UNIVERSITY OF MANNHEIM

University of Mannheim / Department of Economics

Working Paper Series

# *A model specification test for GARCH(1,1) processes*

Anne Leucht Michael H. Neumann Jens-Peter Kreiss

Working Paper 13-11

November 2013

# A model specification test for GARCH(1,1) processes

Anne Leucht
Universität Mannheim
Abteilung Volkswirtschaftslehre
L 7, 3-5
D-68131 Mannheim
Germany
E-mail: leucht@uni-mannheim.de

Michael H. Neumann
Friedrich-Schiller-Universität Jena
Institut für Mathematik
Ernst-Abbe-Platz 2
D-07743 Jena
Germany
E-mail: michael.neumann@uni-jena.de

Jens-Peter Kreiss
Technische Universität Braunschweig
Institut für Mathematische Stochastik
Pockelsstraße 14
D-38106 Braunschweig
Germany
E-mail: j.kreiss@tu-braunschweig.de

**Abstract**

We provide a consistent specification test for GARCH(1,1) models based on a test statistic of Cramér-von Mises type. Since the limit distribution of the test statistic under the null hypothesis depends on unknown quantities in a complicated manner, we propose a model-based (semiparametric) bootstrap method to approximate critical values of the test and verify its asymptotic validity. Finally, we illuminate the finite sample behavior of the test by some simulations.

*2010 Mathematics Subject Classification.* Primary 62F03; secondary 62F40, 62F05.
*JEL subject code.* C12.
*Keywords and Phrases.* Bootstrap, Cramér-von Mises test, GARCH processes, $V$-statistic.
*Short title.* A test for GARCH(1,1) models.
version: November 22, 2013

## 1. Introduction

Conditionally heteroscedastic time series are frequently used in the finance literature to model the evolution of stock prizes, exchange rates and interest rates. Starting with the papers by Engle (1982) on autoregressive conditional heteroscedastic models (ARCH) and Bollerslev (1986) on generalized ARCH (GARCH) models, numerous variants of these models have been proposed for financial time series modeling; see e.g. Francq and Zakoïan (2010) for a detailed overview. The question of parameter estimation in these models has been studied intensively. Exemplarily, we refer the reader to Straumann (2005), Francq and Zakoïan (2010), Tinkl (2013) and references therein.

There is also an overwhelming amount of model specification tests in the econometric literature. However, these methods typically rely on the assumption that the information variables as well as the response variables are observable. This condition is violated in the case of GARCH models, where unobserved quantities enter the information variable. Hence, standard tests cannot be applied and certain additional approximation procedures have to be invoked. It turns out that the literature on specification tests for conditionally heteroscedastic time series is comparatively rare. Berkes, Horváth, and Kokoszka (2004) proposed a Portmanteau goodness-of-fit test for GARCH(1,1) models. Their test statistic is a quadratic form of weighted autocorrelations of the squared residuals of a GARCH(1,1) process fitted to the data, whose dimension increases with the sample size. They showed that its limit distribution is an (infinite) weighted sum of independent $\chi_1^2$-distributed random variables under the null hypothesis but did not consider the behavior under alternatives.

In the present paper, we propose a specification test of Cramér-von Mises type for a GARCH(1,1) hypothesis against general alternatives. Here, we face the particular problem that some of the explanatory variables are not observed and have to be approximated. It turns out that our test statistic can be approximated by a von Mises ($V$-) statistic and it follows from results of Leucht and Neumann (2013a) that the latter converges to a weighted sum of independent $\chi_1^2$ variables. In contrast to Berkes, Horváth, and Kokoszka (2004), where the weights in the limit correspond to the weights in the test statistic itself, here these quantities depend on the properties of the underlying process in a complicated way. Therefore, the asymptotic result cannot be used for determining an appropriate critical value. We propose to apply a model-based (semiparametric) bootstrap method to approximate the null distribution of the test statistic which eventually yields an appropriate critical value for the test. Bootstrap consistency for statistics of $L_2$-type has already been shown in several previous papers. Escanciano (2007a, 2008) showed consistency of the wild bootstrap in the context of tests with an underlying martingale structure under the null hypothesis. Leucht and Neumann (2013a,b) proved consistency of model-based bootstrap and a variant of the dependent wild bootstrap, respectively, for statistics that can be approximated by a $V$-statistic. In contrast to the method of proof used in Leucht and Neumann (2013a), we take this opportunity and present a different approach of proving bootstrap consistency: Rather than imitating the derivation of the limit distribution of the test statistic also on the bootstrap side, we use coupling arguments to show consistency. This approach was successfully applied to $U$- and $V$-statistics of independent random variables by Dehling and Mikosch (1994) and Leucht and Neumann (2009), however, it seems to be new in the context of dependent data. Finally, we would like to mention that our theory can perhaps be generalized to GARCH-models of higher order. To present the main ideas in an as transparent as possible manner, we restrict ourselves to the simple GARCH(1,1)-case.

## 2. Assumptions and some preliminaries on GARCH(1,1)-processes

Suppose that we observe $Y_0, \ldots, Y_n$, where $(Y_t)_{t\in\mathbb{Z}}$ is a (strictly) stationary process satisfying the model equation

$$Y_t \,=\, \sigma_t\, \varepsilon_t,$$

where $\sigma_t$ and $\varepsilon_t$ are stochastically independent and $(\varepsilon_t)_t$ is a sequence of independent and identically distributed (i.i.d.) random variables. We consider the test problem

$$\mathcal{H}_0\colon (Y_t)_{t\in\mathbb{Z}} \in \mathcal{M}_0 \qquad \text{against} \qquad \mathcal{H}_1\colon (Y_t)_{t\in\mathbb{Z}} \in \mathcal{M}\backslash\mathcal{M}_0$$

with

$$\begin{aligned}
\mathcal{M}_0 &= \{(Y_t)_{t\in\mathbb{Z}} \mid Y_t = \sigma_t\, \varepsilon_t \text{ with } E(Y_t^2 \mid Y_{t-1}, \sigma_{t-1}^2) = \sigma_t^2 = \omega + \alpha Y_{t-1}^2 + \beta\sigma_{t-1}^2,\ \theta = (\omega, \alpha, \beta)' \in \Theta\},\\
\mathcal{M} &= \{(Y_t)_{t\in\mathbb{Z}} \mid Y_t = \sigma_t\, \varepsilon_t \text{ with } E(Y_t^2 \mid Y_{t-1}, \sigma_{t-1}^2) = \sigma_t^2 = f(Y_{t-1}, \sigma_{t-1}^2)\}
\end{aligned}$$

and $\Theta = \{\theta = (\omega, \alpha, \beta)' \mid \omega > 0, \alpha,\, \beta \geq 0\}$.

Typical asymmetric alternatives contained in $\mathcal{M}$ are GQARCH(1,1) processes introduced by Sentana (1995), where

$$\sigma_t^2 \,=\, \omega \,+\, \alpha(Y_{t-1} - \delta)^2 \,+\, \beta\, \sigma_{t-1}^2,$$

or GJR-GARCH processes with

$$\sigma_t^2 \,=\, \omega \,+\, \alpha\, Y_{t-1}^2 \,+\, \beta\, \sigma_{t-1}^2 \,+\, \delta\, Y_{t-1}^2 \mathbb{1}_{Y_{t-1}<0},$$

introduced by Glosten, Jagannathan and Runkle (1993), that are frequently used in finance. If $(Y_t)_t$ describes a sequence of log-returns of an asset and if $\delta > 0$ then negative shocks have a larger impact on the conditional volatilities than positive ones. If we had only one of these two particular alternatives in mind, we could simply test whether or not $\delta = 0$. However, other deviations from the null are of a more complicated structure, e.g. the model equation for the volatilities of EGARCH(1,1) processes is given by

$$\ln \sigma_t^2 = \omega + \alpha \left\{ \theta\, \frac{Y_{t-1}}{\sigma_{t-1}} + \zeta \left( \left| \frac{Y_{t-1}}{\sigma_{t-1}} \right| - E \left| \frac{Y_{t-1}}{\sigma_{t-1}} \right| \right) \right\} + \beta\, \ln \sigma_{t-1}^2 \quad \omega, \alpha, \beta, \theta, \zeta \in \mathbb{R};$$

see Nelson (1991). Therefore, we strive for a more general test that is also consistent against unspecified deviations from a GARCH(1,1) model. It can be expected that our test procedure can be generalized to a test for GARCH($p$,$q$) specification but this extension would be very technical and is therefore not carried out here.

A GARCH(1,1) process $(Y_t)_{t\in\mathbb{Z}}$ satisfies the equations

$$\sigma_t^2 \,=\, \omega \,+\, \alpha Y_{t-1}^2 \,+\, \beta\sigma_{t-1}^2 \tag{2.1}$$

and

$$Y_t \,=\, \sigma_t\, \varepsilon_t. \tag{2.2}$$

Under $\mathcal{H}_0$, we denote by $\theta_0 = (\omega_0, \alpha_0, \beta_0)'$ the true parameter and assume

**(A1)** (i) $\omega_0 > 0$, $\alpha_0, \beta_0 \geq 0$,
(ii) $(\varepsilon_t)_{t\in\mathbb{Z}}$ i.i.d., $E\varepsilon_0^2 = 1$, and $E[\ln(\beta_0 \,+\, \alpha_0\varepsilon_0^2)] \,<\, 0$.

According to Theorem 2 in Nelson (1990), there exists a unique strictly stationary and ergodic solution to (2.1) and (2.2) that can be rewritten (see Equation (10) in Nelson (1990)) as

$$\sigma_t^2 \,=\, \omega_0 \left[ 1 \,+\, \sum_{k=1}^{\infty} \prod_{i=1}^{k} (\beta_0 \,+\, \alpha_0\varepsilon_{t-i}^2) \right]. \tag{2.3}$$

Making repeatedly use of the model equations (2.1) and (2.2) we get

$$\begin{aligned}\sigma_t^2 &= \omega_0 + \alpha_0 Y_{t-1}^2 + \beta_0 \sigma_{t-1}^2 \\ &= (\omega_0 + \alpha_0 Y_{t-1}^2) + \beta_0(\omega_0 + \alpha_0 Y_{t-2}^2) + \cdots + \beta_0^{K-1}(\omega_0 + \alpha_0 Y_{t-K}^2) + \beta_0^K \sigma_{t-K}^2. \quad (2.4)\end{aligned}$$

Since $(\sigma_t^2)_{t\in\mathbb{Z}}$ is stationary and since $E[\ln(\beta_0 + \alpha_0\varepsilon_0^2)] < 0$ implies $\beta_0 < 1$, we see that the last summand on the right-hand side of (2.4) tends to zero as $K \to \infty$, i.e., we obtain the alternative representation

$$\sigma_t^2 = \sum_{k=1}^{\infty} \beta_0^{k-1}(\omega_0 + \alpha_0 Y_{t-k}^2) = \frac{\omega_0}{1-\beta_0} + \alpha_0 \sum_{k=1}^{\infty} \beta_0^{k-1} Y_{t-k}^2. \quad (2.5)$$

For any parameter $\theta = (\omega, \alpha, \beta)'$, we define a stationary sequence of approximations of the volatilities that are based on the $Y_t$s but correspond to the model with parameter $\theta$ as

$$\sigma_t^2(\theta) = \frac{\omega}{1-\beta} + \alpha \sum_{k=1}^{\infty} \beta^{k-1} Y_{t-k}^2. \quad (2.6)$$

We have obviously $\sigma_t^2(\theta_0) = \sigma_t^2$. More importantly, for $\theta$ close to $\theta_0$, $(\sigma_t^2(\theta))_{t\in\mathbb{Z}}$ is strictly stationary and ergodic, which implies that, for any fixed value of an estimator $\widehat{\theta}_n$ of $\theta_0$, $\sigma_t^2(\widehat{\theta}_n)$ serves as a suitable approximation to the (nonstationary) estimated volatilities $\widehat{\sigma}_t^2$ that will be specified below. The following lemma shows that the above definition is correct if $\theta$ is sufficiently close to $\theta_0$ and that $\sigma_t^2(\theta)$ converges to $\sigma_t^2$ as $\theta \to \theta_0$.

**Lemma 2.1.** (i) *For* $\theta = (\omega, \alpha, \beta)' \in \Theta$ *satisfying* $E[\ln((\beta_0 \vee \beta) + (\alpha_0 \vee \alpha)\varepsilon_0^2)] < 0$, $(\sigma_t^2(\theta))_{t\in\mathbb{Z}}$ *is the unique stationary solution to*

$$\sigma_t^2(\theta) = \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta), \qquad t \in \mathbb{Z}. \quad (2.7)$$

$\sigma_t^2(\theta)$ *is finite with probability 1.*

(ii) $\sup_{\theta\in\Theta:\, \|\theta-\theta_0\|\le\delta} \left|\sigma_t^2(\theta) - \sigma_t^2\right| \underset{\delta\to 0}{\longrightarrow} 0$ *with probability 1.*

We intend to test a composite hypothesis, i.e. the GARCH(1,1) parameters are unknown and have to be estimated. In accordance with Francq and Zakoïan (2004) we use the quasi-maximum likelihood estimator with a normal reference distribution, which is defined as

$$\widehat{\theta}_n = \arg\min_{\theta\in\Theta_0} \bar{\mathcal{L}}_n(\theta).$$

Here,

$$\Theta_0 = \{\theta = (\omega, \alpha, \beta)' \mid \beta \le \rho_0,\ u_1 \le \min\{\omega, \alpha, \beta\} \le \max\{\omega, \alpha, \beta\} \le u_2\}$$

with some $0 < u_1 < u_2 < \infty$ and $\rho_0 \in (0, 1)$, and $-n\,\bar{\mathcal{L}}_n$ denotes the logarithmic quasi-likelihood function (constant terms are ignored here), given by

$$\bar{\mathcal{L}}_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} \left( \log \bar{\sigma}_t^2(\theta) + \frac{Y_t^2}{\bar{\sigma}_t^2(\theta)} \right),$$

where

$$\bar{\sigma}_t^2(\theta) = \omega + \alpha Y_{t-1} + \beta \bar{\sigma}_{t-1}^2(\theta), \quad t \ge 1.$$

In principle, the initial value $\bar{\sigma}_0^2$ can be chosen arbitrarily. For sake of definiteness, we follow the suggestion (2.7) in Francq and Zakoïan (2004) and set $\bar{\sigma}_0^2(\theta) = Y_0^2$.

