to} 0 \text{ almost surely}; \]

\[ \text{Step 2:} \text{ There exists a } \zeta < 0 \text{ such that sup}_{\theta \in \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) < \zeta/2 + S_n^* \text{ with } S_n^* \overset{p}{\to} 0 \text{ almost surely}; \]

\[ \text{Step 3:} \text{ we show } \mathbb{P}^*\left[\hat{\theta}_n^* \in \mathcal{B}\right] \overset{a.s.}{\to} 0. \]

Regarding Step 1 we find

\[ L_n^*(\theta) - L_n^*(\hat{\theta}_n) = \frac{1}{2n} \sum_{t=1}^n \left\{\eta_t^* - \frac{\hat{\sigma}_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \eta_t^* + \log \frac{\hat{\sigma}_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)}\right\}, \]
where \( \frac{1}{n} \sum_{t=1}^{n} \eta_t^2 \overset{p}{\rightarrow} 1 \) almost surely since
\[
\mathbb{E}^*[\frac{1}{n} \sum_{t=1}^{n} \eta_t^2] = \mathbb{E}^*[\eta_t^2] \overset{a.s.}{\rightarrow} 1 \quad \text{and} \quad \mathbb{V} \text{ar}^*[\frac{1}{n} \sum_{t=1}^{n} \eta_t^2] = \frac{1}{n} \mathbb{V} \text{ar}^*[\eta_t^2] \overset{a.s.}{\rightarrow} 0
\]
by Lemma 4. It remains to show the negligibility of the initial conditions, i.e.

\[
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^{n} \left\{ \log \frac{\hat{\sigma}_t^2(\hat{\theta}_n)}{\hat{\sigma}_t^2(\theta)} - \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right\} \right| \overset{a.s.}{\rightarrow} 0 \quad (A.47)
\]

and

\[
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} - \frac{\hat{\sigma}_t^2(\hat{\theta}_n)}{\hat{\sigma}_t^2(\theta)} \right) \eta_t^2 \right| \overset{p}{\rightarrow} 0 \quad (A.48)
\]

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

\[
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^{n} \left( \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} - \log \frac{\hat{\sigma}_t^2(\hat{\theta}_n)}{\hat{\sigma}_t^2(\theta)} \right) \right| \leq \sup_{\theta \in \Theta} \left| \frac{2}{n} \sum_{t=1}^{n} \log \frac{\hat{\sigma}_t(\hat{\theta}_n)}{\sigma_t(\theta)} \right| = \sup_{\theta \in \Theta} \left| \frac{4}{n} \sum_{t=1}^{n} \log \frac{\hat{\sigma}_t(\hat{\theta}_n)}{\sigma_t(\theta)} \right| = \frac{4}{n} \sum_{t=1}^{n} \log \left( 1 + \frac{\hat{\sigma}_t(\hat{\theta}_n) - \sigma_t(\theta)}{\sigma_t(\theta)} \right)
\]

\[
\leq \frac{4 C_1 \rho^t}{\omega} \leq \frac{4 C_1 \rho^t}{\omega} \leq \frac{4 C_1}{\omega(1 - \rho)n} \overset{a.s.}{\rightarrow} 0
\]
verifying (A.47). Further, Assumption 4(i) and (A.5) imply

\[
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\sigma_t^2(\hat{\theta}_n) - \sigma_t^2(\hat{\theta})}{\sigma_t^2(\theta)} \right) \eta_t^2 \right| \leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\sigma_t^2(\hat{\theta}_n) - \sigma_t^2(\hat{\theta})}{\sigma_t^2(\theta)} \right| \eta_t^2
\]

\[
= \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \sigma_t^2(\hat{\theta}_n) \left( \frac{\sigma_t^2(\hat{\theta}_n) - \sigma_t^2(\hat{\theta})}{\sigma_t^2(\theta)} \right) + \frac{\sigma_t^2(\theta) - \sigma_t^2(\hat{\theta})}{\sigma_t^2(\theta)} |\eta_t^2|
\]

\[
\leq \sup_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \sigma_t^2(\hat{\theta}_n) \left( \frac{\sigma_t(\hat{\theta}_n) - \sigma_t(\hat{\theta})}{\sigma_t(\theta)} \right)^2 + \frac{2|\sigma_t(\hat{\theta}_n) - \sigma_t(\hat{\theta})|}{\sigma_t(\theta)} |\eta_t^2|
\]

\[
\leq \frac{1}{n} \sum_{t=1}^{n} \sigma_t^2(\hat{\theta}_n) \left( \frac{C_1^2 \omega^2}{\omega^2} + 2C_1 \rho^t \right) \frac{1}{\omega^2} \eta_t^2
\]

\[
\leq \left( \frac{2C_1^2}{\omega^4} + \frac{4C_1}{\omega^2} \right) \frac{1}{n} \sum_{t=1}^{n} \rho^t \sigma_t^2(\hat{\theta}_n) |\eta_t^2|.
\]

To verify (A.48) are left to show that \( \frac{1}{n} \sum_{t=1}^{n} \rho^t \sigma_t^2(\hat{\theta}_n) \ll 0 \) almost surely. The conditional mean and conditional variance are given by

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

\[
\text{Var}^* \left[ \frac{1}{n} \sum_{t=1}^{n} \rho^t \sigma_t^2(\hat{\theta}_n) \eta_t^2 \right] = \frac{1}{n^2} \sum_{t=1}^{n} \rho^t \sigma_t^4(\hat{\theta}_n) \text{Var}^* [\eta_t^2]
\]

respectively. Since \( \mathbb{E}^* [\eta_t^2] \xrightarrow{a.s.} 1 \) and \( \text{Var}^* [\eta_t^2] \xrightarrow{a.s.} \kappa - 1 \) by Lemma 4, it remains to show that \( \frac{1}{n} \sum_{t=1}^{n} \rho^m \sigma_t^2(\hat{\theta}_n) \) vanishes almost surely for \( m \in \{1, 2\} \). We have

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

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) results in
\[
\frac{1}{n} \sum_{t=1}^{n} \sigma_{t}^{4m}(\hat{\theta}_n) \text{ a.s.} \leq \frac{1}{n} \sum_{t=1}^{n} T_{t}^{4m} \text{ a.s.} \rightarrow \mathbb{E}[T_{t}^{4m}] < \infty.
\]

In addition, we have for \( \varepsilon > 0 \)
\[
\sum_{t=1}^{\infty} \mathbb{P}[\rho^{2mt} \sigma_{t}^{4m}(\theta_0) > \varepsilon] \leq \sum_{t=1}^{\infty} \frac{\rho^{mt/2} \mathbb{E}[\sigma_{t}^{2}(\theta_0)]}{\varepsilon^{s/(4m)}} = \frac{\mathbb{E}[\sigma_{t}^{2}(\theta_0)]}{\varepsilon^{s/(4m)}(1 - \rho^{s/2})} < \infty
\]
such that the Borel-Cantelli Lemma implies \( \rho^{2mt} \sigma_{t}^{4m} \text{ a.s.} \rightarrow 0 \) as \( t \rightarrow \infty \). Therefore, \( \frac{1}{n} \sum_{t=1}^{n} \rho^{2mt} \sigma_{t}^{4m}(\theta_0) \text{ a.s.} \rightarrow 0 \) follows by Cesàro’s lemma. Combining results, we establish \( \frac{1}{n} \sum_{t=1}^{n} \rho^{mt} \sigma_{t}^{2m}(\hat{\theta}_n) \text{ a.s.} \rightarrow 0 \) for \( m \in \{1, 2\} \), which verifies (A.48) 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 \( 1/k \). Let this subcover be enumerated by \( \mathcal{V}_k(\theta_1), \ldots, \mathcal{V}_k(\theta_K) \) with \( K = K(k) \in \mathbb{N} \). We have
\[
\sup_{\theta \in \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) = \max_{i=1,\ldots,K} \sup_{\theta \in \mathcal{V}_k(\theta_i) \cap \mathcal{B}} L_n^*(\theta) - L_n^*(\hat{\theta}_n)
\]
Next, we fix \( i \in \{1, \ldots, K\} \) and introduce \( c_{\iota} = \frac{\sigma_{t}^{2}(\theta_0 - \iota/k)}{\sigma_{t}^{2}(\hat{\theta}_n)} \) and \( \bar{c}_{\iota} = \frac{\sigma_{t}^{2}(\theta_0 + \iota/k)}{\sigma_{t}^{2}(\hat{\theta}_n)} \) with \( \iota = (1, \ldots, 1)' \in \mathbb{R}^r \) and \( k \) sufficiently large such that \( \theta_0 \pm \iota/k \) and \( \theta_i \pm \iota/k \) belong to \( \Theta \). Assumption 8 and \( \hat{\theta}_n \text{ a.s.} \rightarrow \theta_0 \) yield \( c_{\iota} \leq \frac{\sigma_{t}^{2}(\theta_0)}{\sigma_{t}^{2}(\hat{\theta}_n)} \leq \bar{c}_{\iota} \) almost surely for all \( \theta \in \mathcal{V}_k(\theta_i) \).
With regard to Step 1, we obtain for each \( M > 1 \)
\[ \sup_{\theta \in \mathcal{R}(\theta_0) \cap \mathcal{D}} L_n^*(\theta) - L_n^*(\hat{\theta}_n) \]
\[ = \sup_{\theta \in \mathcal{R}(\theta_0) \cap \mathcal{D}} \left\{ \frac{1}{2n} \sum_{t=1}^{n} \left( 1 - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \bar{\xi}_t^2 + \log \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\theta)} \right) + R_n^*(\theta) \right\} \]
\[ \leq \frac{1}{2n} \sum_{t=1}^{n} \left( 1 - \zeta_t \eta_t^2 + \log \bar{c}_t \right) + \sup_{\theta \in \Theta} \left| R_n^*(\theta) \right| \]
\[ = \frac{1}{2n} \sum_{t=1}^{n} \left( \log \bar{c}_t - \log \bar{c}_\theta \right) + \frac{1}{2n} \sum_{t=1}^{n} \left( 1 - \zeta_t \eta_t^2 + \log \bar{c}_\theta \right) + \sup_{\theta \in \Theta} \left| R_n^*(\theta) \right| \]
\[ = \frac{1}{2n} \sum_{t=1}^{n} \log \frac{\bar{c}_t}{\bar{c}_\theta} + \frac{1}{2n} \sum_{t=1}^{n} \mathbb{I}_{\{\bar{c}_t > M\}} \left( 1 - \zeta_t \eta_t^2 + \log \bar{c}_\theta \right) \]
\[ + \frac{1}{2n} \sum_{t=1}^{n} \mathbb{I}_{\{\bar{c}_t \leq M\}} \left( 1 - \zeta_t \eta_t^2 + \log \bar{c}_\theta \right) + \sup_{\theta \in \Theta} \left| R_n^*(\theta) \right| \]
\[ \leq \frac{1}{2} \frac{1}{n} \sum_{t=1}^{n} \log \frac{\bar{c}_t}{\bar{c}_\theta} + \frac{1}{2} \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_{\{\bar{c}_t > M\}} \left( 1 + \log \bar{c}_\theta \right) + \frac{1}{2} \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_{\{\bar{c}_t \leq M\}} \left( 1 - \eta_t^2 \right) \]
\[ + \frac{1}{2} \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_{\{\bar{c}_t \leq M\}} \left( 1 - \eta_t + \log \bar{c}_\theta \right) + \sup_{\theta \in \Theta} \left| R_n^*(\theta) \right| . \]

Subsequently, we consider each term in turn. The ergodic theorem yields
\[ I = \frac{1}{n} \sum_{t=1}^{n} \log \frac{\sigma_t^2(\theta_0 + \iota/k)}{\sigma_t^2(\theta_0 - \iota/k)} + \frac{1}{n} \sum_{t=1}^{n} \log \frac{\sigma_t^2(\theta_i + \iota/k)}{\sigma_t^2(\theta_i - \iota/k)} \]
\[ \xrightarrow{a.s.} \mathbb{E} \left[ \log \frac{\sigma_t^2(\theta_0 + \iota/k)}{\sigma_t^2(\theta_0 - \iota/k)} \right] + \mathbb{E} \left[ \log \frac{\sigma_t^2(\theta_i + \iota/k)}{\sigma_t^2(\theta_i - \iota/k)} \right] . \]

By the Beppo-Levi theorem, the right hand side decreases to 0 as \( k \) grow large.
Thus, almost surely, \( I \) can be made arbitrarily small by choosing \( k \) sufficiently large.

Hence, \( I = o(1) \) almost surely. Regarding \( II \), take \( k \) sufficiently large such that
\( \theta_0 - \iota/k \in \mathcal{V}(\theta_0) \) yielding \( \zeta_t \leq \sigma_t T^2 / \omega^2 \) with \( \sigma_t = \sigma_t(\theta_0) \). The ergodic theorem, the
inequality \( \log(x) \leq x \) for all \( x > 0 \) and the Cauchy-Schwarz inequality imply

\[
II \xrightarrow{a.s.} \mathbb{E} \left[ \mathbb{1}_{\{\xi_t > M\}} \left( 1 + \log \xi_t \right) \right] \leq \mathbb{E} \left[ \mathbb{1}_{\{\sigma_t^2 T_t^2 > M \omega^2\}} \left( 1 + \log \frac{\sigma_t^2 T_t^2}{\omega^2} \right) \right]
\]

\[
= \mathbb{E} \left[ \mathbb{1}_{\{\sigma_t^2 T_t^2 > M \omega^2\}} \left( 1 - 2 \log \omega + \frac{4}{s} \log \sigma_t^{s/2} + 2 \log T_t \right) \right]
\]

\[
\leq \mathbb{E} \left[ \mathbb{1}_{\{\sigma_t^2 T_t^2 > M \omega^2\}} \left( 1 - 2 \log \omega + \frac{4}{s} \sigma_t^{s/2} + 2 T_t \right) \right]
\]

\[
\leq \left( \mathbb{E} \left[ \left( 1 - 2 \log \omega + \frac{4}{s} \sigma_t^{s/2} + 2 T_t \right)^2 \right] \right)^{\frac{1}{2}} \left( \mathbb{P} \left[ \sigma_t^2 T_t^2 > M \omega^2 \right] \right)^{\frac{1}{2}}.
\]

Employing (A.10) we find that

\[
II_1 \leq 4 \left( 1 + (2 \log \omega)^2 + \frac{16}{s^2} \mathbb{E} \left[ \sigma_t^s \right] + 4 \mathbb{E} \left[ T_t^2 \right] \right) < \infty
\]

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

\[
II_2 \leq \mathbb{P} \left[ T_t^2 > M \omega^2/2 \right] + \mathbb{P} \left[ \sigma_t^2 > M \omega^2/2 \right] \leq \frac{2}{M \omega^2} \mathbb{E} \left[ T_t^2 \right] + \left( \frac{2}{\sqrt{M \omega}} \right)^s \mathbb{E} \left[ \sigma_t^s \right].
\]

Since \( II_1 \) can be made arbitrarily small by the choice of \( M \) we get \( II = o(1) \) almost surely. Further, for given \( M \), Lemma 4 entails

\[
\left| \mathbb{E}^* \left[ III \right] \right| = \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{\xi_t \leq M\}} \mathbb{E} \left[ 1 - \mathbb{E}^* \left[ \eta_t^{s/2} \right] \right] \leq M \left| 1 - \mathbb{E}^* \left[ \eta_t^{s/2} \right] \right| \xrightarrow{a.s.} 0
\]

\[
\mathbb{V} \text{ar}^* \left[ III \right] = \frac{1}{n^2} \sum_{t=1}^{n} \mathbb{1}_{\{\xi_t \leq M\}} \mathbb{E}^2 \left[ \mathbb{E}^* \left[ \eta_t^{s/2} \right] \right] \leq \frac{M^2}{n} \mathbb{V} \text{ar}^* \left[ \eta_t^{s/2} \right] \xrightarrow{a.s.} 0
\]

such that \( III \xrightarrow{p} 0 \) almost surely. Consider IV; the ergodic theorem yields

\[
IV \xrightarrow{a.s.} \mathbb{E} \left[ \mathbb{1}_{\{\xi_t \leq M\}} \left( 1 - \xi_t + \log \xi_t \right) \right]
\]

74
and recall that $c_t = \frac{\sigma^2(\theta_0 - t/k)}{\sigma^2_t(\theta_i)}$. The right-hand side approaches
\[ E \left[ 1 - \frac{\sigma^2_t(\theta_0)}{\sigma^2_t(\theta_i)} + \log \frac{\sigma^2_t(\theta_0)}{\sigma^2_t(\theta_i)} \right] = (A.49) \]
as $M$ and $k$ grow large. Thus, almost surely, IV can be made arbitrarily close to \([A.49]\) by choosing $M$ and $k$ sufficiently large. Further, since $\theta_i \neq \theta_0$ and Assumption 3 implies $\frac{\sigma^2(\theta_0)}{\sigma^2(\theta_i)} \neq 1$ 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 \(A.49\) is strictly smaller than 0. We conclude that there exists a $\zeta_i < 0$ such that $IV < \zeta_i$ holds for sufficiently large $M$ and $k$ and $n$ almost surely. Set $\zeta = \max_{i=1,...,K} \zeta_i$, 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
\[ \mathbb{P}^* [\hat{\theta}^*_n \in \mathcal{B}] \leq \mathbb{P}^* \left[ \zeta/2 + S^*_n > 0 \right] + o(1) \leq \mathbb{P}^* \left[ |S^*_n| > -\zeta/2 \right] + o(1) = o(1) \]
almost surely, which completes Step 3 and establishes the lemma’s claim. \qed

**Lemma 6.** Suppose Assumptions 1–4, 5(i), 5(iii), 6 and 9 hold with $a = \pm 12$, $b = 12$, $c = 6$ and $d = 4$. Then, we have $\frac{1}{n} \sum_{t=1}^n \frac{\partial^2}{\partial \theta \partial \theta} \ell^*_t(\hat{\theta}) \xrightarrow{p} -2J$ almost surely.
Proof. We have

\[ \frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta'} \ell_t(\tilde{\theta}) = \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\epsilon_t^2}{\sigma_t^2(\tilde{\theta})} - 1 \right) \tilde{H}_t(\tilde{\theta}) - \frac{1}{n} \sum_{t=1}^{n} \left( 3 \frac{\epsilon_t^2}{\sigma_t^2(\tilde{\theta})} - 1 \right) \tilde{D}_t(\tilde{\theta}) \tilde{D}_t'(\tilde{\theta}). \]

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

\[ I = \frac{1}{n} \sum_{t=1}^{n} \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\tilde{\theta})} H_t(\tilde{\theta}) \eta_t^2 + \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\tilde{\theta})} \tilde{H}_t(\tilde{\theta}) - \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\tilde{\theta})} H_t(\tilde{\theta}) \right) \eta_t^2 - \frac{1}{n} \sum_{t=1}^{n} \tilde{H}_t(\tilde{\theta}). \]

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

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

for all \( i, j = 1, \ldots, r \). Since \( \tilde{\theta} \) lies between \( \hat{\theta}_n^* \) and \( \hat{\theta}_n \), Theorem 1 and Lemma 5 imply \( \tilde{\theta} \xrightarrow{p} \theta_0 \) almost surely. Since \( \hat{\theta}_n \xrightarrow{a.s.} \theta_0 \) and \( \tilde{\theta} \xrightarrow{p} \theta_0 \) almost surely, we have \( \hat{\theta}_n \in \mathcal{V}_\varepsilon(\theta_0) \) almost surely and \( \tilde{\theta} \in \mathcal{V}_\varepsilon(\theta_0) \) for sufficiently large conditional probability almost surely. In such case, we have for all pairs \( (i, j) \)

\[ L_n^*(i, j) \leq \frac{1}{n} \sum_{t=1}^{n} \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\tilde{\theta})} e_i' H_t(\tilde{\theta}) e_j' \eta_t^2 \leq U_n^*(i, j), \]

76
with

\[ L^*_n(i, j) = \frac{1}{n} \sum_{t=1}^{n} \inf_{\theta_1, \theta_2 \in \mathcal{V}_c(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} e'_i H_t(\theta_2) e_j \eta^*_2 \]

\[ U^*_n(i, j) = \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta_1, \theta_2 \in \mathcal{V}_c(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} e'_i H_t(\theta_2) e_j \eta^*_2. \]

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

\[
\mathbb{E}^* \left[ U^*_n(i, j) \right] = \mathbb{E}^* \left[ \eta^*_2 \right] \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta_1, \theta_2 \in \mathcal{V}_c(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} e'_i H_t(\theta_2) e_j 
\xrightarrow{a.s.} \mathbb{E} \left[ \sup_{\theta_1, \theta_2 \in \mathcal{V}_c(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} e'_i H_t(\theta_2) e_j \right] < \mathbb{E} [e'_i H_t e_j] + \varepsilon.
\]

whereas its conditional variance vanishes:

\[
\text{Var}^* \left[ U^*_n(i, j) \right] = \text{Var}^* \left[ \eta^*_2 \right] \frac{1}{n^2} \sum_{t=1}^{n} \left( \sup_{\theta_1, \theta_2 \in \mathcal{V}_c(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} e'_i H_t(\theta_2) e_j \right)^2 
\leq \text{Var}^* \left[ \eta^*_2 \right] \frac{1}{n^2} \sum_{t=1}^{n} S^4_t T^4_t V^2_t
\leq \text{Var}^* \left[ \eta^*_2 \right] \frac{1}{n^2} \left( \frac{1}{n} \sum_{t=1}^{n} S^4_t \right)^\frac{1}{4} \left( \frac{1}{n} \sum_{t=1}^{n} T^4_t \right)^\frac{1}{4} \left( \frac{1}{n} \sum_{t=1}^{n} V^6_t \right)^\frac{1}{4} \xrightarrow{a.s.} 0.
\]

Similarly, we obtain for the lower bound

\[
\mathbb{E}^* \left[ L^*_n(i, j) \right] \xrightarrow{a.s.} \mathbb{E} \left[ \inf_{\theta_1, \theta_2 \in \mathcal{V}_c(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} e'_i H_t(\theta_2) e_j \right] > \mathbb{E} [e'_i H_t e_j] - \varepsilon
\]

and \( \text{Var}^* \left[ L^*_n(i, j) \right] \xrightarrow{a.s.} 0. \) Taking \( \varepsilon \downarrow 0 \) subsequently, we get \( \frac{1}{n} \sum_{i=1}^{n} \frac{\sigma^2(\hat{\theta}_n)}{\sigma^2(\hat{\theta})} e'_i H_t(\hat{\theta}) e_j \eta^*_2 \xrightarrow{p} \mathbb{E} [e'_i H_t e_j] \) almost surely for all pairs \( (i, j) \), which in turn yields \( I_1 \xrightarrow{p} \mathbb{E}[H_t] \) almost

77
surely. Regarding $I_2$, for any $\theta_1, \theta_2 \in \Theta$ Assumption \( \Box \) implies

$$
\frac{\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} - \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} = \frac{\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} + \sigma_t(\theta_1) \frac{\sigma_t(\theta_2) - \tilde{\sigma}_t(\theta_2)}{\sigma_t(\theta_2)} \\
\leq \frac{\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} + \sigma_t(\theta_1) \frac{\sigma_t(\theta_2) - \tilde{\sigma}_t(\theta_2)}{\sigma_t(\theta_2)} \leq C_1 \rho^t + \sigma_t(\theta_1) C_1 \rho^t = C_1 \rho^t \left( 1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right) \tag{A.50}
$$

and combined with the elementary inequalities \( \Box \) (with $m = 1$) and \( \Box \) yields

$$
\frac{\tilde{\sigma}_t^2(\theta_1) - \sigma_t^2(\theta_1)}{\tilde{\sigma}_t^2(\theta_2)} \leq \frac{\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} + \sigma_t(\theta_1) \left( \frac{\sigma_t(\theta_2) - \tilde{\sigma}_t(\theta_2)}{\sigma_t(\theta_2)} \right) \leq C_1 \rho^t + \sigma_t(\theta_1) \left( 1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right) \tag{A.51}
$$

Together with \( \Box \) it follows that

$$
\|I_2\| \leq \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n) - \tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\hat{\theta})} \right\| \eta_t^2 \\
= \frac{1}{n} \sum_{t=1}^{n} \left\| \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\hat{\theta})} \right\| \left( \tilde{H}_t(\hat{\theta}) - H_t(\hat{\theta}) \right) + \left( \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\hat{\theta})} - \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\hat{\theta})} \right) H_t(\hat{\theta}) \right\| \eta_t^2 \\
\leq \frac{1}{n} \sum_{t=1}^{n} \left\{ \left( \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\hat{\theta})} \right) \left( \tilde{H}_t(\hat{\theta}) - H_t(\hat{\theta}) \right) \right\} \eta_t^2 \\
\leq \frac{1}{n} \sum_{t=1}^{n} \left\{ \left( \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\hat{\theta})} + \frac{2C_1^2}{\omega^2} + \frac{4C_1}{\omega} \right) \rho^t \left( 1 + \frac{\tilde{\sigma}_t^2(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\hat{\theta})} \right) \right\} \eta_t^2 \\
\leq \left( \frac{5C_1}{\omega} + \frac{6C_1^2}{\omega^2} + \frac{2C_1^3}{\omega^3} \right) \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + \frac{\sigma_t(\hat{\theta}_n)}{\tilde{\sigma}_t^2(\hat{\theta})} \right) \left( 1 + \left| H_t(\hat{\theta}) \right| \right) \eta_t^2
$$