We assume

**(A2)** (i) $E|\varepsilon_0|^4 < \infty$ and $\mathrm{var}(\varepsilon_0^2) > 0$,
(ii) $\theta_0$ is in the interior of $\Theta_0$.

Francq and Zakoïan (2004) proved strong consistency and asymptotic normality of the quasi-maximum-likelihood estimator (QMLE) in the framework of GARCH($p$,$q$) processes. Under the above conditions we obtain from their results, in the special case of a GARCH(1,1) process considered here, the Bahadur linearization

$$\widehat{\theta}_n - \theta_0 \;=\; \frac{1}{n}\sum_{t=1}^{n} L_t \,+\, o_P\left(n^{-1/2}\right) \qquad \text{with} \qquad L_t \;=\; (\varepsilon_t^2 - 1)(E[\ddot{W}_0(\theta_0)])^{-1}\frac{\dot{\sigma}_t^2(\theta_0)}{\sigma_t^2(\theta_0)}, \quad (2.8)$$

where $\dot{\sigma}_t^2(\theta) = (\partial\sigma_t^2(\theta)/\partial\omega, \partial\sigma_t^2(\theta)/\partial\alpha, \partial\sigma_t^2(\theta)/\partial\beta)'$, $W_0(\theta) = \log\sigma_0^2(\theta) + Y_0^2/\sigma_0^2(\theta)$ and $\ddot{W}_0$ denotes its Hessian with respect to $\theta$.

*Remark* 1. The practical derivation of the QMLE is based on an optimization problem and therefore computationally intensive. For that reason, Kristensen and Linton (2006) proposed a moment-based approach to estimate the GARCH parameters. They provide explicit expressions for their estimators, however, their method is only reliable for very large sample sizes. Therefore and for sake of definiteness, we stick to the QMLE in the sequel.

## 3. The test statistic and its asymptotics

We propose a test of Cramér-von Mises type. At first glance, the statistic

$$\bar{T}_n \;=\; n\int_{\mathbb{R}^2}\left\{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t^2 - \widehat{\sigma}_t^2\right)\, w(z_1 - Y_{t-1}, z_2 - \widehat{\sigma}_{t-1}^2)\right\}^2 Q(dz_1, dz_2)$$

seems natural, where $\widehat{\sigma}_0^2 = Y_0^2$ and $\widehat{\sigma}_t^2 = \widehat{\omega}_n + \widehat{\alpha}_n Y_{t-1}^2 + \widehat{\beta}_n\widehat{\sigma}_{t-1}^2$ $(t = 1, \ldots, n)$ is a model-based approximation of the unobserved volatility. Here, $w$ is a weight function and $Q$ a probability measure. Since $E(Y_t^2 \mid Y_{t-1}, \sigma_{t-1}^2) - \sigma_t^2 = 0$ under $\mathcal{H}_0$, one would reject the null hypothesis if the value of the test statistic is large. However, the test statistic is of a very complicated structure and critical values cannot be determined directly. A bootstrap-aided testing procedure will be proposed below to circumvent these difficulties. In order to show its asymptotic validity, we would have to impose certain moment constraints, such as finite fourth moments of $Y_t$. The latter assumption would be rather restrictive and would rule out e.g. IGARCH processes $(\alpha + \beta = 1)$ that are frequently applied in financial mathematics; see Lee and Hansen (1994). In contrast, moment assumptions on the innovations are far less restrictive and have already been presumed by Berkes, Horváth, and Kokoszka (2003) as well as Francq and Zakoïan (2004) to obtain asymptotic normality of the quasi-maximum likelihood estimator for the GARCH(1,1) parameter vector. It turns out that moment conditions on the innovations suffice to derive the asymptotics of the slightly modified test statistic

$$\widehat{T}_n \;=\; n\int_{\mathbb{R}^2}\left\{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\widehat{\sigma}_t^2} - 1\right)\, w(z_1 - Y_{t-1}, z_2 - \widehat{\sigma}_{t-1}^2)\right\}^2 Q(dz_1, dz_2).$$

We will show that a bootstrap-aided test based on this statistic is consistent and asymptotically level-$\gamma$. Moreover, it is well known that tests of this type are suitable to detect local alternatives of Pitman type.

We first show that the test statistic can be approximated by a von Mises ($V$-) statistic not depending on the estimators $(\widehat{\sigma}_t^2)_t$ and $\widehat{\theta}_n$ but on the true quantities $(\sigma_t^2)_t$ and $\theta_0$, respectively. The limit distribution of this approximating statistic can then easily be obtained from recent results by Leucht and Neumann (2013a). To this end, we need the kernel of this statistic being continuous which is ensured by the next assumption. Furthermore, in order to keep the effect of approximating the unobserved volatilities in the weight function negligible we require an extra condition on the smoothness of $w$. We make the following assumptions regarding the weight function and the measure $Q$:

**(A3)** $Q$ is a probability measure on $(\mathbb{R}^2, \mathcal{B}(\mathbb{R}^2))$. The weight function $w$ is non-negative, bounded, and measurable. Moreover, there is some $C_w < \infty$ such that

$$\int_{\mathbb{R}^2} |w(z-(y_1,s_1)') \,-\, w(z-(y_2,s_2)')|^2\, Q(dz) \,\leq\, C_w\, \|(y_1,s_1)' \,-\, (y_2,s_2)'\|_2. \qquad (3.1)$$

*Remark* 2. (3.1) is obviously satisfied if $w$ is Lipschitz continuous. It is also satisfied if $w(z-(y,s)') = \mathbb{1}((y,s)' \preceq z)$ and $Q$ has bounded marginal densities $q_1$ and $q_2$. Here and below $(a_1,a_2)' \preceq (b_1,b_2)'$ means that $a_1 \leq b_1$ and $a_2 \leq b_2$.

Subsequently, we will abbreviate the information variable $(Y_{t-1}, \sigma_{t-1}^2(\theta))'$ by $I_{t-1}(\theta)$.

**Lemma 3.1.** *Suppose that $\mathcal{H}_0$ holds true and that (A1) - (A3) are satisfied. Then*

$$\widehat{T}_n \,-\, T_n \,=\, o_P(1),$$

*where*

$$T_n \,=\, \int_{\mathbb{R}^2} \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( \left(\varepsilon_t^2 - 1\right)\, w(z - I_{t-1}(\theta_0)) \,-\, E_{\theta_0}\left[ \frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)}\, w(z - I_0(\theta_0)) \right]' L_t \right) \right\}^2 Q(dz).$$

Note that $T_n$ is a $V$-statistic that is degenerate under $\mathcal{H}_0$, i.e., it can be represented as $T_n = n^{-1} \sum_{s,t=1}^n h(X_s, X_t)$ with $Eh(X_0, x) = 0\ \forall x$, where $X_t = (\varepsilon_t^2, Y_{t-1}, \sigma_{t-1}^2, L_t')'$ and

$$\begin{aligned} h(x,\bar{x}) \;&=\; \int_{\mathbb{R}^2} \left\{ (x_1 - 1)\, w(z - (x_2,x_3)') \,-\, E_{\theta_0}\left[ \frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)}\, w(z - I_0(\theta_0)) \right]' x_4 \right\} \\ &\qquad \times \left\{ (\bar{x}_1 - 1)\, w(z - (\bar{x}_2,\bar{x}_3)') \,-\, E_{\theta_0}\left[ \frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)}\, w(z - I_0(\theta_0)) \right]' \bar{x}_4 \right\} Q(dz) \end{aligned}$$

Thus, its asymptotics can be immediately deduced from a recent result on degenerate $V$-statistics of ergodic data by Leucht and Neumann (2013a). In conjunction with the previous lemma, we obtain the limit distribution of $\widehat{T}_n$.

**Proposition 3.1.** *Suppose that $\mathcal{H}_0$ holds true and that (A1) - (A3) are satisfied. Then*

$$\widehat{T}_n \xrightarrow{d} Z = \sum_{k=1}^\infty \lambda_k Z_k^2.$$

*Here, $(Z_k)_k$ is a sequence of independent standard normal random variables and $(\lambda_k)_k$ denotes the (finite or countably infinite) sequence of nonzero eigenvalues of the equation $\lambda\, \Phi(x) = \int h(x,\bar{x})\, \Phi(\bar{x})\, P_{\theta_0}^X(d\bar{x})$, enumerated according to their multiplicity.*

Now, we consider the behavior of the test statistic under fixed alternatives. To this end, we assume the parameter estimator, that is obtained by the quasi-maximum likelihood approach described in Section 2, to be consistent for some pseudo-true parameter $\bar{\theta}_0$.

**(A4)** $\widehat{\theta}_n \stackrel{P}{\longrightarrow} \bar{\theta}_0 \in \Theta_0$.

Additionally, we impose some regularity conditions on the model under the alternative.

**(A5)** $(Y_t)_{t\in\mathbb{Z}}$ is strictly stationary and ergodic. Moreover, $E|Y_0|^s < \infty$ for some $s > 0$ and $E[Y_0^4/(\sigma_0^2(\bar{\theta}_0))^2] < \infty$.

Note that the first moment condition ensures almost sure finiteness of $(\sigma_t^2(\theta))_{t\in\mathbb{Z}}$ for $\theta$ in a neighborhood of $\bar{\theta}_0$. As expected, the test statistic turns out to be asymptotically unbounded under $\mathcal{H}_1$.

**Proposition 3.2.** *Suppose that (A1) and (A3) - (A5) hold true, with $\theta_0$ replaced by $\bar{\theta}_0$. Then*

(i) $n^{-1}\widehat{T}_n \stackrel{P}{\longrightarrow} \int_{\mathbb{R}^2} \left\{E[(Y_1^2/\sigma_1^2(\bar{\theta}_0) - 1)\, w(z - I_0(\bar{\theta}_0))]\right\}^2 Q(dz)$.

(ii) *If additionally the relation $E[(Y_1^2/\sigma_1^2(\bar{\theta}_0) - 1)\, w(z - I_0(\bar{\theta}_0))] \neq 0$ for all $z \in \Pi$ and some $\Pi$ with $Q(\Pi) > 0$ holds true, then*

$$\widehat{T}_n \stackrel{P}{\longrightarrow} \infty.$$

*Remark* 3. Provided that $Q$ has an everywhere positive density, a weight function that satisfies the additional condition in Proposition 3.2(ii) for all $\mathcal{H}_1$-scenarios is $w(z - I_t) = \mathbb{1}_{I_t \preceq z}$ which is frequently used in Cramér-Mises type tests; cf. Lemma 1(d) in Escanciano (2006).

## 4. A bootstrap-based test

We see from the previous section that the null distribution of the test statistic $\widehat{T}_n$ and also its limit distribution depend on the unknown parameter $\theta_0$ in a complicated way. In particular, the eigenvalues $(\lambda_k)_k$ appearing in the limit are unknown and it is not clear at all how they can be computed in an efficient manner. Therefore, (asymptotic) critical values of a test based on this statistic cannot be derived directly. The bootstrap offers a convenient tool to circumvent these difficulties. In the present context, a model-based bootstrap is probably the first choice since it can be expected to be more precise than alternative model-free methods. We propose the following algorithm:

(1) Compute the residuals

$$e_t = Y_t/\widehat{\sigma}_t, \quad t = 1, \ldots, n.$$

(2) Calculate standardized versions

$$\widehat{\varepsilon}_t = e_t \Big/ \sqrt{n^{-1}\sum_{s=1}^n e_s^2}.$$

(3) Draw independent bootstrap innovations

$$\varepsilon_t^* \sim F_{n,\widehat{\varepsilon}}, \quad \text{where } F_{n,\widehat{\varepsilon}}(x) = n^{-1}\sum_{t=1}^n \mathbb{1}(\widehat{\varepsilon}_t \leq x).$$

(4) Compute $\sigma_t^{*2}$, $Y_t^*$ recursively:

$$\begin{aligned}\sigma_0^{*2} &= \widehat{\omega}_n\left[1+\sum_{k=1}^{\infty}\prod_{i=1}^{k}(\widehat{\beta}_n + \widehat{\alpha}_n\varepsilon_{-i}^{*2})\right],\\ Y_0^* &= \sigma_0^*\,\varepsilon_0^*,\end{aligned}$$

and then, for $t=1,\ldots,n$,

$$\begin{aligned}\sigma_t^{*2} &= \widehat{\omega}_n + \widehat{\alpha}_n Y_{t-1}^{*2} + \widehat{\beta}_n\sigma_{t-1}^{*2},\\ Y_t^* &= \sigma_t^*\,\varepsilon_t^*.\end{aligned}$$

(5) Compute the bootstrap statistic

$$\widehat{T}_n^* = n\int_{\mathbb{R}^2}\left\{\frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t^{*2}}{\widehat{\sigma}_t^{*2}}-1\right)w(z_1-Y_{t-1}^*, z_2-\widehat{\sigma}_{t-1}^{*2})\right\}^2 Q(dz_1,dz_2),$$

with arbitrarily chosen $\widehat{\sigma}_0^{*2}$ and $\widehat{\sigma}_t^{*2}=\widehat{\omega}_n^*+\widehat{\alpha}_n^*\,Y_{t-1}^{*2}+\widehat{\beta}_n^*\,\widehat{\sigma}_{t-1}^{*2}$. Here, $\widehat{\theta}_n^*=(\widehat{\omega}_n^*,\widehat{\alpha}_n^*,\widehat{\beta}_n^*)'$ is the QMLE based on the bootstrap sample.