78
In the case of \( \hat{\theta}_n \in \mathcal{V}(\theta_0) \) and \( \bar{\theta} \in \mathcal{V}(\theta_0) \), we get

\[
\frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + \frac{\sigma^2_t(\hat{\theta}_n)}{\sigma^2_t(\bar{\theta})} \right) \left( 1 + ||H_t(\bar{\theta})|| \right) \eta_t^2 \leq \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2_t T^2_t \right) (1 + V_t) \eta_t^2.
\]

The conditional mean and variance of the right hand side are given by

\[
\mathbb{E}^* \left[ \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2_t T^2_t \right) (1 + V_t) \eta_t^2 \right] = \mathbb{E}^* \left[ \eta_t^2 \right] \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2_t T^2_t \right) (1 + V_t)
\]

\[
\mathbb{Var}^* \left[ \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2_t T^2_t \right) (1 + V_t) \eta_t^2 \right] = \mathbb{Var}^* \left[ \eta_t^2 \right] \frac{1}{n^2} \sum_{t=1}^{n} \rho^t \left( 1 + S^2_t T^2_t \right)^2 (1 + V_t)^2,
\]

respectively. For \( m \in \{1, 2\} \) and \( \varepsilon > 0 \), we have

\[
\sum_{t=1}^{\infty} \mathbb{P} \left[ \rho^m t (1 + S^2_t T^2_t)^m (1 + V_t)^m > \varepsilon \right] \leq \sum_{t=1}^{\infty} \rho^t \frac{\mathbb{E} \left[ (1 + S^2_t T^2_t)(1 + V_t) \right]}{\varepsilon^{1/m}} \frac{\mathbb{E} \left[ (1 + S^2_t T^2_t)(1 + V_t) \right]}{\varepsilon^{1/m}(1 - \rho)} < \infty
\]

such that the Borel-Cantelli Lemma implies \( \rho^m t (1 + S^2_t T^2_t)^m (1 + V_t)^m \xrightarrow{a.s.} 0 \) as \( t \to \infty \).

Therefore, \( \frac{1}{n} \sum_{t=1}^{n} \rho^m t (1 + S^2_t T^2_t)^m (1 + V_t)^m \xrightarrow{a.s.} 0 \) follows by Césaro’s lemma and we get \( \frac{1}{n} \sum_{t=1}^{n} \rho^t (1 + S^2_t T^2_t) (1 + V_t) \eta_t^2 \xrightarrow{p^*} 0 \) 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 \( \hat{\theta} \xrightarrow{p^*} \theta_0 \) 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

\[
II = 3 \frac{1}{n} \sum_{t=1}^{n} \frac{\sigma^2_t(\hat{\theta}_n)}{\sigma^2_t(\bar{\theta})} D_t(\bar{\theta}) D_t'(\bar{\theta}) \eta_t^2 + 3 \frac{1}{n} \sum_{t=1}^{n} \left( \frac{\sigma^2_t(\hat{\theta}_n)}{\sigma^2_t(\bar{\theta})} D_t(\bar{\theta}) D_t'(\bar{\theta}) - \frac{\sigma^2_t(\hat{\theta}_n)}{\sigma^2_t(\bar{\theta})} D_t(\bar{\theta}) D_t'(\bar{\theta}) \right) \eta_t^2
\]

\[
- \frac{1}{n} \sum_{t=1}^{n} D_t(\bar{\theta}) D_t'(\bar{\theta}) + \left( \sum_{t=1}^{n} D_t(\bar{\theta}) D_t'(\bar{\theta}) \right).
\]

79
We treat the subterms of $II$ analogously to the subterms of $I$. We begin with $II_1$ and take $\varepsilon > 0$. Since $\frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} D_i(\theta_2) D'_j(\theta_2)$ is continuous in $\theta_1$ and $\theta_2$ we can take $\mathcal{V}_\varepsilon(\theta_0) \subseteq \mathcal{V}(\theta_0)$ such that

\[
\mathbb{E}[\epsilon'_i D_i D'_j] - \varepsilon < \mathbb{E}\left[ \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} \epsilon'_i D_i(\theta_2) D'_j(\theta_2) \right] \\
\leq \mathbb{E}\left[ \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} \epsilon'_i D_i(\theta_2) D'_j(\theta_2) \right] < \mathbb{E}[\epsilon'_i D_i D'_j] + \varepsilon
\]

for all $i, j = 1, \ldots, r$. Since $\hat{\theta}_n \rightarrow^{a.s.} \theta_0$ and $\tilde{\theta} \rightarrow^{a.s.} \theta_0$ almost surely, we have $\hat{\theta}_n \in \mathcal{V}(\theta_0)$ almost surely and $\tilde{\theta} \in \mathcal{V}_\varepsilon(\theta_0)$ for sufficiently large conditional probability almost surely. In such case, we have for all pairs $(i, j)$

\[
\bar{L}^*_n(i, j) \leq \frac{1}{n} \sum_{t=1}^n \frac{\sigma^2_1(\hat{\theta}_n)}{\sigma^2_2(\tilde{\theta})} \epsilon'_i D_i(\tilde{\theta}) D'_j(\tilde{\theta}) \eta^2_t \leq \bar{U}^*_n(i, j)
\]

with

\[
\bar{L}^*_n(i, j) = \frac{1}{n} \sum_{t=1}^n \inf_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} \epsilon'_i D_i(\theta_2) D'_j(\theta_2) \eta^2_t \\
\bar{U}^*_n(i, j) = \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} \epsilon'_i D_i(\theta_2) D'_j(\theta_2) \eta^2_t.
\]

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

\[
\mathbb{E}^*[\bar{U}^*_n(i, j)] = \mathbb{E}^*[\eta^2_t] \frac{1}{n} \sum_{t=1}^n \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} \epsilon'_i D_i(\theta_2) D'_j(\theta_2) \eta^2_t \\
\rightarrow^{a.s.} \mathbb{E}\left[ \sup_{\theta_1, \theta_2 \in \mathcal{V}_\varepsilon(\theta_0)} \frac{\sigma^2_1(\theta_1)}{\sigma^2_2(\theta_2)} \epsilon'_i D_i(\theta_2) D'_j(\theta_2) \right] < \mathbb{E}[\epsilon'_i D_i D'_j] + \varepsilon.
\]
whereas its conditional variance vanishes:

\[
\var^*[\bar{U}_n^*(i,j)] = \var^*[\eta_t^2] \leq \frac{1}{n^2} \sum_{t=1}^n \left( \sup_{\theta_1, \theta_2 \in \mathcal{Y}_0} \frac{\sigma_t^2(\hat{\theta}_1)}{\sigma_t^2(\hat{\theta}_2)} e_i' D_t(\hat{\theta}_2) D_t'(\hat{\theta}_2) e_j \right)^2 \leq \frac{1}{n^2} \sum_{t=1}^n S_t^4 T_t^4 U_t^4 \leq \frac{1}{n^2} \left( \frac{1}{n} \sum_{t=1}^n S_{12}^2 \right)^{\frac{1}{4}} \left( \frac{1}{n} \sum_{t=1}^n T_{12}^2 \right)^{\frac{1}{4}} \left( \frac{1}{n} \sum_{t=1}^n U_{12}^2 \right)^{\frac{1}{4}} \rightarrow 0.
\]

Similarly, we obtain \( \frac{1}{n} \sum_{t=1}^n \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\hat{\theta})} e_i' D_t(\hat{\theta}) D_t'(\hat{\theta}) e_j \rightarrow 0 \) almost surely for all pairs \((i,j)\), which in turn yields \( II_1 \) almost surely. Regarding \( II_2 \), we employ (A.6) and (A.51) and find

\[
||II_2|| \leq \frac{1}{n} \sum_{t=1}^n \left| \frac{\hat{\sigma}^2_t(\hat{\theta}_n)}{\hat{\sigma}^2_t(\hat{\theta})} \bar{D}_t(\hat{\theta}) \bar{D}_t'(\hat{\theta}) - \frac{\hat{\sigma}^2_t(\hat{\theta}_n)}{\hat{\sigma}^2_t(\hat{\theta})} D_t(\hat{\theta}) D_t'(\hat{\theta}) \right| \eta_t^2 \leq \frac{1}{n} \sum_{t=1}^n \left( \frac{\hat{\sigma}^2_t(\hat{\theta}_n)}{\hat{\sigma}^2_t(\hat{\theta})} \right) \left| \bar{D}_t(\hat{\theta}) \bar{D}_t'(\hat{\theta}) - D_t(\hat{\theta}) D_t'(\hat{\theta}) \right| \eta_t^2 \leq \frac{1}{n} \sum_{t=1}^n \left\{ \left( \frac{\hat{\sigma}^2_t(\hat{\theta}_n)}{\hat{\sigma}^2_t(\hat{\theta})} \right) \left| \bar{D}_t(\hat{\theta}) \bar{D}_t'(\hat{\theta}) - D_t(\hat{\theta}) D_t'(\hat{\theta}) \right| \right\} \eta_t^2 \leq \frac{1}{n} \sum_{t=1}^n \left\{ \left( \frac{\sigma_t^2(\hat{\theta}_n)}{\sigma_t^2(\hat{\theta})} \right) \left| \bar{D}_t(\hat{\theta}) \bar{D}_t'(\hat{\theta}) - D_t(\hat{\theta}) D_t'(\hat{\theta}) \right| \right\} \eta_t^2 \leq \frac{1}{n} \sum_{t=1}^n \left( \frac{2C^2}{\omega^2} + \frac{4C_1}{\omega} \right) \rho_t^2 \left( 1 + \frac{\hat{\sigma}^2_t(\hat{\theta}_n)}{\hat{\sigma}^2_t(\hat{\theta})} \right) \left( \frac{C^2}{\omega^2} + \frac{2C_1}{\omega} \right) \rho_t \left( 1 + ||D_t(\hat{\theta})||^2 \right) \eta_t^2 \leq \left( \frac{6C_1}{\omega} + \frac{2C^4}{\omega^4} + \frac{8C^3}{\omega^3} + \frac{2C^4}{\omega^4} \right) \frac{1}{n} \sum_{t=1}^n \rho_t^2 \left( 1 + \frac{\hat{\sigma}^2_t(\hat{\theta}_n)}{\hat{\sigma}^2_t(\hat{\theta})} \right) \left( 1 + ||D_t(\hat{\theta})||^2 \right) \eta_t^2.
\]
In the case of \( \hat{\theta}_n \in \mathcal{V}(\theta_0) \) and \( \check{\theta} \in \mathcal{V}(\theta_0) \), we get
\[
\frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + \frac{\sigma^2(\hat{\theta}_n)}{\sigma^2(\check{\theta})} \right) \left( 1 + \|D_t(\check{\theta})\|^2 \right) \eta^2 \leq \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2 T^2 \right) (1 + U^2) \eta^2.
\]

The conditional mean and variance of the right hand side are given by
\[
\mathbb{E}^{*} \left[ \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2 T^2 \right) (1 + U^2) \eta^2 \right] = \mathbb{E}^{*} \left[ \eta^2 \right] \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2 T^2 \right) (1 + U^2) \\
\text{Var}^{*} \left[ \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2 T^2 \right) (1 + U^2) \eta^2 \right] = \text{Var}^{*} \left[ \eta^2 \right] \frac{1}{n^2} \sum_{t=1}^{n} \rho^t \left( 1 + S^2 T^2 \right) (1 + U^2)^2,
\]
respectively. For \( m \in \{1, 2\} \) and \( \varepsilon > 0 \), we have
\[
\sum_{t=1}^{\infty} \mathbb{P} \left[ \rho^m t \left( 1 + S^2 T^2 \right)^m (1 + U^2)^m > \varepsilon \right] \leq \sum_{t=1}^{\infty} \rho^t \mathbb{E} \left[ \left( 1 + S^2 T^2 \right) (1 + U^2) \right] \frac{1}{\varepsilon^{1/m}} \mathbb{E} \left[ \left( 1 + S^2 T^2 \right) (1 + U^2) \right] < \infty
\]
such that the Borel-Cantelli Lemma implies \( \rho^m t \left( 1 + S^2 T^2 \right)^m (1 + U^2)^m \overset{a.s.}{\to} 0 \) as \( t \to \infty \).

Therefore, \( \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2 T^2 \right)^m (1 + U^2)^m \overset{a.s.}{\to} 0 \) follows by Césaro’s lemma and we get \( \frac{1}{n} \sum_{t=1}^{n} \rho^t \left( 1 + S^2 T^2 \right) (1 + U^2)^2 \eta^2 \overset{p^*}{\to} 0 \) almost surely. Combining results gives \( \|II_2\| \overset{p^*}{\to} 0 \) almost surely. Similar to the proof of Lemma 2(ii), we establish \( II_3 \overset{p^*}{\to} \mathbb{E} [D_t D_t'] = J \) almost surely using \( \hat{\theta} \overset{p^*}{\to} \theta_0 \) almost surely. Combining results we find
\[
II = 3I_1 + 3I_2 - I_3 \overset{p^*}{\to} 3J + 0 - J = 2J \text{ almost surely.}
\]

In conclusion, we have
\[
\frac{1}{n} \sum_{t=1}^{n} \frac{\partial^2}{\partial \theta \partial \theta'} f^*_t(\check{\theta}) = I - II \overset{p^*}{\to} -2J
\]
almost surely, which completes the proof.

\[\square\]

**Lemma 7.** Suppose Assumptions [1], [2], [3], [5][ii], [6][iv], [8][i] and [10] hold with \( a = -1, 9, \)
\[ b = 9 \text{ and } c = 2 < d. \text{ Then, we have} \]
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \hat{D}_t(\eta_t^2 - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha) \right) \xrightarrow{d^*} N(0, \Upsilon_\alpha) \quad \text{with} \quad \Upsilon_\alpha = \begin{pmatrix} (\kappa - 1) J & p_\alpha \Omega \\ p_\alpha \Omega' & \alpha(1 - \alpha) \end{pmatrix}
\]
almost surely.

**Proof.** Set \( \alpha_n = \mathbb{E}^*\left[ \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right] \) and expand
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \hat{D}_t(\eta_t^2 - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} - \alpha) \right) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \hat{D}_t(\eta_t^2 - \mathbb{E}^*[\eta_t^2]) \right) + \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \hat{D}_t(\mathbb{E}^*[\eta_t^2] - 1) \right).
\]
Consider the second term; with regard to Remark I we have \( \mathbb{E}^*[\eta_t^2] = 1 \) whenever \( \hat{\theta}_n \in \hat{\Theta} \) under Assumption 10. Since \( \hat{\theta}_n \xrightarrow{a.s.} \theta_0 \in \hat{\Theta} \) by Theorem I 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,
\[
\alpha_n = \frac{1}{n} \sum_{t=1}^{n} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \xrightarrow{a.s.} \frac{[n\alpha] + 1}{n} = \alpha + O(n^{-1})
\]
and hence \( \frac{1}{\sqrt{n}} \sum_{t=1}^{n} (\alpha_n - \alpha) \xrightarrow{a.s.} 0 \). We conclude that \( II \) vanishes in almost surely.

Next, consider \( I \); it suffices to show that for each \( \lambda = (\lambda_1, \lambda_2)' \in \mathbb{R}^{r+1} \) with \( ||\lambda|| \neq 0 \)
\[
\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \left( \hat{D}_t(\eta_t^2 - \mathbb{E}^*[\eta_t^2]) \right) \xrightarrow{d^*} N\left(0, \lambda' \Upsilon_\alpha \lambda \right)
\]
almost surely by the Cramér-Wold device. By construction, we have \( \mathbb{E}[Z_{n,t}^*] = 0 \).

83
Further, we obtain

\[ s_n^2 = \sum_{t=1}^{n} \text{Var}^*[Z_{n,t}^*] = \lambda' \left( \text{Var}^*[\eta_t^2] \hat{J}_n + \text{Cov}^*[\eta_t^2, \mathbb{1}_{\{\eta_t^* < \xi_{n,o}\}}] \hat{\Omega}_n \right) \lambda. \]

Lemma 2 states \( \hat{J}_n \overset{a.s.}{\rightarrow} J \) and \( \hat{\Omega}_n \overset{a.s.}{\rightarrow} \Omega \). Employing Lemma 4 yields

\[ \text{Var}^*[\eta_t^2] = \mathbb{E}^*[\eta_t^2] - \mathbb{E}[\eta_t^2] \overset{a.s.}{\rightarrow} \kappa - 1 \]

\[ \text{Var}^*[\mathbb{1}_{\{\eta_t^* < \xi_{n,o}\}}] = \alpha_n(1 - \alpha_n) \overset{a.s.}{\rightarrow} \alpha(1 - \alpha) \]

\[ \text{Cov}^*[\eta_t^2, \mathbb{1}_{\{\eta_t^* < \xi_{n,o}\}}] = \mathbb{E}^*[\eta_t^2 \mathbb{1}_{\{\eta_t^* < \xi_{n,o}\}}] - \mathbb{E}[\eta_t^2] \alpha_n \overset{a.s.}{\rightarrow} p_\alpha. \]

Thus, we get \( s_n^2 \overset{a.s.}{\rightarrow} \lambda'\mathcal{T}_\alpha \lambda \). Next, we verify Lyapounov’s condition. Without loss of generality we can restrain \( d \) to \( d \in (2, 3) \). Employing the elementary inequalities

\[ (x + y)^z \leq 2^z(x^z + y^z) \quad (A.52) \]

and \( |x - y|^z \leq x^z + y^z \) for all \( x, y, z \geq 0 \) we obtain

\[ \sum_{t=1}^{n} \mathbb{E}^*[|Z_{n,t}^*|^d] = \left( \frac{1}{\sqrt{n}} \right)^d \sum_{t=1}^{n} \mathbb{E}^*\left[ \lambda_1^d \mathbb{D}_t(\eta_t^2 - \mathbb{E}^*[\eta_t^2]) + \lambda_2(\mathbb{1}_{\{\eta_t^* < \xi_{n,o}\}} - \alpha_n) \right]^d \]

\[ \leq \left( \frac{1}{\sqrt{n}} \right)^d \sum_{t=1}^{n} \mathbb{E}^*\left[ (\lambda_1^d \mathbb{D}_t|\eta_t^2 - \mathbb{E}^*[\eta_t^2]| + \lambda_2^d) \right]^d \]

\[ \leq \left( \frac{2}{\sqrt{n}} \right)^d \sum_{t=1}^{n} \left( \lambda_1^d \mathbb{D}_t^d |\eta_t^2 - \mathbb{E}^*[\eta_t^2]| + \lambda_2^d \right) \]

\[ \leq n \left( \frac{2}{\sqrt{n}} \right)^d \left( \lambda_1^d \mathbb{D}_t^d \left( \mathbb{E}^*[|\eta_t^2|] + \mathbb{E}^*[|\eta_t^2|^d] \right) + \lambda_2^d \right) \]

As \( d > 2 \), we have \( n \left( \frac{2}{\sqrt{n}} \right)^d \rightarrow 0 \). Lemma 4 gives \( \mathbb{E}^*[|\eta_t|^m] \overset{a.s.}{\rightarrow} \mathbb{E}[|\eta_t|^m] < \infty \) for
$m = 2, 2d$. Employing (A.3) and (A.52) and noting $\hat{\theta}_n \in \mathcal{V}(\theta_0)$ almost surely, we get

$$\frac{1}{n} \sum_{t=1}^{n} ||\hat{D}_t||^d \leq \frac{1}{n} \sum_{t=1}^{n} \left(||D_t(\hat{\theta}_n)|| + \frac{C_1\rho^t}{\omega} \left(1 + ||D_t(\hat{\theta}_n)||\right)\right)^d$$

$$\overset{a.s.}{\leq} \frac{1}{n} \sum_{t=1}^{n} \left(U_t + \frac{C_1\rho^t}{\omega}(1 + U_t)\right)^d \leq 2^d \left(\frac{1}{n} \sum_{t=1}^{n} U_t^d + \frac{C_1^d}{\omega^d} \frac{1}{n} \sum_{t=1}^{n} \{\rho^t(1 + U_t)\}^d\right).$$

The uniform ergodic theorem and Assumption 9(ii) imply $\frac{1}{n} \sum_{t=1}^{n} U_t^d \overset{a.s.}{\to} \mathbb{E}[U_t^d] < \infty$. Further, (A.4) leads to $\rho^t(1 + U_t) \overset{a.s.}{\to} 0$ as $t \to \infty$, which in turn implies $\{\rho^t(1 + U_t)\}^d \overset{a.s.}{\to} 0$ as $t \to \infty$. Cesàro’s lemma yields $\frac{1}{n} \sum_{t=1}^{n} \{\rho^t(1 + U_t)\}^d \overset{a.s.}{\to} 0$ and we have $\lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} ||\hat{D}_t||^d < \infty$ almost surely. Combining results establishes $\sum_{t=1}^{n} \mathbb{E}^*[|Z_{n,t}|^d] \overset{a.s.}{\to} 0$. The Central Limit Theorem for triangular arrays (c.f. 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, which completes the proof. \hfill $\square$

**Lemma 8.** Suppose Assumptions 1–9 hold with $a = \pm 6, b = 6, c = 2$ and $d = 4$. Then, we have $I_n^*(z) \overset{p}{\to} \frac{z^2}{2} f(\xi_\alpha)$ in probability.

**Proof.** Using Lemma 3 and Remark 5 the conditional expectation is equal to

$$\mathbb{E}^*[I_n^*(z)] = \sum_{t=1}^{n} \int_{0}^{z/\sqrt{n}} \mathbb{E}^*[\mathbb{I}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{I}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}}] \, ds$$

$$= n \int_{0}^{z/\sqrt{n}} \left(\mathbb{P}_n(\hat{\xi}_{n,\alpha} + s) - \mathbb{P}_n(\hat{\xi}_{n,\alpha} -)\right) \, ds$$

$$= \int_{0}^{z} \sqrt{n} \left(\mathbb{P}_n(\hat{\xi}_{n,\alpha} + u/\sqrt{n}) - \mathbb{P}_n(\hat{\xi}_{n,\alpha} -)\right) \, du$$

$$\overset{p}{\to} \int_{0}^{z} uf(\xi_\alpha) \, du = \frac{z^2}{2} f(\xi_\alpha).$$

85
The conditional variance vanishes in probability as

\[
\text{Var}^*[I_n^*(z)] = \sum_{t=1}^{n} \text{Var}^* \left[ \int_0^{z/\sqrt{n}} \left( \mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) ds \right] \\
= n \text{Var}^* \left[ \int_0^{z/\sqrt{n}} \left( \mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) ds \right] \\
\leq n \frac{|z|}{\sqrt{n}} \mathbb{E}^* \left[ \int_0^{z/\sqrt{n}} \left( \mathbb{1}_{\{\eta_t^* \leq \hat{\xi}_{n,\alpha} + s\}} - \mathbb{1}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) ds \right] \\
= \frac{|z|}{\sqrt{n}} \int_0^{z/\sqrt{n}} n \left( \hat{F}_n(\hat{\xi}_{n,\alpha} + s) - \hat{F}_n(\hat{\xi}_{n,\alpha} - s) \right) ds \\
= \frac{|z|}{\sqrt{n}} \mathbb{E}^*[I_n^*(z)] \overset{p}{\to} 0,
\]

where the inequality follows from the fact that

\[\text{Var}(Y) \leq |c| \mathbb{E}[Y] \quad (A.54)\]

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

**Lemma 9.** Suppose Assumptions 1–10 hold with \( a = \pm 12 \), \( b = 12 \), \( c = 6 \) and \( d > 2 \). Then, we have \( J_{n,1}^*(z) = \frac{1}{2} \hat{\xi}_{n,\alpha} f(\hat{\xi}_{n,\alpha})(\hat{\theta}_n^* - \hat{\theta}_n)' \tilde{J}(\hat{\theta}_n^* - \hat{\theta}_n) \overset{op}{\to} (1) \) in probability.