(6) Repeat steps (3) to (5) $B$ times and, for a nominal size of $\gamma\in(0,1)$, choose $t_\gamma^*$ as any $(1-\gamma)$-quantile of the empirical distribution of $\widehat{T}_{n,1}^*,\ldots,\widehat{T}_{n,B}^*$.

(7) Reject the null hypothesis if $\widehat{T}_n > t_\gamma^*$.

In order to validate asymptotic correctness of the algorithm above, we do not imitate all the proofs of Section 3. Instead, an appropriate coupling of $X_t=(\varepsilon_t, Y_{t-1}, \sigma_{t-1}^2, L_t')'$ and $X_t^*=(\varepsilon_t^{*2}, Y_{t-1}^*, \sigma_{t-1}^{*2}, L_t^{*\prime})'$ (with $\sigma_{t-1}^{*2}=\sigma_{t-1}^{*2}(\widehat{\theta}_n)$ and $\sigma_t^{*2}(\theta)$ being the bootstrap analogue to $\sigma_t^2(\theta)$) directly results in a coupling of the corresponding test statistics on the original and on the bootstrap side. Here, $L_t^*=(\varepsilon_t^{*2}-1)(E^*[\ddot{W}_0^*(\widehat{\theta}_n)])^{-1}\dot{\sigma}_t^{*2}(\widehat{\theta}_n)/\sigma_t^{*2}(\widehat{\theta}_n)$.

To express distributional convergence in conjunction with the additional qualification "almost surely" properly and to describe closeness of two distributions both depending on $n$, we use the Lévy metric $d_L$ which is defined, for distribution functions $G$ and $H$ on $\mathbb{R}$, as

$$d_L(G,H) = \inf\{\varepsilon:\quad G(x-\varepsilon)-\varepsilon\le H(x)\le G(x+\varepsilon)+\varepsilon\quad\forall x\in\mathbb{R}\}.$$

Applied to random variables $U$ and $V$ with c.d.f. $F_U$ and $F_V$, we also use the notation $d_L(U,V)$. A first step towards our proof of bootstrap consistency is done by the following lemma.

**Lemma 4.1.** *Assume that $\mathcal{H}_0$ holds true and that (A1) - (A3) are fulfilled. Then*

$$d_L(F_{n,\widehat{\varepsilon}}, F_\varepsilon) \xrightarrow{a.s.} 0.$$

Now we are in the position to construct a coupling of the random variables $X_t = (\varepsilon_t^2, Y_{t-1}, \sigma_{t-1}^2, L_t')'$ appearing in the approximating $V$-statistic $T_n$ with the random variables $X_t^*=(\varepsilon_t^{*2}, Y_{t-1}^*, \sigma_{t-1}^{*2}, L_t^{*\prime})'$.

**Lemma 4.2.** *Assume that $\mathcal{H}_0$ holds true and that (A1) - (A3) are fulfilled. On a sufficiently rich probability space $(\widetilde{\Omega},\widetilde{\mathcal{A}},\widetilde{P})$, there exist independent random vectors $((\widetilde{\varepsilon}_t,\widetilde{\varepsilon}_t^*)')_{t\in\mathbb{Z}}$ such* *that*

$$\widetilde{\varepsilon}_t \overset{d}{=} \varepsilon_t,$$
$$\widetilde{\varepsilon}_t^* \overset{d}{=} \varepsilon_t^*,$$

*and, with* $\widetilde{L}_t$, $\widetilde{Y}_t$, $\widetilde{\sigma}_t^2$ *and* $\widetilde{L}_t^*$, $\widetilde{Y}_t^*$, $\widetilde{\sigma}_t^{*2}$ *being versions based on the* $\widetilde{\varepsilon}_s$ *and* $\widetilde{\varepsilon}_s^*$, *respectively,*

$$E_{\widetilde{P}}\left[(\widetilde{\varepsilon}_t^* - \widetilde{\varepsilon}_t)^2 + \|\widetilde{L}_t^* - \widetilde{L}_t\|^2 + |\widetilde{Y}_{t-1}^* - \widetilde{Y}_{t-1}| \wedge 1 + |\widetilde{\sigma}_{t-1}^{*2} - \widetilde{\sigma}_{t-1}^2| \wedge 1\right] \xrightarrow{a.s.} 0.$$

As a consequence of the above coupling, the following assertion provides a useful approximation for the two hypothetical volatility processes.

**Corollary 4.1.** *Assume that* $\mathcal{H}_0$ *holds true and that (A1) - (A3) are fulfilled. On the probability space* $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$ *from Lemma 4.2,*

$$E_{\widetilde{P}}\left[\sup_{\theta\in\Theta_0} \left|\widetilde{\sigma}_t^{*2}(\theta) - \widetilde{\sigma}_t^2(\theta)\right| \wedge 1\right] \xrightarrow{\widetilde{P}} 0,$$

*where* $\widetilde{\sigma}_t^{*2}(\theta)$ *and* $\widetilde{\sigma}_t^2(\theta)$ *are versions of the* $\sigma_t^{*2}(\theta)$ *and* $\sigma_t^2(\theta)$, *respectively, based on the* $\widetilde{Y}_t^*$ *and* $\widetilde{Y}_t$.

The asymptotics of the test statistic $\widehat{T}_n$ heavily relies on the linearization of the QMLE $\widehat{\theta}_n$. We now establish this property for the bootstrap QMLE.

**Lemma 4.3.** *Assume that* $\mathcal{H}_0$ *holds true and that (A1) - (A3) are fulfilled. Then*

$$\widehat{\theta}_n^* - \widehat{\theta}_n = \frac{1}{n}\sum_{t=1}^n L_t^* + o_{P^*}\left(\frac{1}{\sqrt{n}}\right).$$

These results enable us to derive a bootstrap analogue to Lemma 3.1.

**Lemma 4.4.** *Assume that* $\mathcal{H}_0$ *holds true and that (A1) - (A3) are fulfilled. Then, with* $T_n^*$ *being the bootstrap analogue of* $T_n$,

$$\widehat{T}_n^* = T_n^* + o_{P^*}(1).$$

Hence, bootstrap validity under the null hypothesis can be deduced from the following result.

**Theorem 4.1.** *Assume that* $\mathcal{H}_0$ *holds true and that (A1) - (A3) are fulfilled. Then*

$$d_L(T_n^*, T_n) \xrightarrow{P} 0.$$

To show hat our bootstrap method is also valid under fixed alternatives, we additionally assume

**(A6)** (i) $E^*[(\varepsilon_1^*)^4] = O_P(1)$,
(ii) $E^*[(\sigma_1^{*2}/\sigma_t^{*2}(\widehat{\theta}_n^*) - 1)^2] \xrightarrow{P} 0$.
(iii) There exists some $\delta > 0$ such that $P(E^*[\ln(\widehat{\beta}_n + \widehat{\alpha}_n \varepsilon_1^{*2})] \leq -\delta) \longrightarrow_{n\to\infty} 1$.

**Lemma 4.5.** *Suppose that (A3) and (A6) are fulfilled. Then*

$$E^*\left[n^{-1}\widehat{T}_n^*\right] \xrightarrow{P} 0.$$

Thus, the above algorithm leads to a consistent, asymptotic level-$\gamma$ test.

**Corollary 4.2.** (i) *Assume that $\mathcal{H}_0$ holds true, that (A1) to (A3) are fulfilled and that additionally $E[h(X_1, X_1)] > 0$. Then*

$$P(\widehat{T}_n > t_\gamma^*) \underset{n\to\infty}{\longrightarrow} \gamma.$$

(ii) *Under $\mathcal{H}_1$ and if additionally the prerequisites of Proposition 3.2(ii) and (A6) are satisfied, then*

$$P(\widehat{T}_n > t_\gamma^*) \underset{n\to\infty}{\longrightarrow} 1.$$

*Remark* 4. Our test has similarities to the methodology proposed by Escanciano (2008) in the general context of mean and variance specification testing. While our test is based on the marked empirical process $n^{-1/2}\sum_{t=1}^n (Y_t^2/\widehat{\sigma}_t^2 - 1)\, w(z_1 - Y_{t-1}, z_2 - \widehat{\sigma}_t^2)$, Escanciano's test is based on the processes $(n-j-1)^{-1/2}\sum_{t=j}^n (Y_t^2 - \widehat{\sigma}_t^2)\, w(Y_{t-j}, z)$, for $j = 1, 2, \ldots$, i.e., he uses only observable random variables in the weight function. However, in a previous version of that paper, Escanciano (2007b), he allows for models with infinitely many explanatory variables. In principle, using representation (2.5) above it seems that these results could be applied in our case, too. In order to apply his result to our test problem, we would have to assume that certain moments of the observed process exist; see his assumption A1(b). However, this typically is not guaranteed in financial time series as we already discussed at the beginning of Section 3.

Escanciano (2008) proposed a wild bootstrap procedure to determine critical values of $L_2$-type tests. Instead of adapting his approach, we decided to apply a model-based approach here since this kind of resampling procedure often outperforms model-free bootstrap methods.

## 5. Numerical Examples

We illustrate the finite sample behavior of the proposed test by some simulations. We use the indicator function $w(z - I_t) = \mathbb{1}_{I_t \preceq z}$ as a weight function which is admissible in view of Remark 3, and choose $Q = \mathcal{N}(0, 25) \otimes \mathcal{N}(0, 25)$. Straightforward calculations show that in this case the test statistic simplifies to

$$\widehat{T}_n = \frac{1}{n}\sum_{s,t=1}^n \left(\frac{Y_s}{\widehat{\sigma}_s^2} - 1\right)\left(\frac{Y_t}{\widehat{\sigma}_t^2} - 1\right)(1 - \Phi_{0,25}(\max\{Y_s, Y_t\}))\left(1 - \Phi_{0,25}(\max\{\widehat{\sigma}_s^2, \widehat{\sigma}_t^2\})\right),$$

where $\Phi_{0,25}(\cdot) = \Phi(\cdot/5)$ and $\Phi$ denotes the cumulative distribution function of $\mathcal{N}(0, 1)$. To study of the performance of our test under the null hypothesis as well as certain alternative scenarios we choose the innovations to be standard normal, draw samples of size $n = 500$ and $n = 1000$ and run a Monte-Carlo simulation $N = 500$ times, each with $B = 500$ bootstrap replications. In order to meet our assumption of stationarity, we discarded 500 pre-sample data values of the corresponding processes. The implementation was carried out with the aid of the statistical software package *R*; see R Core Team (2012). To estimate the GARCH parameters we use the routine of Tinkl (2013).

Figure 1 shows a realization of a GARCH(1,1) process whereas Figures 2 and 3 show realizations of a GQARCH and a GJR-GARCH process, respectively.

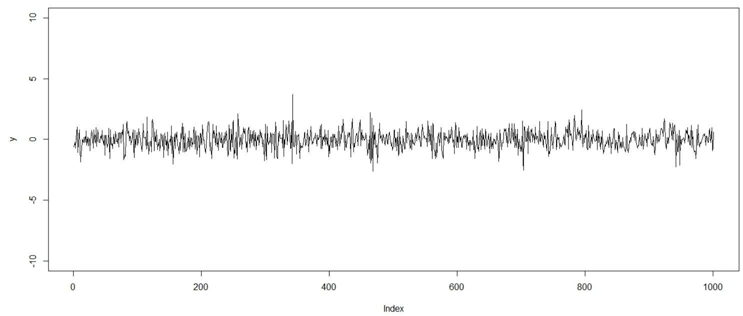


FIGURE 1. GARCH process with parameter $\theta = (0.2, 0.25, 0.35)'$.

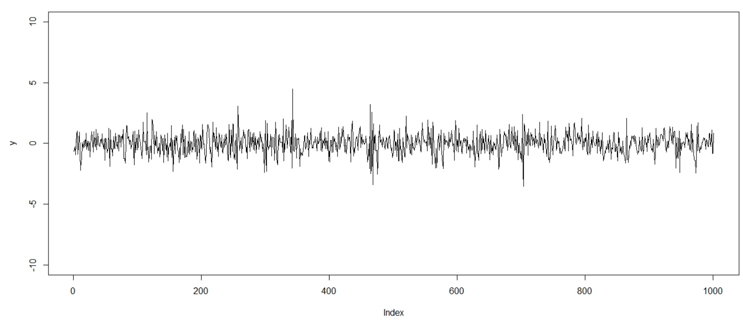


FIGURE 2. GQARCH process with parameters $\theta = (0.2, 0.25, 0.35)'$ and $\delta = 0.5$.

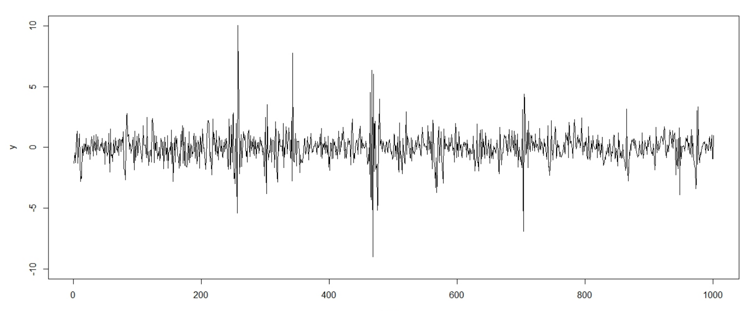


FIGURE 3. GJR-GARCH process with parameters $\theta = (0.2, 0.25, 0.35)'$ and $\delta = 0.5$.

The rejection frequencies of our test under two null scenarios and for nominal significance levels $\gamma = 0.05$ and $\gamma = 0.1$ are summarized in Table 1. Table 2 and Table 3 report the finite sample behavior of our procedure under GQARCH and GJR-GARCH alternatives.

Table 1. Rejection frequencies

| | | $\theta = (0.20, 0.15, 0.25)'$ | $\theta = (0.20, 0.25, 0.35)'$ |
|---|---|---|---|
| $n = 500$ | $\gamma = 0.05$ | 0.046 | 0.068 |
| | $\gamma = 0.10$ | 0.094 | 0.120 |
| $n = 1000$ | $\gamma = 0.05$ | 0.052 | 0.060 |
| | $\gamma = 0.10$ | 0.100 | 0.112 |

GQARCH

| | | | $\theta = (0.20, 0.15, 0.25)'$ | $\theta = (0.20, 0.25, 0.35)'$ |
|---|---|---|---|---|
| $n = 500$ | $\delta = 0.25$ | $\gamma = 0.05$ | 0.254 | 0.528 |
| | | $\gamma = 0.10$ | 0.386 | 0.634 |
| | $\delta = 0.50$ | $\gamma = 0.05$ | 0.660 | 0.940 |
| | | $\gamma = 0.10$ | 0.816 | 0.982 |
| $n = 1000$ | $\delta = 0.25$ | $\gamma = 0.05$ | 0.624 | 0.846 |
| | | $\gamma = 0.10$ | 0.748 | 0.920 |
| | $\delta = 0.50$ | $\gamma = 0.05$ | 0.982 | 1.000 |
| | | $\gamma = 0.10$ | 0.994 | 1.000 |

GJR-GARCH

| | | | $\theta = (0.20, 0.15, 0.25)'$ | $\theta = (0.20, 0.25, 0.35)'$ |
|---|---|---|---|---|
| $n = 500$ | $\delta = 0.25$ | $\gamma = 0.05$ | 0.424 | 0.392 |
| | | $\gamma = 0.10$ | 0.558 | 0.502 |
| | $\delta = 0.50$ | $\gamma = 0.05$ | 0.874 | 0.774 |
| | | $\gamma = 0.10$ | 0.936 | 0.866 |
| $n = 1000$ | $\delta = 0.25$ | $\gamma = 0.05$ | 0.754 | 0.674 |
| | | $\gamma = 0.10$ | 0.840 | 0.774 |
| | $\delta = 0.50$ | $\gamma = 0.05$ | 0.996 | 0.984 |
| | | $\gamma = 0.10$ | 0.998 | 0.994 |

It can be seen that the prescribed size is kept very well. The power behavior is convincing for all of our alternatives. Having a particular alternative in mind, the power can even be increased by a tailor-made choice of the weights $w$ and $Q$; cf. Anderson and Darling (1954) in the case of generalized Cramér-von Mises statistics.

## 6. Proofs

Throughout this section, we shortly write $\int$ instead of $\int_{\mathbb{R}^2}$. Moreover, $C$ denotes a generic, finite constant that may change its value from one line to another.

*Proof of Lemma 2.1.* *(i)*

First of all, finiteness of $\sigma_t^2(\theta)$ follows from a simple coupling argument. According to (2.5), $\sigma_t^2(\theta)$ consists only of nonnegative summands. Hence, the series $\sum_{k=1}^{\infty} \beta^{k-1} Y_{t-k}^2$ converges, possibly to infinity. We show next that this series is actually finite with probability 1 under the assumption $E[\ln((\beta_0 \vee \beta) + (\alpha_0 \vee \alpha)\varepsilon_0^2)] < 0$. To this end, we compare $\sigma_t^2(\theta)$ with

$$\check{\sigma}_t^2 \;=\; (\omega_0 \vee \omega) \left[ 1 \,+\, \sum_{k=1}^{\infty} \prod_{i=1}^{k} \left( (\beta_0 \vee \beta) \,+\, (\alpha_0 \vee \alpha) \varepsilon_{t-i}^2 \right) \right]. \tag{6.1}$$

In contrast to $\sigma_t^2(\theta)$, $\breve{\sigma}_t^2$ is the solution to a system of GARCH(1,1) equations,

$$\begin{aligned}\breve{\sigma}_t^2 &= (\omega_0 \vee \omega) + (\alpha_0 \vee \alpha)\breve{Y}_{t-1}^2 + (\beta_0 \vee \beta)\breve{\sigma}_{t-1}^2, \\ \breve{Y}_t &= \breve{\sigma}_t \varepsilon_t\end{aligned}$$

Hence, we can use available theory and we obtain from Theorem 2 of Nelson (1990) that $\breve{\sigma}_t^2$ is finite with probability 1. Comparing (2.3) with (6.1) we see that $\sigma_t^2 \leq \breve{\sigma}_t^2$ almost surely. Hence, $Y_t^2 \leq \breve{Y}_t^2$, and since $\breve{\sigma}_t^2$ can be rewritten as

$$\breve{\sigma}_t^2 = \sum_{k=1}^{\infty} (\beta_0 \vee \beta)^{k-1} \left( (\omega_0 \vee \omega) + (\alpha_0 \vee \alpha)\breve{Y}_{t-k}^2 \right),$$

we see that $\sigma_t^2(\theta) \leq \breve{\sigma}_t^2$. Hence, $\sigma_t^2(\theta)$ is finite with probability 1. Now we obtain that

$$\begin{aligned}\sigma_t^2(\theta) &= \omega + \alpha Y_{t-1}^2 + \beta \sum_{k=1}^{\infty} \beta^{k-1}(\omega + \alpha Y_{t-1-k}^2) \\ &= \omega + \alpha Y_{t-1}^2 + \beta \sigma_{t-1}^2(\theta),\end{aligned}$$

i.e., $(\sigma_t^2(\theta))_{t\in\mathbb{Z}}$ solves the system of equations (2.7). As for uniqueness, assume that $(\widetilde{\sigma}_t^2)_{t\in\mathbb{Z}}$ is any arbitrary stationary solution to (2.7). Then we obtain from a repeated application of this equation that

$$\left|\sigma_t^2(\theta) - \widetilde{\sigma}_t^2\right| \leq \beta^K \left|\sigma_{t-K}^2(\theta) - \widetilde{\sigma}_{t-K}^2\right|.$$

Since our assumption $E[\ln((\beta_0 \vee \beta) + (\alpha_0 \vee \alpha)\varepsilon_0^2)] < 0$ implies that $\beta < 1$, we conclude that $\widetilde{\sigma}_t^2 = \sigma_t^2(\theta)$ a.s. for all $t \in \mathbb{Z}$.

*(ii)*

Since $E\varepsilon_0^2 < \infty$, the function $(\alpha, \beta) \mapsto E[\ln(\beta + \alpha\varepsilon_0^2)]$ is continuous. Hence, there exists a sufficiently small $\delta_0 > 0$ such that $E[\ln((\beta_0 + \delta_0) + (\alpha_0 + \delta_0)\varepsilon_0^2)] < 0$. If $\theta \in \Theta$ and $\|\theta - \theta_0\| \leq \delta_0$, then there exists a stationary solution $(\sigma_t^2(\theta))_{t\in\mathbb{Z}}$ to (2.7). We have the representations

$$\sigma_t^2(\theta) = \sum_{k=1}^{\infty} \beta^{k-1}(\omega + \alpha Y_{t-k}^2) \tag{6.2}$$

and

$$\sigma_t^2 = \sum_{k=1}^{\infty} \beta_0^{k-1}(\omega_0 + \alpha_0 Y_{t-k}^2). \tag{6.3}$$

If $\theta \to \theta_0$, then all summands on the right-hand side of (6.2) converge to their counterparts in (6.3). Moreover, they are majorized by $(\beta_0 \vee \beta)^{k-1}((\omega_0 \vee \omega) + (\alpha_0 \vee \alpha)Y_{t-k}^2)$. Since

$$\sum_{k=1}^{\infty} (\beta_0 \vee \beta)^{k-1}((\omega_0 \vee \omega) + (\alpha_0 \vee \alpha)Y_{t-k}^2) = \breve{\sigma}_t^2 < \infty,$$

we obtain by majorized convergence that

$$\sup_{\theta\in\Theta:\, \|\theta-\theta_0\|\leq\delta} \left|\sigma_t^2(\theta) - \sigma_t^2\right| \underset{\delta\to 0}{\longrightarrow} 0.$$

□

*Proof of Lemma 3.1.* (a) *Eliminating the effect of choosing an arbitrary initial volatility* First, we show that the effect of choosing an arbitrary initial volatility $\widehat{\sigma}_0^2$ is asymptotically negligible. Let $\widetilde{T}_n$ be the statistic based on $\sigma_t^2(\widehat{\theta}_n)$ instead of $\widehat{\sigma}_t^2$, i.e.

$$\widetilde{T}_n \;=\; \int \left\{ \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} \,-\, 1 \right) \, w(z \,-\, I_{t-1}(\widehat{\theta}_n)) \right\}^2 Q(dz). \tag{6.4}$$

We decompose the square root of the integrand in $\widehat{T}_n$ as

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \frac{Y_t^2}{\widehat{\sigma}_t^2} - 1 \right) \, w(z \,-\, (Y_{t-1}, \widehat{\sigma}_t^2)') \\
&= \; \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \frac{Y_t^2}{\widehat{\sigma}_t^2} \,-\, \frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} \right) \, w(z \,-\, (Y_{t-1}, \widehat{\sigma}_t^2)') \\
&\quad + \; \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} \,-\, 1 \right) \, \Big( w(z \,-\, (Y_{t-1}, \widehat{\sigma}_t^2)') - w(z \,-\, I_{t-1}(\widehat{\theta}_n)) \Big) \\
&\quad + \; \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \left( \frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} \,-\, 1 \right) \, w(z \,-\, I_{t-1}(\widehat{\theta}_n)) \\
&= \; S_{n,1}(z) \,+\, S_{n,2}(z) \,+\, S_{n,3}(z).
\end{aligned}$$

By the Cauchy-Schwarz inequality, we get

$$\widehat{T}_n \;=\; \widetilde{T}_n \,+\, o_P(1) \tag{6.5}$$

if $\int S_{n,i}^2(z)\, Q(dz) \;=\; o_P(1)$ for $i \;=\; 1,2$ and if $\widetilde{T}_n \;=\; \int S_{n,3}^2(z)\, Q(dz) \;=\; O_P(1)$. The latter follows from part (b) of this proof and Proposition 3.1.

If $\|\widehat{\theta}_n - \theta_0\|$ is sufficiently small, say less than $\delta > 0$, then $\sigma_t^2(\widehat{\theta}_n)$ is finite on a set with probability one. In view of

$$|\widehat{\sigma}_t^2 \,-\, \sigma_t^2(\widehat{\theta}_n)| \mathbb{1}_{\|\widehat{\theta}_n - \theta_0\| < \delta} \;\le\; \widehat{\beta}_n^t \, |\widehat{\sigma}_0^2 \,-\, \sigma_0^2(\widehat{\theta}_n)| \mathbb{1}_{\|\widehat{\theta}_n - \theta_0\| < \delta},$$

we get, since $\widehat{\beta}_n \le \rho_0 < 1$ and since the estimator $\widehat{\theta}_n$ is consistent,

$$\begin{aligned}
\int S_{n,1}^2(z)\, Q(dz) \;\; &\le \;\; \frac{C}{n} \sum_{s,t=1}^{n} \left| \frac{Y_s^2}{\sigma_s^2(\widehat{\theta}_n)} \frac{\sigma_s^2(\widehat{\theta}_n) - \widehat{\sigma}_s^2}{\widehat{\sigma}_s^2} \right| \, \left| \frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} \frac{\sigma_t^2(\widehat{\theta}_n) - \widehat{\sigma}_t^2}{\widehat{\sigma}_t^2} \right| \\
&\le \;\; O_P(1) \, \frac{1}{n} \sum_{s,t=1}^{n} \rho_0^{s+t} \left| \frac{Y_s^2}{\sigma_s^2(\widehat{\theta}_n)} \frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} \right| \,+\, o_P(1).
\end{aligned}$$

Again by consistency of $\widehat{\theta}_n$,

$$\sum_{s=1}^{n} \rho_0^s \left| \frac{Y_s^2}{\sigma_s^2(\widehat{\theta}_n)} \right| \;\le\; \sum_{s=1}^{n} \rho_0^s \, \varepsilon_s^2 \sup_{\theta \in U_\delta(\theta_0)} \frac{\sigma_s^2(\theta_0)}{\sigma_s^2(\theta)} \,+\, o_P(1).$$

This and $E[\varepsilon_0^2 \sup_{\theta:\, \|\theta - \theta_0\| \le \delta} |\sigma_0^2(\theta_0)/\sigma_0^2(\theta)|] < \infty$ for some $\delta > 0$, where the latter inequality follows from (4.26) in Francq and Zakoïan (2004), imply that $\int S_{n,1}^2(z) Q(dz) \;=\; o_P(1)$. Similarly, making use of (3.1) from (A3) we obtain

$$\int S_{n,2}^2(z)\, Q(dz) \;\le\; C \, \left| \widehat{\sigma}_0^2 - \sigma_0^2(\widehat{\theta}_n) \right| \; \frac{1}{n} \sum_{s,t=1}^{n} \rho_0^{(s+t)/2} \; \left| \frac{Y_s^2}{\sigma_s^2(\widehat{\theta}_n)} - 1 \right| \left| \frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} - 1 \right| + o_P(1) \;=\; o_P(1).$$