**Proof.** We set \( \hat{\xi}_{n,\alpha}(z) = \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \) and define

\[
T_n^*(z, u) = \sum_{t=1}^{n} \tau_{n,t}^*(z, u) \\
\tau_{n,t}^*(z, u) = \int_0^{(1-\tilde{\lambda}_t^{-1}(u))\eta_t^*} (\mathbb{1}_{\{\eta_t^* - \hat{\xi}_{n,\alpha}(z) \leq v\}} - \mathbb{1}_{\{\eta_t^* - \hat{\xi}_{n,\alpha}(z) < 0\}}) dv \\
\tilde{\lambda}_t(u) = \frac{\tilde{\sigma}_t(\hat{\theta}_n + n^{-1/2}u)}{\tilde{\sigma}_t(\hat{\theta}_n)}.
\]
Employing the elementary equality

\[ \int_0^v \left( \mathbb{1}_{u \leq s} - \mathbb{1}_{u < 0} \right) ds = (u - v) \left( \mathbb{1}_{v \leq u < 0} - \mathbb{1}_{0 \leq u < v} \right), \]  

(A.55)

we can decompose \( \tau_t^*(z, u) \) into two non-negative components:

\[ \tau_t^*(z, u) = \tau_{t,1}^*(z, u) + \tau_{t,2}^*(z, u) \]

\[ \tau_{t,1}^*(z, u) = \left( \lambda_t^{-1}(u) \eta_t^* - \tilde{\xi}_{n,\alpha}(z) \right) \mathbb{1}_{\{ \tilde{\lambda}_t(u) \tilde{\xi}_{n,\alpha}(z) \leq \eta_t^* < \tilde{\xi}_{n,\alpha}(z) \}} \]

\[ \tau_{t,2}^*(z, u) = \left( \tilde{\lambda}_t^{-1}(u) \eta_t^* - \lambda_t^{-1}(u) \eta_t \right) \mathbb{1}_{\{ \tilde{\xi}_{n,\alpha}(z) \leq \eta_t^* < \lambda_t(u) \tilde{\xi}_{n,\alpha}(z) \}}. \]

Let \( A > 0 \); We establish the result in three steps:

**Step 1:**

\[ T_{n,1}^*(z, u) = \sum_{t=1}^n \tau_{t,1}^*(z, u) \xrightarrow{p^*} \frac{1}{2} \xi^2 f(\xi) \mathbb{E} \left[ \mathbb{1}_{\{ D^*_t'u > 0 \}} u'D_tD'_u \right] \]

\[ T_{n,2}^*(z, u) = \sum_{t=1}^n \tau_{t,2}^*(z, u) \xrightarrow{p^*} \frac{1}{2} \xi^2 f(\xi) \mathbb{E} \left[ \mathbb{1}_{\{ D^*_t'u < 0 \}} u'D_tD'_u \right] \]

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

**Step 2:**

\[ \sup_{\|u\| \leq A} \left| T_{n}^*(z, u) - \frac{1}{2} \xi^2 f(\xi) u'J u \right| \xrightarrow{p^*} 0 \] in probability for all \( z \in \mathbb{R} \);

**Step 3:**

\[ J_{n,1}^*(z) = \frac{1}{2} \xi^2 f(\xi) (\hat{\theta}_n^* - \hat{\theta}_n)'J(\hat{\theta}_n^* - \hat{\theta}_n) + o_p(1) \] in probability.

For **Step 1** we fix \((z, u)\). To lighten notation we therefore drop the \( u \) and \( z \) arguments,
e.g. $\bar{\xi}_{n,\alpha}$ and $\tilde{\lambda}_t$. Employing (A.8) we find that the conditional mean of $\tau^*_{t,1}$ is

\[
\mathbb{E}^*[\tau^*_{t,1}] = \mathbb{1}_{\{\tilde{\lambda}_t \bar{\xi}_{n,\alpha} < \bar{\xi}_{n,\alpha}\}} \mathbb{E}^*[\left(\tilde{\lambda}_t^{-1}\eta^*_t - \bar{\xi}_{n,\alpha}\right) \mathbb{1}_{\{\tilde{\lambda}_t \bar{\xi}_{n,\alpha} \leq \eta^*_t < \bar{\xi}_{n,\alpha}\}]
\]

\[
= \mathbb{1}_{\{\tilde{\lambda}_t \bar{\xi}_{n,\alpha} < \bar{\xi}_{n,\alpha}\}} \left(\tilde{\lambda}_t^{-1} \int_{[\bar{\lambda}_t \bar{\xi}_{n,\alpha} \bar{\xi}_{n,\alpha}]} xd\hat{F}_n(x) - \bar{\xi}_{n,\alpha} \int_{[\bar{\lambda}_t \bar{\xi}_{n,\alpha} \bar{\xi}_{n,\alpha}]} d\hat{F}_n(x)\right)
\]

\[
= \mathbb{1}_{\{\tilde{\lambda}_t \bar{\xi}_{n,\alpha} < \bar{\xi}_{n,\alpha}\}} \left(\tilde{\lambda}_t^{-1} \left(\tilde{\xi}_{n,\alpha} \tilde{F}_n(\tilde{\xi}_{n,\alpha}) - \tilde{\lambda}_t \tilde{\xi}_{n,\alpha} \tilde{F}_n(\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} - ) - \int_{[\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} \tilde{\xi}_{n,\alpha}]} \hat{F}_n(x)dx\right)
\]

\[
- \bar{\xi}_{n,\alpha} \left(\tilde{F}_n(\tilde{\xi}_{n,\alpha} - ) - \tilde{F}_n(\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} - )\right)\right]
\]

\[
= \mathbb{1}_{\{\tilde{\lambda}_t \bar{\xi}_{n,\alpha} < \bar{\xi}_{n,\alpha}\}} \tilde{\lambda}_t^{-1} \left(\tilde{\xi}_{n,\alpha} \tilde{F}_n(\tilde{\xi}_{n,\alpha} - ) (1 - \tilde{\lambda}_t) - \int_{[\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} \tilde{\xi}_{n,\alpha}]} \hat{F}_n(x)dx\right)
\]

\[
= \mathbb{1}_{\{\tilde{\lambda}_t \bar{\xi}_{n,\alpha} < \bar{\xi}_{n,\alpha}\}} \tilde{\lambda}_t^{-1} \int_{[\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} \tilde{\xi}_{n,\alpha}]} \left(\tilde{F}_n(\tilde{\xi}_{n,\alpha} - ) - \hat{F}_n(x)\right)dx.
\]

whereas the conditional mean of $\tau^*_{t,2}$ is given by

\[
\mathbb{E}^*[\tau^*_{t,2}] = \mathbb{1}_{\{\tilde{\xi}_{n,\alpha} < \tilde{\lambda}_t \bar{\xi}_{n,\alpha}\}} \mathbb{E}^*[\left(\tilde{\xi}_{n,\alpha} - \tilde{\lambda}_t^{-1}\eta^*_t\right) \mathbb{1}_{\{\tilde{\xi}_{n,\alpha} \leq \eta^*_t < \bar{\lambda}_t \bar{\xi}_{n,\alpha}\}]
\]

\[
= \mathbb{1}_{\{\tilde{\xi}_{n,\alpha} < \tilde{\lambda}_t \bar{\xi}_{n,\alpha}\}} \left(\tilde{\xi}_{n,\alpha} \int_{[\tilde{\xi}_{n,\alpha} \tilde{\lambda}_t \tilde{\xi}_{n,\alpha}]} d\hat{F}_n(x) - \tilde{\lambda}_t^{-1} \int_{[\tilde{\xi}_{n,\alpha} \tilde{\lambda}_t \tilde{\xi}_{n,\alpha}]} xd\hat{F}_n(x)\right)
\]

\[
= \mathbb{1}_{\{\tilde{\xi}_{n,\alpha} < \tilde{\lambda}_t \bar{\xi}_{n,\alpha}\}} \left(\tilde{\xi}_{n,\alpha} \hat{F}_n(\tilde{\xi}_{n,\alpha} - ) - \tilde{\lambda}_t \tilde{\xi}_{n,\alpha} \hat{F}_n(\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} - ) - \int_{[\tilde{\xi}_{n,\alpha} \tilde{\lambda}_t \tilde{\xi}_{n,\alpha}]} \hat{F}_n(x)dx\right)
\]

\[
- \tilde{\lambda}_t^{-1} \left(\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} \hat{F}_n(\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} - ) - \tilde{\xi}_{n,\alpha} \hat{F}_n(\tilde{\xi}_{n,\alpha} - ) - \int_{[\tilde{\xi}_{n,\alpha} \tilde{\lambda}_t \tilde{\xi}_{n,\alpha}]} \hat{F}_n(x)dx\right)\right]
\]

\[
= \mathbb{1}_{\{\tilde{\xi}_{n,\alpha} < \tilde{\lambda}_t \bar{\xi}_{n,\alpha}\}} \tilde{\lambda}_t^{-1} \left(\tilde{\xi}_{n,\alpha} \hat{F}_n(\tilde{\xi}_{n,\alpha} - ) (1 - \tilde{\lambda}_t) + \int_{[\tilde{\xi}_{n,\alpha} \tilde{\lambda}_t \tilde{\xi}_{n,\alpha}]} \hat{F}_n(x)dx\right)
\]

\[
= \mathbb{1}_{\{\tilde{\xi}_{n,\alpha} < \tilde{\lambda}_t \bar{\xi}_{n,\alpha}\}} \tilde{\lambda}_t^{-1} \int_{[\tilde{\xi}_{n,\alpha} \tilde{\lambda}_t \tilde{\xi}_{n,\alpha}]} \left(\hat{F}_n(x) - \hat{F}_n(\tilde{\xi}_{n,\alpha} - )\right)dx.
\]
Further, employing \((A.3)\) we get
\[
\sup_{\|u\| \leq A} \sqrt{n} \left| \hat{\lambda}_t(u) - 1 \right| = \sup_{\|u\| \leq A} \sqrt{n} \left| \frac{\tilde{\sigma}_t(\hat{\theta}_n + n^{-1/2}u)}{\tilde{\sigma}_t(\hat{\theta}_n)} - 1 \right|
\]
\[
= \sup_{\|u\| \leq A} \sqrt{n} \left| \frac{1}{\tilde{\sigma}_t(\hat{\theta})} \tilde{\sigma}_t(\hat{\theta}) u \right| = \sup_{\|u\| \leq A} \left| \tilde{D}_t(\hat{\theta})u \right|
\]
\[
\leq \sup_{\|u\| \leq A} \|u\| \left( \|D_t(\hat{\theta})\| + \frac{C_1 \rho^f}{\omega} \left( 1 + \|D_t(\hat{\theta})\| \right) \right) \overset{a.s.}{\leq} A \left( U_t + \frac{C_1 \rho^f}{\omega} (1 + U_t) \right)
\]

with \(\hat{\theta}\) lying between \(\hat{\theta}_n\) and \(\hat{\theta}_n + n^{-1/2}u\). Thus, we have
\[
\sup_{\|u\| \leq A} \sqrt{n} \left( \hat{\lambda}_t(u) - 1 \right) = O_p(1). \quad (A.56)
\]

Similarly, we find
\[
\sup_{\|u\| \leq A} \sqrt{n} \left( \hat{\lambda}_t^{-1}(u) - 1 \right) = O_p(1). \quad (A.57)
\]

In addition, for any \(\theta_1, \theta_2 \in \Theta\) Assumption \([4.3]\) yields
\[
\left| \frac{\tilde{\sigma}_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} - \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right| = \left| \frac{\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \frac{\sigma_t(\theta_2) - \tilde{\sigma}_t(\theta_2)}{\tilde{\sigma}_t(\theta_2)} \right|
\]
\[
\leq \left| \frac{\tilde{\sigma}_t(\theta_1) - \sigma_t(\theta_1)}{\tilde{\sigma}_t(\theta_2)} \right| + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \left| \frac{\sigma_t(\theta_2) - \tilde{\sigma}_t(\theta_2)}{\tilde{\sigma}_t(\theta_2)} \right|
\]
\[
\leq \frac{C_1 \rho^f}{\omega} + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \frac{C_1 \rho^f}{\omega} = \frac{C_1 \rho^f}{\omega} \left( 1 + \frac{\sigma_t(\theta_1)}{\sigma_t(\theta_2)} \right) \quad (A.58)
\]
and together with (A.7), we obtain

\[ \sup_{||u|| \leq A} \left| \sqrt{n}(\tilde{\lambda}_t(u) - 1) - \tilde{D}_t' u \right| = O_p(n^{-1/2}). \] (A.59)
Using equations (A.56), (A.57) and (A.59) as well as Lemma 3, we find

\[ \mathbb{E}^*[T^*_{n,1}] = \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} < \tilde{\xi}_{n,\alpha}\}} \tilde{\lambda}_t^{-1} \int_{[\tilde{\lambda}_t^{-1}\tilde{\xi}_{n,\alpha},0]} \left( \hat{F}_n(\xi_{n,\alpha} - ) - \hat{F}_n(\xi_{n,\alpha} + x) \right) dx \]

\[ = \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} < \tilde{\xi}_{n,\alpha}\}} \tilde{\lambda}_t^{-1} \int_{\sqrt{n}(\tilde{\lambda}_t^{-1}\tilde{\xi}_{n,\alpha},0)} \sqrt{n} \left( \hat{F}_n(\xi_{n,\alpha} + \frac{z}{\sqrt{n}} - ) - \hat{F}_n(\xi_{n,\alpha} + \frac{z + y}{\sqrt{n}}) \right) dy \]

\[ = \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} < \tilde{\xi}_{n,\alpha}\}} \tilde{\lambda}_t^{-1} \int_{\sqrt{n}(\tilde{\lambda}_t^{-1}\tilde{\xi}_{n,\alpha},0)} -f(\xi_{\alpha})y dy + o_p(1) \]

\[ = \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} < \tilde{\xi}_{n,\alpha}\}} \tilde{\lambda}_t^{-1} \frac{1}{2} n(\tilde{\lambda}_t - 1)^2 \xi_n^2 f(\xi_{\alpha}) + o_p(1) \]

\[ = \frac{1}{2} \xi_n^2 f(\xi_{\alpha}) \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\lambda}_t \tilde{\xi}_{n,\alpha} < \tilde{\xi}_{n,\alpha}\}} \tilde{\lambda}_t^{-1} (\tilde{\lambda}_t - 1)^2 + o_p(1) \]

\[ = \frac{1}{2} \xi_n^2 f(\xi_{\alpha}) \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\lambda}_t > 1\}} (\tilde{\lambda}_t - 1)^2 + o_p(1) \]

\[ = \frac{1}{2} \xi_n^2 f(\xi_{\alpha}) \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{D}_t u > 0\}} u'D_t \tilde{D}_t u + o_p(1) \]

\[ \rightarrow \frac{1}{2} \xi_n^2 f(\xi_{\alpha}) \mathbb{E} \left[ \mathbb{I}_{\{D_t u > 0\}} u'D_t \tilde{D}_t u \right] \]