(b) *Proof of* $\widetilde{T}_n - T_n = o_P(1)$

Now we decompose the square root of the integrand in $\widetilde{T}_n$ as

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} - 1\right) w(z - I_{t-1}(\widehat{\theta}_n)) \\
&= \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\sigma_t^2(\theta_0)} - 1\right) w(z - I_{t-1}) - E_{\theta_0}\left[\frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)} w(z - I_0)\right]' L_t \\
&\quad + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\sigma_t^2(\theta_0)} - 1\right)\Big(w(z - I_{t-1}(\widehat{\theta}_n)) - w(z - I_{t-1})\Big) \\
&\quad + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} - \frac{Y_t^2}{\sigma_t^2(\theta_0)}\right)\Big(w(z - I_{t-1}(\widehat{\theta}_n)) - w(z - I_{t-1})\Big) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\varepsilon_t^2\frac{\dot{\sigma}_t^2(\theta_0)}{\sigma_t^2(\theta_0)}\, w(z - I_{t-1}) - E_{\theta_0}\left[\frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)}\, w(z - I_0)\right]'\right)\frac{1}{n}\sum_{s=1}^{n} L_s \\
&\quad + \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} - \frac{Y_t^2}{\sigma_t^2(\theta_0)} + \varepsilon_t^2\left[\frac{\dot{\sigma}_t^2(\theta_0)}{\sigma_t^2(\theta_0)}\right]'(\widehat{\theta}_n - \theta_0)\right) w(z - I_{t-1}) \\
&\quad - \frac{1}{\sqrt{n}}\sum_{t=1}^{n}\varepsilon_t^2\left[\frac{\dot{\sigma}_t^2(\theta_0)}{\sigma_t^2(\theta_0)}\right]'\left(\widehat{\theta}_n - \theta_0 - \frac{1}{n}\sum_{s=1}^{n} L_s\right) w(z - I_{t-1}) \\
&=: R_{n,0}(z) + R_{n,1}(z) + R_{n,2}(z) - R_{n,3}(z) + R_{n,4}(z) - R_{n,5}(z), \qquad (6.6)
\end{aligned}$$

say, where we use the abbreviation $I_t$ instead of $I_t(\theta_0)$. Since

$$E_{\theta_0}\int R_{n,0}^2(z)\,Q(dz) = E_{\theta_0}\int\left\{(\varepsilon_1^2 - 1)w(z - I_0) - E_{\theta_0}\left[\frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)} w(z - I_0)\right]' L_t\right\}^2 Q(dz) < \infty$$

by (A2), (A3), and (4.29) in Francq and Zakoïan (2004), we have that $\int R_{n,0}^2(z)\,Q(dz) = O_P(1)$ and it remains to show that

$$\int R_{n,i}^2(z)\,Q(dz) = o_P(1), \qquad \text{for } i = 1, \ldots, 5. \qquad (6.7)$$

The main tool to estimate $\int R_{n,1}^2(z)\,Q(dz)$ will be the Bernstein-type inequality for martingales given in Proposition 2.1 in Freedman (1975). Since this inequality requires bounded random variables, we have to truncate $Y_t^2/\sigma_t^2 - 1 = \varepsilon_t^2 - 1$. To this end, we define

$$\xi_{n,t} = \left(\varepsilon_t^2 - 1\right)\mathbb{1}\left(\left|\varepsilon_t^2 - 1\right| \le \sqrt{n}\right) - E_{\theta_0}\left(\left(\varepsilon_t^2 - 1\right)\mathbb{1}\left(\left|\varepsilon_t^2 - 1\right| \le \sqrt{n}\right)\right)$$

and

$$\bar{\xi}_{n,t} = \left(\varepsilon_t^2 - 1\right)\mathbb{1}\left(\left|\varepsilon_t^2 - 1\right| > \sqrt{n}\right).$$

Then, with

$$W_{s,t}(\theta) = \int [w(z - I_{s-1}(\theta)) - w(z - I_{s-1})][w(z - I_{t-1}(\theta)) - w(z - I_{t-1})]\,Q(dz),$$

we obtain

$$
\begin{aligned}
&\int R_{n,1}^2(z)\,Q(dz) \\
&\le\ \frac{3}{n}\sum_{s,t=1}^{n} E_{\theta_0}\left[(\varepsilon_s^2-1)\,\mathbb{1}\left(\left|\varepsilon_s^2-1\right|\le\sqrt{n}\right)\right] E_{\theta_0}\left[(\varepsilon_t^2-1)\,\mathbb{1}\left(\left|\varepsilon_t^2-1\right|\le\sqrt{n}\right)\right] W_{s,t}(\widehat{\theta}_n) \\
&\qquad + \frac{3}{n}\sum_{s,t=1}^{n} \bar{\xi}_{n,s}\,\bar{\xi}_{n,t}\,W_{s,t}(\widehat{\theta}_n) \\
&\qquad + \frac{3}{n}\sum_{t=1}^{n} \xi_{n,t}^2\,W_{t,t}(\widehat{\theta}_n) \\
&\qquad + \frac{6}{n}\sum_{t=2}^{n} \xi_{n,t}\left\{\sum_{s=1}^{t-1}\xi_{n,s}\,W_{s,t}(\widehat{\theta}_n)\right\} \\
&=:\ T_{n,1} + \cdots + T_{n,4}.
\end{aligned}
\tag{6.8}
$$

Since $E_{\theta_0}[\varepsilon_t^2-1)\mathbb{1}(|\varepsilon_t^2-1|\le\sqrt{n})] = -E_{\theta_0}[(\varepsilon_t^2-1)\mathbb{1}(|\varepsilon_t^2-1|>\sqrt{n})]$, we obtain that

$$
\sqrt{n}\left|E_{\theta_0}((\varepsilon_t^2-1)\mathbb{1}(|\varepsilon_t^2-1|\le\sqrt{n}))\right| \ \le\ E_{\theta_0}\left[(\varepsilon_t^2-1)^2\mathbb{1}(|\varepsilon_t^2-1|>\sqrt{n})\right] \underset{n\to\infty}{\longrightarrow} 0
$$

which implies by $|W_{s,t}(\theta)|\le 4\|w\|_\infty^2$ that

$$
T_{n,1} \ =\ o_P(1). \tag{6.9}
$$

Furthermore, we have

$$
\begin{aligned}
&P\left(|\varepsilon_t^2-1| \ >\ \sqrt{n} \quad \text{for some } t\in\{1,\ldots,n\}\right) \\
&\le\ \sum_{t=1}^{n} P\left(|\varepsilon_t^2-1| \ >\ \sqrt{n}\right) \\
&\le\ E_{\theta_0}\left[(\varepsilon_t^2-1)^2\mathbb{1}(|\varepsilon_t^2-1|>\sqrt{n})\right] \underset{n\to\infty}{\longrightarrow} 0,
\end{aligned}
$$

which leads to

$$
T_{n,2} \ =\ o_P(1). \tag{6.10}
$$

Since $W_{s,t}(\theta)$ is bounded,

$$
W_{s,t}(\theta) \ =\ O\left(\sqrt{|\sigma_{s-1}^2(\theta)-\sigma_{s-1}^2(\theta_0)|}\,\sqrt{|\sigma_{t-1}^2(\theta)-\sigma_{t-1}^2(\theta_0)|}\right) \tag{6.11}
$$

and $E[|\varepsilon_0|^4]<\infty$, we obtain by Lemma 2.1 that

$$
T_{n,3} = o_P(1). \tag{6.12}
$$

The estimation of the term $T_{n,4}$ turns out to be much more delicate. The fact that $W_{s,t}(\widehat{\theta}_n)$ is of order $O_P(\|\widehat{\theta}_n-\theta_0\|)$ might suggest that $T_{n,4}$ is negligible. However, these weights depend via $\widehat{\theta}_n$ on the whole sample and we cannot use any standard inequality for sums of martingale differences directly. To proceed, we choose a sequence of increasingly fine grids $\Theta_n=\{\theta_{n,1},\ldots,\theta_{n,M_n}\}$ on $\bar{\Theta}_n := \Theta_0\cap\{\theta : \|\theta-\theta_0\|\le\gamma_n n^{-1/2}\}$, where $\gamma_n \longrightarrow_{n\to\infty} \infty$ and $\gamma_n = O(n^\gamma)$, for some $\gamma<1/2$. Terms such as $\sum_{t=2}^n \xi_{n,t}\left\{\sum_{s=1}^{t-1}\xi_{n,s}W_{s,t}(\theta_{n,i})\right\}$ have now the desired martingale structure and we will show that

$$
\max_{1\le i\le M_n}\left|\frac{1}{n}\sum_{t=2}^{n}\xi_{n,t}\left\{\sum_{s=1}^{t-1}\xi_{n,s}W_{s,t}(\theta_{n,i})\right\}\right| \ =\ o_P(1). \tag{6.13}
$$

In order to get a meaningful result, we will choose the grids sufficiently fine such that

$$\min_{1\le i\le M_n}\left|\frac{1}{n}\sum_{t=2}^{n}\xi_{n,t}\left\{\sum_{s=1}^{t-1}\xi_{n,s}\left(W_{s,t}(\widehat{\theta}_n)\;-\;W_{s,t}(\theta_{n,i})\right)\right\}\right| \;=\; o_P(1). \tag{6.14}$$

Then (6.13) and (6.14) eventually yield that

$$T_{n,4} \;=\; o_P(1). \tag{6.15}$$

To prove (6.14), we have to find a reasonably good estimate for $|W_{s,t}(\widehat{\theta}_n) - W_{s,t}(\theta_{n,i})|$. To this end, we decompose

$$\begin{aligned}
&\left|W_{s,t}(\widehat{\theta}_n)\;-\;W_{s,t}(\theta_{n,i})\right| \\
&\quad\le\; \left|\int\left[w(z-I_{s-1}(\widehat{\theta}_n))\;-\;w(z-I_{s-1}(\theta_{n,i}))\right]\right. \\
&\qquad\qquad\left.\left[w(z-I_{t-1}(\widehat{\theta}_n))\;-\;w(z-I_{t-1})\right]\;Q(dz)\right| \\
&\qquad+\;\left|\int\left[w(z-I_{s-1}(\theta_{n,i}))\;-\;w(z-I_{s-1})\right]\right. \\
&\qquad\qquad\left.\left[w(z-I_{t-1}(\widehat{\theta}_n))\;-\;w(z-I_{t-1}(\theta_{n,i}))\right]\;Q(dz)\right| \\
&\quad\le\; \sqrt{\int\left[w(z-I_{s-1}(\widehat{\theta}_n))\;-\;w(z-I_{s-1}(\theta_{n,i}))\right]^2\;Q(dz)}\;\sqrt{W_{t,t}(\widehat{\theta}_n)} \\
&\qquad+\;\sqrt{\int\left[w(z-I_{t-1}(\widehat{\theta}_n))\;-\;w(z-I_{t-1}(\theta_{n,i}))\right]^2\;Q(dz)}\;\sqrt{W_{s,s}(\theta_{n,i})} \\
&\quad=\; O\left(\sqrt{\left|\sigma_{s-1}^2(\widehat{\theta}_n)\;-\;\sigma_{s-1}^2(\theta_{n,i})\right|}\;+\;\sqrt{\left|\sigma_{t-1}^2(\widehat{\theta}_n)\;-\;\sigma_{t-1}^2(\theta_{n,i})\right|}\right),
\end{aligned}$$

where the latter equation follows from (A3) and the rough estimate $|W_{t,t}(\theta)| \le 4\|w\|_\infty^2$. Therefore,

$$\left|W_{s,t}(\widehat{\theta}_n)\;-\;W_{s,t}(\theta_{n,i})\right| \;\le\; C\;\left(\sqrt{\dot{\sigma}_{s-1}^2(\widehat{\theta}_{n,i,s})}\;+\;\sqrt{\dot{\sigma}_{t-1}^2(\widehat{\theta}_{n,i,t})}\right)\;\sqrt{\|\widehat{\theta}_n-\theta_{n,i}\|},$$

for some random $\widehat{\theta}_{n,i,s}$ between $\widehat{\theta}_n$ and $\theta_{n,i}$ ($s=1,\ldots,n$, $i=1,\ldots,M_n$). To deal with terms such as $\dot{\sigma}_{s-1}^2(\widehat{\theta}_{n,i,s})$, we define $\theta_{\max,n} = \theta_0 + \gamma_n n^{-1/2}(1,1,1)'$ and make use of the fact that $\sigma_t^2(\theta) \le \sigma_t^2(\theta_{\max,n})$ holds for all $\theta \in \bar{\Theta}_n$. Using $|\xi_{n,t}| \le 2\sqrt{n}$ and again $|W_{s,t}(\theta)| \le 4\|w\|_\infty^2$ we obtain that

$$\begin{aligned}
&\left|\frac{1}{n}\sum_{t=2}^{n}\xi_{n,t}\left\{\sum_{s=1}^{t-1}\xi_{n,s}\left(W_{s,t}(\widehat{\theta}_n)\;-\;W_{s,t}(\theta_{n,i})\right)\right\}\right| \\
&\quad\le\; n\,C\;\|w\|_\infty^{2-4\kappa}\sum_{t=1}^{n}\left(\sigma_{t-1}^2(\theta_{\max,n})\sup_{\theta\in\bar{\Theta}_n}\left\{\frac{\|\dot{\sigma}_{t-1}^2(\theta)\|}{\sigma_{t-1}^2(\theta)}\right\}\right)^{\kappa}\;\|\widehat{\theta}_n\;-\;\theta_{n,i}\|^{\kappa}
\end{aligned}$$

holds for all $\kappa \in (0,1/2)$. In view of Proposition 1 and (4.29) in Francq and Zakoïan (2004) we have $E_{\theta_0}\left(\sigma_{t-1}^2(\theta_{\max,n})\sup_{\theta\in\bar{\Theta}_n}\{\|\dot{\sigma}_{t-1}^2(\theta)\|/\sigma_{t-1}^2(\theta)\}\right)^{\kappa_0} \;<\; \infty$, for some $\kappa_0 \in (0,1/2)$. Hence, choosing the grid such that

$$\sup_{\theta\in\bar{\Theta}_n}\min_{1\le i\le M_n}\|\theta-\theta_{n,i}\|^{\kappa_0} \;=\; o(n^{-2})$$

we eventually obtain (6.14). Note that the grid can be chosen such that $M_n = O(n^m)$ for some $m \in \mathbb{N}$ which will be essential for the calculations below.

To prove (6.13), we first study the size of the quantities $n^{-1/2} \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i})$. Note that this is not a sum of martingale differences since the weights of $W_{s,t}(\theta_{n,i})$ depend on $I_{t-1}(\theta_{n,i})$ and $I_{t-1}(\theta_0)$. By an approximation of $I_{t-1}(\theta_{n,i})$ and $I_{t-1}(\theta_0)$ on a sufficiently fine grid with nonrandom points we obtain by the Bernstein-type inequality given in Proposition 2.1 in Freedman (1975) that

$$P\left( \max_{1\le i\le M_n} \max_{2\le t\le n} \left| n^{-1/2} \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right| > \nu_n \right) \underset{n\to\infty}{\longrightarrow} 0, \tag{6.16}$$

for $\nu_n = n^{-\delta}$ with arbitrary $\delta \in (0, 1/2 - \gamma)$. We define the stopping time

$$\tau = \inf \left\{ k: \quad \max_{1\le i\le M_n} \left| n^{-1/2} \sum_{s=1}^{k} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right| > \nu_n \right\}.$$

Then, (6.16) is equivalent to

$$P(\tau < n) \underset{n\to\infty}{\longrightarrow} 0. \tag{6.17}$$

Moreover, in case of $\tau \ge n$, we have, with $\mathcal{F}_t = \sigma(\sigma_t, Y_t, \sigma_{t-1}, Y_{t-1}, \dots)$,

$$\begin{aligned} V_n^2(\theta_{n,i}) &= \frac{1}{n^2} \sum_{t=2}^{n} E\left( \left( \xi_{n,t} \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right)^2 \middle| \mathcal{F}_t \right) \\ &\le \frac{1}{n} \sum_{t=2}^{n} \left( n^{-1/2} \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right)^2 E[\xi_{n,t}^2] \\ &\le \nu_n^2 \, \kappa, \end{aligned}$$

where $\kappa = E[(\varepsilon_1^2 - 1)^2]$. Now we obtain, again from the Bernstein-type inequality in Freedman (1975, Proposition 2.1) that

$$\begin{aligned} &P\left( \left| \frac{1}{n} \sum_{t=2}^{n} \xi_{n,t} \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right\} \right| > \nu_n \, a_n \quad \text{and } \tau \ge n \right) \\ &\le 2 \exp\left\{ - \frac{\nu_n^2 \, a_n^2}{2\,(2\nu_n^2 \, a_n + \nu_n^2 \, \kappa)} \right\}. \end{aligned} \tag{6.18}$$

Therefore, choosing $a_n = c\sqrt{\log n}$ with an appropriate $c$ we obtain that

$$\sum_{i=1}^{M_n} P\left( \left| \frac{1}{n} \sum_{t=2}^{n} \xi_{n,t} \left\{ \sum_{s=1}^{t-1} \xi_{n,s} W_{s,t}(\theta_{n,i}) \right\} \right| > \nu_n \, a_n \quad \text{and } \tau \ge n \right) = o(1). \tag{6.19}$$

Now, (6.17) and (6.19) imply (6.15).

Furthermore, (6.9), (6.10), (6.12), and (6.15) yield

$$\int R_{n,1}^2(z)\, Q(dz) = o_P(1). \tag{6.20}$$

The estimation of the remaining terms is much easier. A Taylor expansion gives

$$\begin{aligned} &\int R_{n,2}^2(z)\, Q(dz) \\ &\le \int \left\{ \frac{\|\widehat{\theta}_n - \theta_0\|}{\sqrt{n}} \sum_{t=1}^{n} \varepsilon_t^2 \frac{\|\dot{\sigma}_t^2(\bar{\theta}_{n,t})\|}{\sigma_t^2(\widehat{\theta}_n)} \left| w(z - I_{t-1}(\widehat{\theta}_n)) - w(z - I_{t-1}) \right| \right\}^2 Q(dz) \end{aligned}$$

with some random $\bar{\theta}_{n,t}$ between $\widehat{\theta}_n$ and $\theta_0$. Note that

$$E\left[\varepsilon_t^2 \frac{\|\dot{\sigma}_t^2(\bar{\theta}_{n,t})\|}{\sigma_t^2(\widehat{\theta}_n)} \mathbb{1}_{\|\widehat{\theta}_n-\theta\|\le\eta}\right] = E\left[\varepsilon_t^2 \frac{\|\dot{\sigma}_t^2(\bar{\theta}_{n,t})\|}{\sigma_t^2(\bar{\theta}_{n,t})} \frac{\sigma_t^2(\bar{\theta}_{n,t})}{\sigma_t^2(\theta_0)} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\widehat{\theta}_n)} \mathbb{1}_{\|\widehat{\theta}_n-\theta\|\le\eta}\right] < \infty$$

for some $\eta > 0$ by (4.29) in Francq and Zakoïan (2004) and

$$E\left[\sup_{\theta\in U(\theta_0)} \frac{\sigma_t^2(\theta)}{\sigma_t^2(\theta_0)}\right]^4 + E\left[\sup_{\theta\in U(\theta_0)} \frac{\sigma_t^2(\theta_0)}{\sigma_t^2(\theta)}\right]^4 < \infty, \tag{6.21}$$

where the latter relation can be deduced similarly to (4.26) in Francq and Zakoïan (2004) for sufficiently small $U(\theta_0)$. Hence, by (6.11) in conjunction with Lemma 2.1, we get

$$\int R_{n,2}^2(z)\, Q(dz) = o_P(1). \tag{6.22}$$

The application of a CLT for martingale differences, Theorem 2.3 of McLeish (1974), leads to

$$\begin{aligned} &\int R_{n,3}^2(z)\, Q(dz) \\ &= O_P(1) \int \left\| \frac{1}{n} \sum_{t=1}^n \left( \varepsilon_t^2 \frac{\dot{\sigma}_t^2(\theta_0)}{\sigma_t^2(\theta_0)}\, w(z - I_{t-1}) - E_{\theta_0}\left[\varepsilon_1^2 \frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)}\, w(z - I_0)\right]\right)\right\|_2^2 Q(dz) \\ &= o_P(1). \end{aligned} \tag{6.23}$$

The latter relation follows essentially from convergence of the empirical distribution based on $X_1, \ldots, X_n$ to $P^{X_0}$; see also the proof of equation (17) in Leucht and Neumann (2013a) for details. Since

$$\left(\frac{Y_t^2}{\sigma_t^2(\widehat{\theta}_n)} - \frac{Y_t^2}{\sigma_t^2(\theta_0)}\right) = -\varepsilon_t^2 \left(\frac{\dot{\sigma}_t^2(\bar{\theta}_t)}{\sigma_t^2(\widehat{\theta}_n)}\right)' (\widehat{\theta}_n - \theta_0)$$

we obtain

$$\begin{aligned} &\int R_{n,4}^2(z)\, Q(dz) \\ &\le \frac{\|\widehat{\theta}_n - \theta_0\|_2^2\, \|w\|_\infty^2}{n} \left(\sum_{t=1}^n \varepsilon_t^2 \sup_{\theta:\, \|\theta-\theta_0\|_2\le\|\widehat{\theta}_n-\theta_0\|_2} \left\|\frac{\dot{\sigma}_t^2(\theta)}{\sigma_t^2(\widehat{\theta}_n)} - \frac{\dot{\sigma}_t^2(\theta_0)}{\sigma_t^2(\theta_0)}\right\|_2\right)^2 \\ &= o_P(1) \end{aligned} \tag{6.24}$$

again by (4.29) in Francq and Zakoïan (2004) and (6.21) in conjunction with Lebesgue's dominated convergence theorem and consistency of $\widehat{\theta}_n$. Finally, by (2.8),

$$\begin{aligned} \int R_{n,5}^2(z)\, Q(dz) &\le \frac{\|\widehat{\theta}_n - \theta_0 - n^{-1}\sum_{s=1}^n L_s\|_2^2\, \|w\|_\infty^2}{n} \left\|\sum_{t=1}^n \varepsilon_t^2 \frac{\dot{\sigma}_t^2(\theta_0)}{\sigma_t^2(\theta_0)}\right\|_2^2 \\ &= o_P(1). \end{aligned} \tag{6.25}$$

We see from (6.20) and (6.22) to (6.25) that (6.7) actually holds true, which completes the proof. □

*Proof of Proposition 3.2.* (i) The overall structure of the proof is similar to the one of Lemma 3.1. Therefore we only sketch the proof and stress the main differences to the previous one. Within the proof of Lemma 3.1, we repeatedly apply relations (4.26) and (4.29) of Francq and Zakoïan (2004). Under the alternative, the observations do not arise from a GARCH(1,1) process. Therefore, these results are not applicable directly. Still, replacing $\theta_0$ by $\bar\theta_0$ these relations remain valid since the proofs of Francq and Zakoïan (2004) only rely on the definition of $\sigma_t^2(\theta),\ t\in\mathbb{Z},\theta\in\Theta_0$, and the moment conditions stated in (A5).

As under $\mathcal{H}_0$, the effect of choosing an arbitrary initial volatility $\widehat\sigma_0$ is asymptotically negligible. Using the notation of part (a) of the proof of Lemma 3.1, $n^{-1}\int S_{n,i}^2(z)\,Q(dz)$, $i=1,\,2$, can be verified in the same manner as before. Thus, we obtain negligibility of the effect of the starting value if additionally $n^{-1}\widetilde{T}_n = O_P(1)$, which follows from the calculations below.

We decompose the square root of the integrand of $n^{-1}\widetilde{T}_n$ defined in (6.4) as

$$\begin{aligned}
&\frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\sigma_t^2(\widehat\theta_n)}-1\right)w(z-I_{t-1}(\widehat\theta_n))\\
&= \frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\sigma_t^2(\bar\theta_0)}-1\right)w(z-I_{t-1}(\bar\theta_0))\\
&\quad+\frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\sigma_t^2(\bar\theta_0)}-1\right)\Big(w(z-I_{t-1}(\widehat\theta_n))\;-\;w(z-I_{t-1}(\bar\theta_0))\Big)\\
&\quad+\frac{1}{n}\sum_{t=1}^{n}\left(\frac{Y_t^2}{\sigma_t^2(\widehat\theta_n)}-\frac{Y_t^2}{\sigma_t^2(\bar\theta_0)}\right)w(z-I_{t-1}(\widehat\theta_n))\\
&= U_{n,1}(z)\,+\,U_{n,2}(z)\,+\,U_{n,3}(z).
\end{aligned}$$

Similarly to (6.23) we obtain $\int U_{n,1}^2(z)Q(dz)\overset{P}{\longrightarrow}\int\big(E[(Y_1^2/\sigma_1^2(\bar\theta_0)-1)\,w(z-I_0(\bar\theta_0))]\big)^2\,Q(dz)$. Moreover, under (A3) and (A5) we obtain

$$\int U_{n,2}^2(z)\,Q(dz)\ \le\ O_P(1)\,\frac{1}{n}\sum_{t=1}^{n}\Big\{1\wedge|\sigma_t^2(\widehat\theta_n)-\sigma_t^2(\bar\theta_0)|\Big\}.$$

In analogy to the proof of Lemma 2.1(ii), asymptotic negligibility of the remaining term can be verified. Finally, we get

$$|U_{n,3}(z)|\ \le\ o_P(1)\,\frac{\|w\|_\infty^2}{n}\sum_{t=1}^{n}\frac{Y_t^2}{\sigma_t^2(\bar\theta_0)}\,\frac{\|\dot\sigma_t^2(\bar\theta_{n,t})\|}{\sigma_t^2(\widehat\theta_n)}$$

for some random $\bar\theta_{n,t}$ between $\widehat\theta_n$ and $\bar\theta_0$. Thus, $\int U_{n,3}^2(z)Q(dz)$ is of order $o_P(1)$ under (A5) and finally $n^{-1}\widehat{T}_n=\int E[(Y_1^2/\sigma_1^2(\bar\theta_0)-1)w(z-I_0(\bar\theta_0))]^2Q(dz)+o_P(1)$.

(ii) The assertion follows immediately from part (i) and the extra condition presumed under (ii).