91
and similarly, we obtain

\[
\mathbb{E}^* \left[ T_{n,2}^* \right] = \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\xi}_{n,\alpha}, \tilde{\lambda}_{t,\alpha} \}} \lambda_t^{1-1} \int_{[0,1]} \left( \hat{F}_n (\xi_{n,\alpha} + x) - \hat{F}_n (\xi_{n,\alpha} - x) \right) dx
\]

\[
= \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\xi}_{n,\alpha}, \tilde{\lambda}_{t,\alpha} \}} \lambda_t^{1-1} \int_{[0,1]} \sqrt{n} \left( \hat{F}_n (\xi_{n,\alpha} + \frac{z}{\sqrt{n}}) - \hat{F}_n (\xi_{n,\alpha} - \frac{z}{\sqrt{n}}) \right) dy
\]

\[
= \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\xi}_{n,\alpha}, \tilde{\lambda}_{t,\alpha} \}} \lambda_t^{1-1} \int_{[0,1]} f (\xi) y dy + o_p(1)
\]

\[
= \frac{1}{n} \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\xi}_{n,\alpha}, \tilde{\lambda}_{t,\alpha} \}} \lambda_t^{1-1} \frac{1}{2} (\lambda_t - 1)^2 \mathbb{E} \left[ V_{\alpha}^{\lambda_2} \right] + o_p(1)
\]

\[
= \frac{1}{2} \xi_2^2 \mathbb{E} \left[ V_{\alpha}^{\lambda_2} \right] \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\xi}_{n,\alpha}, \tilde{\lambda}_{t,\alpha} \}} \lambda_t^{1-1} (\lambda_t - 1)^2 + o_p(1)
\]

\[
= \frac{1}{2} \xi_2^2 \mathbb{E} \left[ V_{\alpha}^{\lambda_2} \right] \sum_{t=1}^{n} \mathbb{I}_{\{\tilde{\xi}_{n,\alpha}, \tilde{\lambda}_{t,\alpha} \}} \lambda_t^{1-1} (\lambda_t - 1)^2 + o_p(1)
\]

Further, notice that

\[
\text{Var}^* \left[ T_{n,1}^* \right] = \sum_{t=1}^{n} \text{Var}^* \left[ \tau_{t,1}^* \right] = \sum_{t=1}^{n} \left( \text{Var}^* \left[ \tau_{t,1}^* \right] + \text{Var}^* \left[ \tau_{t,2}^* \right] + 2 \text{Cov}^* \left[ \tau_{t,1}^*, \tau_{t,2}^* \right] \right) + o_p(1)
\]

\[
= \sum_{t=1}^{n} \text{Var}^* \left[ \tau_{t,1}^* \right] + \sum_{t=1}^{n} \text{Var}^* \left[ \tau_{t,2}^* \right] = \text{Var}^* \left[ T_{n,1}^* \right] + \text{Var}^* \left[ T_{n,2}^* \right]
\]

as \( \tau_{t,1}^* \tau_{t,2}^* = 0 \) and \( \mathbb{E}^* \left[ \tau_{t,1}^* \right] \mathbb{E}^* \left[ \tau_{t,2}^* \right] = 0 \). Thus, to complete Step 1 it is sufficient to show that the conditional variance of \( T_{n,1}^* \) vanishes in probability. Since

\[
\lambda_t \tau_t^* = \left( (\eta_t^* - \bar{\xi}_{n,\alpha}) - (\bar{\lambda}_t - 1) \right) \left( \mathbb{I}_{\{\tilde{\lambda}_{t,\alpha} \leq \eta_t^* - \bar{\xi}_{n,\alpha} < 0\} - \mathbb{I}_{\{0 \leq \eta_t^* - \bar{\xi}_{n,\alpha} < (\bar{\lambda}_t - 1) \tilde{\xi}_{n,\alpha}\}} \right)
\]
we employ \((A.54)\) and find

$$\text{Var}^*[\tau^*_t] = \lambda_t^{-2} \text{Var}^*[\tilde{\lambda}_t \tau^*_t] \leq \lambda_t^{-2} |(\tilde{\lambda}_t - 1)\xi_{n,\alpha}| \text{E}^*[\tilde{\lambda}_t \tau^*_t] \leq \lambda_t^{-1} |\tilde{\lambda}_t - 1| |\xi_{n,\alpha}| \text{E}^*[\tau^*_t] .$$

Hence, we obtain

$$\text{Var}^*[T^*_n] = \sum_{t=1}^{n} \text{Var}^*[\tau^*_t] \leq \sum_{t=1}^{n} \lambda_t^{-1} |\tilde{\lambda}_t - 1| |\xi_{n,\alpha}| \left( \text{E}^*[\tau^*_1] + \text{E}^*[\tau^*_2] \right)$$

$$\leq \sum_{t=1}^{n} \lambda_t^{-1} |\tilde{\lambda}_t - 1| |\xi_{n,\alpha}| \left( \tilde{\lambda}_t \int_{[\tilde{\lambda}_t \xi_{n,\alpha}]} \left( \hat{\lambda}_n(\xi_{n,\alpha} -) - \hat{\lambda}_n(x) \right) dx \right)$$

$$+ \sum_{t=1}^{n} \lambda_t^{-1} |\tilde{\lambda}_t - 1| |\xi_{n,\alpha}| \left( \tilde{\lambda}_t \int_{[\xi_{n,\alpha} \tilde{\lambda}_t \xi_{n,\alpha}]} \left( \hat{\lambda}_n(x) - \hat{\lambda}_n(\xi_{n,\alpha} -) \right) dx \right)$$

$$= \frac{1}{n} \sum_{t=1}^{n} \lambda_t^{-2} |\tilde{\lambda}_t - 1| |\xi_{n,\alpha}| \left[ (\tilde{\lambda}_t - 1)^2 \xi_{n,\alpha} f(\xi_{n,\alpha}) + o_p(1) \right]$$

$$+ \frac{1}{n} \sum_{t=1}^{n} \lambda_t^{-2} |\tilde{\lambda}_t - 1| \left[ (\tilde{\lambda}_t - 1)^2 \xi_{n,\alpha} f(\xi_{n,\alpha}) + o_p(1) \right]$$

$$= \frac{1}{2} \sum_{t=1}^{n} \lambda_t^{-2} |\tilde{\lambda}_t - 1|^3 |\xi_{n,\alpha}|^3 f(\xi_{n,\alpha}) + o_p(1) = \frac{1}{2} |\xi_{n,\alpha}|^3 f(\xi_{n,\alpha}) \sum_{t=1}^{n} \lambda_t^{-2} |\tilde{\lambda}_t - 1|^3 + o_p(1)$$

and we conclude that \( \text{Var}^*[T^*_n] \xrightarrow{n \to \infty} 0 \), which completes Step 1.

Regarding Step 2 the triangle inequality yields

$$\sup_{||u|| \leq A} \left| T_n^*(z, u) - \text{plim}_{n \to \infty} T_n^*(z, u) \right| \leq \sup_{||u|| \leq A} \left| T_{n,1}^*(z, u) - \text{plim}_{n \to \infty} T_{n,1}^*(z, u) \right|$$

$$+ \sup_{||u|| \leq A} \left| T_{n,2}^*(z, u) - \text{plim}_{n \to \infty} T_{n,2}^*(z, u) \right| .$$

(A.60)

Let \( N \geq 1 \) be an integer. Analogously to the third step in Lemma 3, we divide the (hyper-)cube \([-A, A]^p\) into \( L = (2N)^p \) (hyper-)cubes with side length \( A/N \). Again, in
case of a (hyper-)cube \( \ell \), \( u_\bullet(\ell) \) and \( u^*(\ell) \) denote the lower left and upper right vertex of \( \ell \). Given \( \ell \in \{1, \ldots, L\} \) and \( u \) satisfying \( u_\bullet(\ell) \leq u \leq u^*(\ell) \) (element-by-element comparison), Assumption 8 implies \( \tilde{\lambda}_\ell(u_\bullet(\ell)) \leq \tilde{\lambda}_\ell(u) \leq \tilde{\lambda}_\ell(u^*(\ell)) \). Further, given \( z \), Theorem 1 results in \( \bar{\xi}_{n,\alpha}(z) \xrightarrow{a.s.} \xi_{\alpha} < 0 \). Thus, we have

\[
T_{n,1}^*(z, u_\bullet(\ell)) \leq T_{n,1}^*(z, u) \leq T_{n,1}^*(z, u^*(\ell))
\]

\[
T_{n,2}^*(z, u^*(\ell)) \leq T_{n,2}^*(z, u) \leq T_{n,2}^*(z, u_\bullet(\ell))
\]

almost surely. Let \( k \in \{1, 2\} \); we obtain

\[
\sup_{||u|| \leq A} \left| T_{n,k}^*(z, u) - \lim_{n \to \infty} T_{n,k}^*(z, u) \right| \leq \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^*(\ell)) - \lim_{n \to \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right|
\]

\[
+ \max_{1 \leq \ell \leq L} \sup_{u_\bullet(\ell) \leq u \leq u^*(\ell)} \left| T_{n,k}^*(z, u^*(\ell)) - T_{n,k}^*(z, u) \right|
\]

\[
+ \max_{1 \leq \ell \leq L} \sup_{u_\bullet(\ell) \leq u \leq u^*(\ell)} \left| \lim_{n \to \infty} \left( T_{n,k}^*(z, u^*(\ell)) - T_{n,k}^*(z, u) \right) \right|
\]

with

\[
A_n \leq \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^*(\ell)) - T_{n,k}^*(u_\bullet(\ell)) \right|
\]

\[
\leq \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^*(\ell)) - \lim_{n \to \infty} T_{n,k}^*(z, u^*(\ell)) \right|
\]

\[
+ \max_{1 \leq \ell \leq L} \left| T_{n,k}^*(z, u^*(\ell)) - \lim_{n \to \infty} T_{n,k}^*(z, u_\bullet(\ell)) \right|
\]

\[
+ \max_{1 \leq \ell \leq L} \left| \lim_{n \to \infty} \left( T_{n,k}^*(z, u^*(\ell)) - T_{n,k}^*(z, u_\bullet(\ell)) \right) \right|
\]

\[
B_n \leq \max_{1 \leq \ell \leq L} \left| \lim_{n \to \infty} \left( T_{n,k}^*(z, u^*(\ell)) - T_{n,k}^*(z, u_\bullet(\ell)) \right) \right|
\]
Hence, we establish the following bound

\[
\sup_{\|u\| \leq A} \left| T_{n,k}^* (z, u) - \lim_{n \to \infty} T_{n,k}^* (z, u) \right| \leq 2I + II + III
\]

with

\[
I = \max_{1 \leq \ell \leq L} \left| T_{n,k}^* (z, u^* (\ell)) - \lim_{n \to \infty} T_{n,k}^* (z, u^* (\ell)) \right|
\]
\[
II = \max_{1 \leq \ell \leq L} \left| T_{n,k}^* (z, u_\bullet (\ell)) - \lim_{n \to \infty} T_{n,k}^* (z, u_\bullet (\ell)) \right|
\]
\[
III = \max_{1 \leq \ell \leq L} \left| \lim_{n \to \infty} T_{n,k}^* (z, u^* (\ell)) - \lim_{n \to \infty} T_{n,k}^* (z, u_\bullet (\ell)) \right|
\]

Regarding \(III\), we have for every \(u\) satisfying \(\|u\| \leq A\) that

\[
\lim_{n \to \infty} T_{n,k}^* (z, u) = \begin{cases} \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E} \left[ \mathbb{1}_{\{D_i u > 0\}} u' D_i D_i' u \right] & \text{if } k = 1 \\ \frac{1}{2} \xi_\alpha^2 f(\xi_\alpha) \mathbb{E} \left[ \mathbb{1}_{\{D_i u < 0\}} u' D_i D_i' u \right] & \text{if } k = 2 \end{cases}
\]

is continuous in \(u\). Together with \(\|u^* (\ell) - u_\bullet (\ell)\| \leq \frac{A}{N}\) for every \(\ell \in \{1, \ldots, L\}\), it follows that \(III\) can be made arbitrarily small by choosing \(N\) sufficiently large.