□

*Proof of Lemma 4.1.* Recall that $e_t$ and $\widehat{\varepsilon}_t$ denote the raw and the standardized residuals, respectively. We define

$$\begin{aligned} F_\varepsilon(x) &= P(\varepsilon_0 \le x), \\ F_{n,\varepsilon}(x) &= \frac{1}{n}\sum_{t=1}^n \mathbb{1}(\varepsilon_t \le x), \\ F_{n,e}(x) &= \frac{1}{n}\sum_{t=1}^n \mathbb{1}(e_t \le x), \\ F_{n,\widehat{\varepsilon}}(x) &= \frac{1}{n}\sum_{t=1}^n \mathbb{1}(\widehat{\varepsilon}_t \le x). \end{aligned}$$

First of all, we obtain from the Glivenko-Cantelli theorem that

$$d_L(F_{n,\varepsilon}, F_\varepsilon) \stackrel{a.s}{\longrightarrow} 0. \tag{6.26}$$

Next we will prove that

$$\frac{1}{n}\sum_{t=1}^n |e_t - \varepsilon_t| \stackrel{a.s.}{\longrightarrow} 0, \tag{6.27}$$

which then implies

$$d_L(F_{n,e}, F_{n,\varepsilon}) \stackrel{a.s.}{\longrightarrow} 0. \tag{6.28}$$

To this end, we split up

$$\begin{aligned} &\frac{1}{n}\sum_{t=1}^n |e_t - \varepsilon_t| \\ &\quad\le \frac{1}{n}\sum_{t=1}^n |\varepsilon_t| \left|\sigma_t(\theta_0)/\sigma_t(\widehat{\theta}_n) - 1\right| + \frac{1}{n}\sum_{t=1}^n |\varepsilon_t|\sigma_t(\theta_0)\left|\frac{1}{\widehat{\sigma}_t} - \frac{1}{\sigma_t(\widehat{\theta}_n)}\right|. \end{aligned}$$

It follows from (A2) and (4.26) of Francq and Zakoïan (2004) that $E[|\varepsilon_0| \sup_{\theta:\, \|\theta-\theta_0\|\le\delta} |\sigma_0(\theta_0)/\sigma_0(\theta) - 1|] < \infty$ for some $\delta > 0$. Therefore we obtain from Lemma 2.1 that $E[|\varepsilon_0| \sup_{\theta:\, \|\theta-\theta_0\|\le\delta} |\sigma_0(\theta_0)/\sigma_0(\theta) - 1|] \longrightarrow_{\delta\to 0} 0$. This implies, in conjunction with $\widehat{\theta}_n \stackrel{a.s.}{\longrightarrow} \theta_0$ and by the ergodic theorem (see e.g. Theorem 2.3 on page 48 in Bradley (2007)) that

$$P\left(\sup_{n\ge n_0} \frac{1}{n}\sum_{t=1}^n |\varepsilon_t||\sigma_t(\theta_0)/\sigma_t(\widehat{\theta}_n) - 1| > \varepsilon\right) \underset{n_0\to\infty}{\longrightarrow} 0$$

holds for all $\varepsilon > 0$. Therefore, we obtain that

$$\frac{1}{n}\sum_{t=1}^n |\varepsilon_t| \left|\sigma_t(\theta_0)/\sigma_t(\widehat{\theta}_n) - 1\right| \stackrel{a.s.}{\longrightarrow} 0. \tag{6.29}$$

Since $|\sigma_t^2(\widehat{\theta}_n) - \widehat{\sigma}_t^2| \le \widehat{\beta}_n^t |\sigma_0^2(\widehat{\theta}_n) - \widehat{\sigma}_0^2|$ and $\widehat{\sigma}_t^2 \ge \widehat{\omega}_n$ (for $t \ge 1$) we obtain

$$\frac{1}{n}\sum_{t=1}^n |\varepsilon_t|\sigma_t(\theta_0)\left|\frac{1}{\widehat{\sigma}_t} - \frac{1}{\sigma_t(\widehat{\theta}_n)}\right| \le \frac{1}{n}\sum_{t=1}^n |\varepsilon_t| \frac{\sigma_t(\theta_0)}{\sigma_t(\widehat{\theta}_n)} \, \frac{\widehat{\beta}_n^t |\sigma_0^2(\widehat{\theta}_n) - \widehat{\sigma}_0^2|}{\widehat{\omega}_n}.$$

Furthermore, since

$$\frac{|\sigma_0^2(\widehat{\theta}_n) - \widehat{\sigma}_0^2|}{\widehat{\omega}_n} \stackrel{a.s.}{\longrightarrow} \frac{|\sigma_0^2(\theta_0) - \widehat{\sigma}_0^2|}{\omega_0}$$

and

$$\widehat{\beta}_n \xrightarrow{a.s.} \beta_0 < 1$$

we conclude that

$$\frac{1}{n}\sum_{t=1}^{n} |\varepsilon_t| \sigma_t(\theta_0) \left| \frac{1}{\widehat{\sigma}_t} - \frac{1}{\sigma_t(\widehat{\theta}_n)} \right| \xrightarrow{a.s.} 0. \tag{6.30}$$

Next, (6.29) and (6.30) yield (6.27) and therefore also (6.28). Moreover, we obtain by the strong law of large numbers (SLLN) that $n^{-1}\sum_{t=1}^{n} \varepsilon_t^2 \xrightarrow{a.s.} E[\varepsilon_0^2] = 1$. In analogy to the proof of (6.27) it can be shown that

$$\frac{1}{n}\sum_{t=1}^{n} |e_t^2 - \varepsilon_t^2| \xrightarrow{a.s.} 0.$$

Therefore, we end up with

$$\frac{1}{n}\sum_{t=1}^{n} e_t^2 \xrightarrow{a.s.} 1. \tag{6.31}$$

Hence,

$$d_L(F_{n,\widehat{\varepsilon}}, F_{n,e}) \xrightarrow{a.s.} 0. \tag{6.32}$$

From (6.26), (6.28), and (6.32) we conclude that

$$d_L(F_{n,\widehat{\varepsilon}}, F_{\varepsilon}) \xrightarrow{a.s.} 0, \tag{6.33}$$

as required. □

*Proof of Lemma 4.2.* Recall that $\varepsilon_t^*$ has the distribution function $F_{n,\widehat{\varepsilon}}$. Since $E^*[\varepsilon_t^{*2}] = E[\varepsilon_t^2] = 1$ it follows from our Lemma 4.1 and Lemma 8.3 in Bickel and Freedman (1981) that

$$d_2(\varepsilon_t^*, \varepsilon_t) \xrightarrow{a.s.} 0, \tag{6.34}$$

where $d_2(U,V) = \inf\{E(\widetilde{U} - \widetilde{V})^2 : \widetilde{U} \stackrel{d}{=} U, \widetilde{V} \stackrel{d}{=} V\}$ denotes Mallows' distance between the random variables $U$ and $V$. Since $(\varepsilon_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t^*)_{t\in\mathbb{Z}}$ are both sequences of i.i.d. random variables, we can construct, on an appropriate probability space $(\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P})$, a sequence of i.i.d. random vectors $((\widetilde{\varepsilon}_t^*, \widetilde{\varepsilon}_t)')_{t\in\mathbb{Z}}$ such that

$$\begin{aligned} \widetilde{\varepsilon}_t &\stackrel{d}{=} \varepsilon_t, \\ \widetilde{\varepsilon}_t^* &\stackrel{d}{=} \varepsilon_t^*, \end{aligned}$$

and

$$E_{\widetilde{P}}[(\widetilde{\varepsilon}_t^* - \widetilde{\varepsilon}_t)^2] = d_2(\varepsilon_t^*, \varepsilon_t) \xrightarrow{a.s.} 0. \tag{6.35}$$

The proof of

$$E_{\widetilde{P}}\left[ |\widetilde{Y}_{t-1}^* - \widetilde{Y}_{t-1}| \wedge 1 + |\widetilde{\sigma}_{t-1}^{*2} - \widetilde{\sigma}_{t-1}^2| \wedge 1 \right] \xrightarrow{a.s.} 0 \tag{6.36}$$

is more delicate since we have to deal here with infinite series. Let $\varepsilon > 0$ be arbitrary. We define approximations

$$\begin{aligned} \widetilde{\sigma}_{t,K}^2 &= \omega_0 \left[ 1 + \sum_{k=1}^{K} \prod_{i=1}^{k} (\beta_0 + \alpha_0 \widetilde{\varepsilon}_{t-i}^2) \right], \\ \widetilde{\sigma}_{t,K}^{*2} &= \widehat{\omega}_n \left[ 1 + \sum_{k=1}^{K} \prod_{i=1}^{k} (\widehat{\beta}_n + \widehat{\alpha}_n \widetilde{\varepsilon}_{t-i}^{*2}) \right]. \end{aligned}$$

It follows from (6.35) and $\widehat{\theta}_n \xrightarrow{a.s.} \theta_0$ that

$$\widetilde{P}\left(|\widetilde{\sigma}^2_{t,K} - \widetilde{\sigma}^{*2}_{t,K}| > \varepsilon\right) \xrightarrow{a.s.} 0$$

holds for all $K \in \mathbb{N}$. Since the infinite series defining $\widetilde{\sigma}^2_t$ converges, we have

$$\widetilde{P}\left(|\widetilde{\sigma}^2_t - \widetilde{\sigma}^2_{t,K}| > \varepsilon\right) \leq \varepsilon$$

if $K = K(\varepsilon)$ is sufficiently large. Furthermore, since $E^*[\ln(\widehat{\beta}_n + \widehat{\alpha}_n \varepsilon_t^{*2})] \xrightarrow{a.s.} E[\ln(\beta_0 + \alpha_0 \varepsilon_0^2)] < 0$, we have that

$$\widetilde{P}\left(|\widetilde{\sigma}^{*2}_t - \widetilde{\sigma}^{*2}_{t,K^*}| > \varepsilon\right) \leq \varepsilon, \quad a.s.$$

for all $n \geq N$, where $K^*$ is sufficiently large and nonrandom and $N$ is random but finite. For $\bar{K} = \max\{K, K^*\}$, we obtain that

$$\widetilde{P}\left(|\widetilde{\sigma}^2_t - \widetilde{\sigma}^{*2}_t| > 3\,\varepsilon\right) \leq 3\,\varepsilon, \quad a.s.,$$

for all $n$ larger than some random but finite value. This, however, implies that

$$E_{\widetilde{P}}\left[|\widetilde{\sigma}^{*2}_{t-1} - \widetilde{\sigma}^2_{t-1}| \wedge 1\right] \xrightarrow{a.s.} 0.$$

The proof of the fact,

$$E_{\widetilde{P}}\left[|\widetilde{Y}^*_{t-1} - \widetilde{Y}_{t-1}| \wedge 1\right] \xrightarrow{a.s.} 0 \tag{6.37}$$

is similar and therefore omitted.

It remains to establish a coupling of $(\widetilde{L}^*_t)_t$ and $(\widetilde{L}_t)_t$. To this end, we first show that $E^*[\ddot{W}^*_0(\widehat{\theta}_n)] \xrightarrow{a.s.} E_{\theta_0}[\ddot{W}_0(\theta_0)]$ with $W^*_0(\theta) := \log \sigma_0^{*2}(\theta) + Y_0^{*2}/\sigma_0^{*2}(\theta)$ and $\sigma_t^{*2}(\theta) = \omega/(1-\beta) + \alpha \sum_{k=1}^{\infty} \beta^{k-1} Y^{*2}_{t-k}$. In accordance with (4.13) in Francq and Zakoïan (2004) we get

$$\ddot{W}^*_t(\theta) := \left(1 - \frac{Y_t^{*2}}{\sigma_t^{*2}(\theta)}\right) \frac{1}{\sigma_t^{*2}(\theta)} \frac{\partial^2 \sigma_t^{*2}(\theta)}{\partial\theta\partial\theta'} + \left(2\frac{Y_t^{*2}}{\sigma_t^{*2}(\theta)} - 1\right) \frac{\partial \sigma_t^{*2}(\theta)}{\partial\theta} \frac{\partial \sigma_t^{*2}}{\partial\theta'}.$$

We obtain explicit formulas for the derivatives appearing in the formula above by substituting the original random variables by their bootstrap counterparts in the equations (4.15), (4.16), and (4.20) to (4.22) of that paper. They depend on the parameters and lagged $Y^*_s$ in a smooth manner and we get the desired convergence by dominated convergence theorem from almost sure convergence of $\widehat{\theta}_n$ to $\theta$, (6.37) and from the bootstrap analogues to (4.25) in Francq and Zakoïan (2004) and to Lemma 2.3 in Berkes, Horváth, and Kokoszka (2003).