Given \(N\) (and \(L\)), \(I \xrightarrow{p} 0\) in probability and \(II \xrightarrow{p} 0\) in probability by Step 1, which completes Step 2.

Consider Step 3; for each \(\epsilon > 0\) we obtain

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

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
the Step 2, which completes Step 3 and 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 Zakoian (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 \( \{ \hat{\theta}_t^* - \hat{\theta}_n, \eta_t^* : u < t \} \) and \( F_n \) has, almost surely, a single point mass.

**Lemma 10.** Suppose Assumptions 1–10 with \( a = \pm 12, b = 12, c = 6 \) and \( d > 2 \). Then, we have

\[ J_{n,2}^* = z_{\xi_\alpha} f(\xi_\alpha) \sqrt{n} (\hat{\theta}_n^* - \hat{\theta}_n) + o_p(1) \]

in probability.

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

\[ J_{n,2}^*(z) = \sum_{t=1}^{n} \left( 1 - \frac{\hat{\sigma}_t(\hat{\theta}_n)}{\hat{\sigma}_t(\hat{\theta}_n^*)} \eta_t^* \left( \mathbb{I}_{\{\eta_t^* < \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}\}} - \mathbb{I}_{\{\eta_t^* < \hat{\xi}_{n,\alpha}\}} \right) \right). \]  

(A.61)

A Taylor expansion around \( \hat{\theta}_n \) yields

\[ 1 - \frac{\hat{\sigma}_t(\hat{\theta}_n)}{\hat{\sigma}_t(\hat{\theta}_n^*)} = 1 - \frac{\hat{\sigma}_t(\hat{\theta}_n)}{\hat{\sigma}_t(\hat{\theta}_n^*)} (\hat{\theta}_n^* - \hat{\theta}_n) \]

\[ + \frac{1}{2} (\hat{\theta}_n^* - \hat{\theta}_n) \frac{\hat{\sigma}_t(\hat{\theta}_n)}{\hat{\sigma}_t(\hat{\theta}_n^*)} \left( \frac{1}{\hat{\sigma}_t(\hat{\theta}_n)} \frac{\partial^2 \hat{\sigma}_t(\hat{\theta}_n)}{\partial \theta \partial \theta} - \frac{2}{\hat{\sigma}_t^2(\hat{\theta}_n)} \frac{\partial \hat{\sigma}_t(\hat{\theta}_n)}{\partial \theta} \frac{\partial \hat{\sigma}_t(\hat{\theta}_n)}{\partial \theta} \right) (\hat{\theta}_n^* - \hat{\theta}_n) \]

\[ = \hat{D}_t^* (\hat{\theta}_n^* - \hat{\theta}_n) + \frac{1}{2} (\hat{\theta}_n^* - \hat{\theta}_n) \frac{\hat{\sigma}_t(\hat{\theta}_n)}{\hat{\sigma}_t(\hat{\theta}_n^*)} \left( \tilde{H}_t(\hat{\theta}_n^*) - 2 \hat{D}_t(\hat{\theta}_n^*) \hat{D}_t^*(\hat{\theta}_n^*) \right) (\hat{\theta}_n^* - \hat{\theta}_n). \]
where \( \theta \) lies between \( \hat{\theta}_n^* \) and \( \hat{\theta}_n \). Plugging this result into (A.61) gives

\[
J_{n,2}^* = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} J_{n,t}^{*2} \tilde{D}_t \sqrt{n(\hat{\theta}_n^* - \hat{\theta}_n)} + \frac{1}{2} \sqrt{n(\hat{\theta}_n^* - \hat{\theta}_n)} \frac{1}{n} \sum_{t=1}^{n} \tilde{\sigma}_t(\hat{\theta}_n) \left( \tilde{H}_t(\hat{\theta}) - 2\tilde{D}_t(\hat{\theta})\tilde{D}'_t(\hat{\theta}) \right) J_{n,t}^{*2} \sqrt{n(\hat{\theta}_n^* - \hat{\theta}_n)}.
\]

With regard to Proposition 1, it suffices to show that

\[
I_p^* \xrightarrow{p} \xi_\alpha z f(\xi_\alpha)\Omega' \text{ in probability and } II_p^* \xrightarrow{p} 0 \text{ in probability.}
\]

The conditional mean and variance of the first term are

\[
\mathbb{E}^*[I] = \sqrt{n} \mathbb{E}^*[\tilde{J}_{n,t}^{*2}] \frac{1}{n} \sum_{t=1}^{n} \tilde{D}_t' = \sqrt{n} \mathbb{E}^*[\tilde{J}_{n,t}^{*2}] \tilde{\Omega}_n'
\]

and

\[
\text{Var}^*[I] = \text{Var}^*[\tilde{J}_{n,t}^{*2}] \frac{1}{n} \sum_{t=1}^{n} \tilde{D}_t \tilde{D}_t' = \text{Var}^*[\tilde{J}_{n,t}^{*2}] \tilde{J}_n.
\]

Lemma 2 states \( \tilde{\Omega}_n \xrightarrow{a.s.} \Omega \) and \( \tilde{J}_n \xrightarrow{a.s.} J \). Employing (A.8), Theorem 1 and Lemma 3 with Remark 5, the conditional mean of \( \tilde{J}_{n,t}^{*2} \) satisfies

\[
\sqrt{n} \mathbb{E}^*[\tilde{J}_{n,t}^{*2}] = \sqrt{n} \int_{\left[\xi_{n,\alpha}, \xi_{n,\alpha} + \frac{y}{\sqrt{n}}\right]} x d\tilde{F}_n(x)
\]

\[
= \left( \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \right) \sqrt{n} \tilde{F}_n(\hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}}) - \hat{\xi}_{n,\alpha} \sqrt{n} \tilde{F}_n(\hat{\xi}_{n,\alpha}) - \sqrt{n} \int_{\left[\hat{\xi}_{n,\alpha}, \hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}}\right]} \tilde{F}_n(x) dx
\]

\[
= \left( \hat{\xi}_{n,\alpha} + \frac{z}{\sqrt{n}} \right) \left( \sqrt{n} \tilde{F}_n(\hat{\xi}_{n,\alpha}) + z f(\xi_{\alpha}) + o_p(1) \right) - \hat{\xi}_{n,\alpha} \sqrt{n} \tilde{F}_n(\hat{\xi}_{n,\alpha})
\]

\[
- \int_{[0,z]} \tilde{F}_n \left( \hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}} \right) dy
\]

\[
= \hat{\xi}_{n,\alpha} z f(\xi_{\alpha}) + z \tilde{F}_n(\hat{\xi}_{n,\alpha}) - \int_{[0,z]} \tilde{F}_n \left( \hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}} \right) dy + o_p(1)
\]

\[
= \xi_{\alpha} z f(\xi_{\alpha}) - \int_{[0,z]} \left[ \tilde{F}_n \left( \hat{\xi}_{n,\alpha} + \frac{y}{\sqrt{n}} \right) - \tilde{F}_n(\hat{\xi}_{n,\alpha}) \right] dy + o_p(1) = \xi_{\alpha} z f(\xi_{\alpha}) + o_p(1).
\]
Further, we get \( \sqrt{n}E^* \left[ \left( j_{n,t}^{(2)} \right)^2 \right] \xrightarrow{p} |z|^2 f(\xi_\alpha) \). To appreciate why we have for \( z \geq 0 \)

\[ \sqrt{n}E^* \left[ \left( j_{n,t}^{(2)} \right)^2 \right] = \sqrt{n} \int_{\left[ \hat{\xi}_{\alpha,n} \xi_{\alpha,n} + \frac{z}{\sqrt{n}} \right]} x^2 \, d\hat{F}_n(x) \]

\[ = \left( \hat{\xi}_{\alpha,n} + \frac{z}{\sqrt{n}} \right)^2 \sqrt{n} \hat{F}_n \left( \hat{\xi}_{\alpha,n} + \frac{z}{\sqrt{n}} \right) - \hat{\xi}_{\alpha,n}^2 \sqrt{n} \hat{F}_n(\hat{\xi}_{\alpha,n}) \]

\[ - \sqrt{n} \int_{\left[ \hat{\xi}_{\alpha,n} \xi_{\alpha,n} + \frac{z}{\sqrt{n}} \right]} \hat{F}_n(x) \, d(x^2) \]

\[ = \left( \hat{\xi}_{\alpha,n} + \frac{z}{\sqrt{n}} \right)^2 \left( \sqrt{n} \hat{F}_n(\hat{\xi}_{\alpha,n}) + zf(\xi_\alpha) + o_p(1) \right) - \hat{\xi}_{\alpha,n}^2 \sqrt{n} \hat{F}_n(\hat{\xi}_{\alpha,n}) \]

\[ - 2 \sqrt{n} \int_{\left[ \hat{\xi}_{\alpha,n} \xi_{\alpha,n} + \frac{z}{\sqrt{n}} \right]} x \hat{F}_n(x) \, dx \]

\[ = \hat{\xi}_{\alpha,n} z f(\xi_\alpha) + 2 \hat{\xi}_{\alpha,n} z \hat{F}_n(\hat{\xi}_{\alpha,n}) - 2 \int_{[0,z]} \left( \hat{\xi}_{\alpha,n} + \frac{y}{\sqrt{n}} \right) \hat{F}_n \left( \hat{\xi}_{\alpha,n} + \frac{y}{\sqrt{n}} \right) \, dy + o_p(1) \]

\[ = \xi_\alpha z f(\xi_\alpha) - 2 \int_{[0,z]} \left[ \left( \hat{\xi}_{\alpha,n} + \frac{y}{\sqrt{n}} \right) \hat{F}_n \left( \hat{\xi}_{\alpha,n} + \frac{y}{\sqrt{n}} \right) - \hat{\xi}_{\alpha,n} \hat{F}_n(\hat{\xi}_{\alpha,n}) \right] \, dy + o_p(1) \]

\[ = \xi_\alpha z f(\xi_\alpha) + o_p(1) \]
and in the case of $z < 0$ we obtain

$$\sqrt{n} \mathbb{E}^* \left[ (j_{n,t}^{(2)})^2 \right] = \sqrt{n} \int_{[\xi_{n,a} + \frac{z}{\sqrt{n}} \xi_{n,a}]} x^2 \ d\hat{F}_n(x)$$

$$= \xi_{n,a}^2 \sqrt{n} \hat{F}_n(\xi_{n,a} -) - \left( \xi_{n,a} + \frac{z}{\sqrt{n}} \right)^2 \sqrt{n} \hat{F}_n\left( \xi_{n,a} + \frac{z}{\sqrt{n}} - \right)$$

$$- \sqrt{n} \int_{[\xi_{n,a} + \frac{z}{\sqrt{n}} \xi_{n,a}]} \hat{F}_n(x) \ dx$$

$$= \xi_{n,a}^2 \sqrt{n} \hat{F}_n(\xi_{n,a} -) - \left( \xi_{n,a} + \frac{z}{\sqrt{n}} \right)^2 \left( \sqrt{n} \hat{F}_n(\xi_{n,a}) + zf(\xi_a) + o_p(1) \right)$$

$$- 2 \sqrt{n} \int_{[\xi_{n,a} + \frac{z}{\sqrt{n}} \xi_{n,a}]} x \hat{F}_n(x) \ dx$$

$$= - z \xi_{n,a} f(\xi_a) - 2 \xi_{n,a} z \hat{F}_n(\xi_{n,a}) - 2 \int_{[z,0]} \left( \xi_{n,a} + \frac{y}{\sqrt{n}} \right) \hat{F}_n\left( \xi_{n,a} + \frac{y}{\sqrt{n}} - \right) \ dy + o_p(1)$$

$$= - z \xi_{n,a} f(\xi_a) - 2 \int_{[0,z]} \left[ \left( \xi_{n,a} + \frac{y}{\sqrt{n}} \right) \hat{F}_n\left( \xi_{n,a} + \frac{y}{\sqrt{n}} - \right) - \xi_{n,a} \hat{F}_n(\xi_{n,a}) \right] \ dy + o_p(1)$$

$$= - z \xi_{n,a} f(\xi_a) + o_p(1).$$

Hence, $\text{Var}^* [j_{n,t}^{(2)}] \overset{P}{\to} 0$ and combining results we establish $I \overset{P^*}{\to} \xi_{n,a} z f(\xi_a) \Omega'$ in probability. Consider the second term; since $\hat{\theta}_n \overset{a.s.}{\to} \theta_0$ (Theorem 1) and $\hat{\theta}_n^* \overset{P^*}{\to} \theta_0$ almost surely (Lemma 5), we have $\mathbb{P}^* \left[ \hat{\theta} \notin \mathcal{V}(\theta_0) \right] \overset{a.s.}{\to} 0$. Thus, for every $\epsilon > 0$ we obtain

$$\mathbb{P}^* \left[ ||I|| \geq \epsilon \right]$$

$$\leq \mathbb{P}^* \left[ \left| \frac{1}{n} \sum_{t=1}^n \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\hat{\theta})} \left( \hat{H}_t(\hat{\theta}) - 2 \hat{D}_t(\hat{\theta}) \tilde{D}_t(\hat{\theta}) \right) j_{n,t}^{(2)} \right| \geq \epsilon \right] + \mathbb{P}^* \left[ \hat{\theta} \notin \mathcal{V}(\theta_0) \right]$$

$$\leq \mathbb{P}^* \left[ \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \tilde{\sigma}_t(\hat{\theta}_n) \left( \sup_{\theta \in \mathcal{V}(\theta_0)} ||\hat{H}_t(\theta)|| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} ||\hat{D}_t(\theta)||^2 \right) j_{n,t}^{(2)} \right] + o(1)$$

$$\leq \frac{1}{\epsilon} \mathbb{E}^* \left[ \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \tilde{\sigma}_t(\hat{\theta}_n) \left( \sup_{\theta \in \mathcal{V}(\theta_0)} ||\hat{H}_t(\theta)|| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} ||\hat{D}_t(\theta)||^2 \right) j_{n,t}^{(2)} \right] + o(1)$$

$$= \frac{1}{\epsilon} \mathbb{E}^* \left[ j_{n,t}^{(2)} \right] \frac{1}{n} \sum_{t=1}^n \sup_{\theta \in \mathcal{V}(\theta_0)} \tilde{\sigma}_t(\hat{\theta}_n) \left( \sup_{\theta \in \mathcal{V}(\theta_0)} ||\hat{H}_t(\theta)|| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} ||\hat{D}_t(\theta)||^2 \right) + o(1)$$

99
almost surely, where the last inequality follows from Markov’s inequality. Because
\[ E^*\left[ |j_{n,t}^{(2)}| \right] \leq E^*\left[ \left| j_{n,t}^{(2)} \right|^2 \right]^{\frac{1}{2}} \overset{p}{\to} 0, \]
it remains to show that
\[ \frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \mathcal{V}(\theta_0)} \tilde{\sigma}_t(\hat{\theta}_n) \left( \sup_{\theta \in \mathcal{V}(\theta_0)} \| H_t(\theta) \| + 2 \sup_{\theta \in \mathcal{V}(\theta_0)} \| \tilde{D}_t(\theta) \|^2 \right) \]
is stochastically bounded. Using (A.7) we find
\[ \sup_{\theta \in \mathcal{V}(\theta_0)} \| H_t(\theta) \| \leq \sup_{\theta \in \mathcal{V}(\theta_0)} \left( \| H_t(\theta) \| + \frac{C_1 \rho^t}{\omega} \left( 1 + \| H_t(\theta) \| \right) \right) \leq V_t + \frac{C_1 \rho^t}{\omega} (1 + V_t). \]

Employing (A.58) we further have
\[ \sup_{\theta \in \mathcal{V}(\theta_0)} \frac{\tilde{\sigma}_t(\hat{\theta}_n)}{\tilde{\sigma}_t(\theta)} \leq \sup_{\theta \in \mathcal{V}(\theta_0)} \left( \frac{\sigma_t(\hat{\theta}_n)}{\sigma_t(\theta)} + \frac{C_1 \rho^t}{\omega} \left( 1 + \frac{\sigma_t(\hat{\theta}_n)}{\sigma_t(\theta)} \right) \right) \leq S_t T_t + \frac{C_1 \rho^t}{\omega} (1 + S_t T_t). \]

In addition, (A.3) and (A.10) imply
\[ \sup_{\theta \in \mathcal{V}(\theta_0)} \| \tilde{D}_t(\theta) \|^2 \leq \sup_{\theta \in \mathcal{V}(\theta_0)} \left( \| D_t(\theta) \| + \frac{C_1 \rho^t}{\omega} \left( 1 + \| D_t(\theta) \| \right) \right)^2 \leq \sup_{\theta \in \mathcal{V}(\theta_0)} 3 \left( \| D_t(\theta) \|^2 + \frac{C_2 \rho^{2t}}{\omega^2} \left( 1 + \| D_t(\theta) \|^2 \right) \right) \leq 3 U_t^2 + \frac{3 C_1 \rho^{2t}}{\omega^2} \left( 1 + U_t^2 \right). \]
Hence,

\[
\frac{1}{n} \sum_{t=1}^{n} \sup_{\theta \in \Psi(\theta_{0})} \frac{\tilde{\sigma}_{t}(\theta)}{\sigma_{t}(\theta)} \left( \sup_{\theta \in \Psi(\theta_{0})} \left| \tilde{H}_{t}(\theta) \right| + 2 \sup_{\theta \in \Psi(\theta_{0})} \left| \tilde{D}_{t}(\theta) \right| \right) \\
\leq \frac{1}{n} \sum_{t=1}^{n} \left( S_{t}T_{t} + \frac{C_{1}\rho^{2}}{\omega} (1 + S_{t}T_{t}) \right) \left( V_{t} + \frac{C_{1}\rho^{2}}{\omega} (1 + V_{t}) + 6U_{t}^{2} + \frac{6C_{2}^{2}\rho^{2t}}{\omega^{2}} (1 + U_{t}^{2}) \right) \\
= \frac{1}{n} \sum_{t=1}^{n} S_{t}T_{t}V_{t} + \frac{6}{n} \sum_{t=1}^{n} S_{t}T_{t}U_{t}^{2} + \frac{C_{1}}{\omega} \frac{1}{n} \sum_{t=1}^{n} \rho^{t}S_{t}T_{t} + \frac{C_{1}}{\omega} \frac{1}{n} \sum_{t=1}^{n} \rho^{t}S_{t}T_{t}V_{t} \\
+ \frac{C_{1}}{\omega} \frac{1}{n} \sum_{t=1}^{n} \rho^{t}V_{t} + \frac{C_{1}^{6}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{2t}S_{t}T_{t} + \frac{C_{1}^{6}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{2t}S_{t}T_{t}V_{t} + \frac{C_{1}}{\omega} \frac{1}{n} \sum_{t=1}^{n} \rho^{t}S_{t}T_{t} \\
+ \frac{C_{1}^{2}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{2t}V_{t} + \frac{C_{1}^{2}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{2t}S_{t}T_{t} + \frac{C_{1}^{2}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{2t}S_{t}T_{t}V_{t} + \frac{C_{1}^{2}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{2t}S_{t}T_{t} \\
+ \frac{C_{1}^{3}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{3t}V_{t}^{2} + \frac{C_{1}^{3}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{3t}S_{t}T_{t} + \frac{C_{1}^{3}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{3t}S_{t}T_{t}V_{t} + \frac{C_{1}^{3}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{3t}S_{t}T_{t}U_{t}^{2} \\
+ \frac{C_{1}^{2}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{2t} + \frac{C_{1}^{6}}{\omega^{2}} \frac{1}{n} \sum_{t=1}^{n} \rho^{3t} \\
\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
\]

From Assumption 3 the uniform ergodic theorem and Hölder’s inequality, we obtain

\[II_{1} \leq \left( \frac{1}{n} \sum_{t=1}^{n} S_{t}^{3} \right)^{\frac{1}{3}} \left( \frac{1}{n} \sum_{t=1}^{n} T_{t}^{3} \right)^{\frac{1}{3}} \left( \frac{1}{n} \sum_{t=1}^{n} V_{t}^{3} \right)^{\frac{1}{3}} \overset{a.s.}{\rightarrow} \left( \mathbb{E}[S_{t}^{3}] \right)^{\frac{1}{3}} \left( \mathbb{E}[T_{t}^{3}] \right)^{\frac{1}{3}} \left( \mathbb{E}[V_{t}^{3}] \right)^{\frac{1}{3}} < \infty\]

and similarly we can show that \( \lim_{n \to \infty} II_{2} < \infty \) 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 \( \frac{1}{n} \sum_{t=1}^{n} \rho^t S_t \xrightarrow{a.s.} 0 \) 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 \( II_4, \ldots, II_{16} \) vanish almost surely. Further, \( II_{17} \leq \frac{1}{n} \frac{C_{17}^t}{\omega(1-\rho^2)} \xrightarrow{a.s.} 0 \) and similarly, we can prove that \( II_{18} \) vanish almost surely, which completes the proof. \( \square \)

## B Recursive-design Bootstrap

This appendix devotes attention to the recursive design bootstrap. The bootstrap schemes described in Algorithms 3 and 4 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 \( \epsilon_{t-1}^*, \ldots, \epsilon_1^* \).

### Algorithm 3. (Recursive-design residual bootstrap)

1. For \( t = 1, \ldots, n \) generate \( \eta_t^* \sim \hat{F}_n \) and the bootstrap observation \( \epsilon_t^* = \sigma_t^* \eta_t^* \) with \( \sigma_t^* = \sigma_t^*(\hat{\theta}_n) \) and \( \sigma_t^*(\theta) = \sigma(\epsilon_{t-1}^*, \ldots, \epsilon_1^*, \hat{\epsilon}_0, \ldots; \theta) \)

2. Calculate the bootstrap estimator

\[
\hat{\theta}_n^* = \arg\max_{\theta \in \Theta} L_n^*(\theta), \quad \text{(B.1)}
\]

with the bootstrap criterion function given by

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

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

\[
\hat{\xi}_{n,\alpha}^* = \arg\min_{z \in \mathbb{R}} \frac{1}{n} \sum_{t=1}^{n} \rho_\alpha(\hat{\eta}_t^* - z). \quad \text{(B.2)}
\]
4. Obtain the bootstrap estimator of the conditional VaR

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

\text{(B.3)}

\textbf{Algorithm 4.} \textit{(Recursive-design Bootstrap Confidence Intervals for VaR)}

1. Acquire a set of \( B \) bootstrap replicates, i.e. \( \hat{\text{VaR}}_{n,\alpha}^{*(b)} \) for \( b = 1, \ldots, B \), by repeating Algorithm [3].

2.1. Obtain the EP interval

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

\text{(B.4)}

with \( \hat{G}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1} \left\{ \sqrt{n} \left| \hat{\text{VaR}}_{n,\alpha}^{*(b)} - \hat{\text{VaR}}_{n,\alpha} \right| \leq x \right\} \).

2.2. Calculate the RT interval

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

\text{(B.5)}

2.3. Compute the SY interval

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

\text{(B.6)}

with \( \hat{H}_{n,B}^*(x) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1} \left\{ \sqrt{n} \left| \hat{\text{VaR}}_{n,\alpha}^{*(b)} - \hat{\text{VaR}}_{n,\alpha} \right| \leq x \right\} \).

In contrast to the fixed-design bootstrap, the bootstrap sample \( \epsilon_1^*, \ldots, \epsilon_n^* \), 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 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.

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. Smeekes (2017). A justification of conditional confidence intervals. Working paper, Maastricht University. http://researchers-sbe.unimaas.nl/stephansmeekes/research/.

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*.

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.

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.

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.

Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.

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.

Gao, F. and F. Song (2008). Estimation risk in GARCH VaR and ES estimates. Econometric Theory 24(5), 1404–1424.

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.

Hidalgo, J. and P. Zaffaroni (2007). A goodness-of-fit test for ARCH(∞) 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.

Koenker, R. and Z. Xiao (2006). Quantile autoregression. Journal of the American Statistical Association 101(475), 980–990.
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.

Shimizu, K. (2009). *Bootstrapping Stationary ARMA-GARCH Models*. Springer.

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.