Similarly we obtain

$$E_{\widetilde{P}}\left\|\dot{\widetilde{\sigma}}^{*2}_t(\widehat{\theta}_n)/\widetilde{\sigma}^{*2}_t(\widehat{\theta}_n) - \dot{\widetilde{\sigma}}^2_t(\theta_0)/\widetilde{\sigma}^2_t(\theta_0)\right\|^2 \xrightarrow{a.s.} 0$$

which in turn leads to

$$E_{\widetilde{P}}\|\widetilde{L}^*_t - \widetilde{L}_t\|^2 \xrightarrow{a.s.} 0. \tag{6.38}$$

This finally completes the proof. □

*Proof of Corollary 4.1.* We have

$$\begin{aligned}\widetilde{\sigma}^2_t(\theta) &= \left(\frac{\omega}{1-\beta} + \alpha \sum_{k=1}^{K} \beta^{k-1} \widetilde{Y}^2_{t-k}\right) + \alpha \sum_{k=K+1}^{\infty} \beta^{k-1} \widetilde{Y}^2_{t-k} \\ &=: \widetilde{\sigma}^2_{t,K}(\theta) + R_{t,K}(\theta)\end{aligned}$$

and, analogously,

$$\begin{aligned}\widetilde{\sigma}_t^{*2}(\theta) &= \left(\frac{\omega}{1-\beta} + \alpha\sum_{k=1}^{K}\beta^{k-1}\widetilde{Y}_{t-k}^{*2}\right) + \alpha\sum_{k=K+1}^{\infty}\beta^{k-1}\widetilde{Y}_{t-k}^{*2} \\ &=: \widetilde{\sigma}_{t,K}^{*2}(\theta) + R_{t,K}^{*}(\theta).\end{aligned}$$

Since $\beta \leq \rho_0 < 1$ and $\alpha \leq u_2 < \infty$ for all $(\omega, \alpha, \beta)' \in \Theta_0$ we obtain from Lemma 4.2, for arbitrary $K < \infty$,

$$\begin{aligned}&E_{\widetilde{P}}\left[\sup_{\theta\in\Theta_0}\left|\widetilde{\sigma}_{t,K}^{*2}(\theta) - \widetilde{\sigma}_{t,K}^{2}(\theta)\right| \wedge 1/3\right] \\ &\leq u_2\sum_{k=1}^{K}E_{\widetilde{P}}\left[\left|\widetilde{Y}_{t-k}^{*2} - \widetilde{Y}_{t-k}^{2}\right| \wedge 1/3\right] \stackrel{a.s.}{\longrightarrow} 0.\end{aligned} \tag{6.39}$$

To estimate the remainder terms $R_{t,K}(\theta)$ and $R_{t,K}^{*}(\theta)$, we choose any $\zeta \in (1, 1/\rho_0)$. Then

$$\begin{aligned}\sum_{k=K+1}^{\infty}\widetilde{P}(\widetilde{Y}_{t-k}^{2} > \zeta^k) &= \sum_{k=1}^{\infty}P(Y_0^2 > \zeta^{K+k}) \\ &= \sum_{k=1}^{\infty}P\left(\frac{2\log Y_0}{\log\zeta} - K > k\right) \\ &\leq E\left[\left(\frac{2\log Y_0}{\log\zeta} - K\right)_+\right],\end{aligned}$$

which tends to zero as $K \to \infty$. On the other hand, if $\widetilde{Y}_{t-k}^{2} \leq \zeta^k$ for all $k > K$, we obtain that

$$\sup_{\theta\in\Theta_0}R_{t,K}(\theta) \leq u_2\sum_{k=K+1}^{\infty}\rho_0^{k-1}\zeta^k = (\rho_0\zeta)^K\frac{u_2\zeta}{1-\rho_0\zeta}.$$

Since this upper estimate also tends to zero as $K \to \infty$, we conclude

$$E_{\widetilde{P}}\left[\sup_{\theta\in\Theta_0}R_{t,K} \wedge 1/3\right] \underset{K\to\infty}{\longrightarrow} 0. \tag{6.40}$$

Finally, since

$$E^*\left[\left(\frac{2\log Y_0^*}{\log\zeta} - K\right)_+\right] \stackrel{P}{\longrightarrow} E\left[\left(\frac{2\log Y_0}{\log\zeta} - K\right)_+\right]$$

we get in analogy to (6.40) that

$$E_{\widetilde{P}}\left[\sup_{\theta\in\Theta_0}R_{t,K}^{*} \wedge 1/3\right] \stackrel{\widetilde{P}}{\longrightarrow} 0. \tag{6.41}$$

The assertion follows now from (6.39) to (6.41). □

*Proof of Lemma 4.3.* The assertion can be proved along the lines of the proof of Theorem 2.2 in Francq and Zakoïan (2004) if $\widehat{\theta}_n^* - \widehat{\theta}_n = o_{P^*}(1)$. Therefore we proceed in two steps.

(i) *Weak consistency of $\widehat{\theta}_n^*$.*

Let $\epsilon > 0$ be arbitrary. By strong consistency of $\widehat{\theta}_n$ for $\theta_0$, it suffices to show that

$$P^*\left(\widehat{\theta}_n^* \in U_\epsilon\right) \xrightarrow{P} 0, \tag{6.42}$$

where $U_\epsilon = \{\theta\colon \|\theta - \theta_0\| \geq \epsilon\} \cap \Theta_0$.

Before we deduce this conclusion via coupling arguments, some preliminary considerations involving the behavior of the log-likelihood process $(\bar{\mathcal{L}}_n(\theta))_{\theta\in\Theta_0}$ on the original side are in order. It can be seen from the proof of Theorem 2.1 in Francq and Zakoïan (2004) that, for all $\theta \neq \theta_0$, there exist sufficiently small $\eta(\theta) > 0$, $\delta(\theta) > 0$ and a sufficiently large $M(\theta) < \infty$ such that, for $U(\theta) = \{\bar{\theta}\colon \|\bar{\theta} - \theta\| < \eta(\theta)\} \cap \Theta_0$,

$$E_{\theta_0}\left[\inf_{\bar{\theta}\in U(\theta)} \left(\ln\sigma_t^2(\bar{\theta}) \,+\, Y_t^2/\sigma_t^2(\bar{\theta})\right) \wedge M(\theta)\right] \;\geq\; E_{\theta_0}\left[\ln\sigma_0^2(\theta_0) \,+\, 1\right] \,+\, \delta(\theta). \tag{6.43}$$

Since the set $U_\epsilon$ is a compact subset of $\mathbb{R}^3$ and is covered by the open sets $\{\bar{\theta}\colon \|\bar{\theta}-\theta\| < \eta(\theta)\}$, $\theta \in U_\epsilon$, we can extract a finite subcover, that is, there exist $\theta^{(1)}, \ldots, \theta^{(N)} \in U_\epsilon$ such that

$$U_\epsilon \subseteq \bigcup_{i=1}^N U(\theta^{(i)}).$$

Let $\mathcal{L}_{n,M}(\theta) = n^{-1}\sum_{t=1}^n \{(\ln\sigma_t^2(\theta) + Y_t^2/\sigma_t^2(\theta)) \wedge M(\theta)\}$. Since the underlying process is strictly stationary and ergodic we obtain by the ergodic theorem, for $M = \max\{M(\theta^{(1)}), \ldots, M(\theta^{(N)})\}$,

$$\begin{aligned} &P_{\theta_0}\left(\inf_{\theta\in U_\epsilon} \mathcal{L}_{n,M}(\theta) \;\geq\; \frac{1}{n}\sum_{t=1}^n \left(\ln\sigma_t^2(\theta_0) \,+\, Y_t^2/\sigma_t^2(\theta_0)\right)\right) \qquad (6.44)\\ &\leq \quad \sum_{i=1}^N P_{\theta_0}\left(\frac{1}{n}\sum_{t=1}^n \inf_{\bar{\theta}\in U(\theta^{(i)})} \left\{\ln\sigma_t^2(\bar{\theta}) \,+\, Y_t^2/\sigma_t^2(\bar{\theta})\right\} \wedge M \;\geq\; \frac{1}{n}\sum_{t=1}^n \left(\ln\sigma_t^2(\theta_0) \,+\, Y_t^2/\sigma_t^2(\theta_0)\right) \,+\, \frac{\delta(\theta)}{2}\right)\\ &\underset{n\to\infty}{\longrightarrow} \quad 0. \end{aligned}$$

We now show that the log-likelihood process $\bar{\mathcal{L}}_n^*$, given by

$$\bar{\mathcal{L}}_n^*(\theta) \;=\; \frac{1}{n}\sum_{t=1}^n \ln\bar{\sigma}_t^{*2}(\theta) \,+\, \frac{Y_t^{*2}}{\bar{\sigma}_t^{*2}(\theta)},$$

behaves in a similar manner as $\bar{\mathcal{L}}_n$. Here, $\bar{\sigma}_t^{*2}(\theta)$ denotes the bootstrap analogue to $\bar{\sigma}_t^2(\theta)$. Since we have to deal with a triangular scheme on the bootstrap side, we do not have tools such as the ergodic theorem at hand and a direct imitation of the consistency proof from the original side would be presumably rather cumbersome. Fortunately, some coupling arguments can be employed to complete the proof in a simple manner.

Recall that $\widehat{\theta}_n^*$ is defined as a minimizer of $\bar{\mathcal{L}}_n^*$. In the following we show that

$$P^*\left(\inf_{\theta\in U_\epsilon} \bar{\mathcal{L}}_n^*(\theta) \;>\; \bar{\mathcal{L}}_n^*(\widehat{\theta}_n)\right) \xrightarrow{P} 1, \tag{6.45}$$

which then implies (6.42).

Let $\mathcal{L}_n^*(\theta) = n^{-1}\sum_{t=1}^n \ln\sigma_t^{*2}(\theta) + Y_t^{*2}/\sigma_t^{*2}(\theta)$ be the analogue to $\bar{\mathcal{L}}_n^*(\theta)$, where only $\bar{\sigma}_t^{*2}(\theta)$ is replaced by the stationary approximation $\sigma_t^{*2}(\theta)$. We can prove in complete analogy to (i) in the proof of Theorem 2.1 in Francq and Zakoïan (2004) that

$$\sup_{\theta\in\Theta_0} \left|\bar{\mathcal{L}}_n^*(\theta) \,-\, \mathcal{L}_n^*(\theta)\right| \;=\; o_{P^*}(1). \tag{6.46}$$

Next we prove that, with $\widetilde{\mathcal{L}}_n^*(\widehat{\theta}_n) = n^{-1}\sum_{t=1}^n \ln\widetilde{\sigma}_t^{*2}(\widehat{\theta}_n) + \widetilde{Y}_t^{*2}/\widetilde{\sigma}_t^{*2}(\widehat{\theta}_n)$,

$$\left|\widetilde{\mathcal{L}}_n^*(\widehat{\theta}_n) \,-\, \widetilde{\mathcal{L}}_n(\theta_0)\right| \stackrel{\widetilde{P}}{\longrightarrow} 0. \tag{6.47}$$

We split up

$$\begin{aligned}
&\left|\widetilde{\mathcal{L}}_n^*(\widehat{\theta}_n) \,-\, \widetilde{\mathcal{L}}_n(\theta_0)\right| \\
\le\ & \frac{1}{n}\sum_{t=1}^n \left|\widetilde{\varepsilon}_t^{*2} \,-\, \widetilde{\varepsilon}_t^2\right| \\
&+ \frac{1}{n}\sum_{t=1}^n \left|\Big(\ln\widetilde{\sigma}_t^{*2}(\widehat{\theta}_n)\wedge M\Big) \,-\, \Big(\ln\widetilde{\sigma}_t^{2}(\widehat{\theta}_n)\wedge M\Big)\right| \\
&+ \frac{1}{n}\sum_{t=1}^n \Big(\ln\widetilde{\sigma}_t^{2}(\widehat{\theta}_n) \,-\, M\Big)_+ \\
&+ \frac{1}{n}\sum_{t=1}^n \Big(\ln\widetilde{\sigma}_t^{*2}(\widehat{\theta}_n) \,-\, M\Big)_+ \\
&+ \frac{1}{n}\sum_{t=1}^n \left|\ln\widetilde{\sigma}_t^{2}(\widehat{\theta}_n) \,-\, \ln\widetilde{\sigma}_t^{2}(\theta_0)\right| \\
=\ & T_{n,1} \,+\, \cdots \,+\, T_{n,5}.
\end{aligned}$$

It follows from Lemma 4.2 that

$$T_{n,1} \,+\, T_{n,2} \stackrel{\widetilde{P}}{\longrightarrow} 0.$$

We obtain from monotone convergence that

$$E_{\widetilde{P}}\left[\sup_{\theta:\ \|\theta-\theta_0\|\le\delta} (\ln\widetilde{\sigma}_t^2(\theta) - M)_+\right] \mathop{\longrightarrow}_{M\to\infty} 0, \tag{6.48}$$

for some $\delta > 0$. Since $\widehat{\theta}_n \stackrel{a.s.}{\longrightarrow} \theta_0$ we conclude that

$$T_{n,3} \stackrel{\widetilde{P}}{\longrightarrow} 0.$$

Furthermore, it can be deduced similarly to Theorem 3 in Nelson (1991) that $E^*\widetilde{\sigma}_0^{*s}(\widehat{\theta}_n) \le C$ with probability tending to one for some $s > 0$ and $C < \infty$. Hence, using Lemma 4.2 and (6.48)

$$T_{n,4} \stackrel{\widetilde{P}}{\longrightarrow} 0.$$

Finally, it follows from $E[\sup_{\theta:\ \|\theta-\theta_0\|\le\delta} |\ln\sigma_0^2(\theta) - \ln\sigma_0^2(\theta_0)|] \to_{\delta\to 0} 0$ and $\widehat{\theta}_n \stackrel{a.s.}{\longrightarrow} \theta_0$ that

$$T_{n,5} \stackrel{\widetilde{P}}{\longrightarrow} 0,$$

which completes the proof of (6.47).

Define $\widetilde{\mathcal{L}}_{n,M}^*(\theta) \,=\, n^{-1}\sum_{t=1}^n \Big(\ln\widetilde{\sigma}_t^{*2}(\theta) \,+\, \widetilde{Y}_t^{*2}/\widetilde{\sigma}_t^{*2}(\theta)\Big) \wedge M$. We obtain from Corollary 4.1 that

$$\sup_{\theta\in\Theta_0}\left|\widetilde{\mathcal{L}}_{n,M}^*(\theta) \,-\, \widetilde{\mathcal{L}}_{n,M}(\theta)\right| \stackrel{\widetilde{P}}{\longrightarrow} 0. \tag{6.49}$$

From (6.43), (6.44), and (6.46) to (6.49) we obtain (6.45) and therefore (6.42).

(ii) *Linearization of* $\widehat{\theta}_n^*$.

With the same arguments as in the proof of Theorem 2.2 of Francq and Zakoïan (2004) we obtain

$$(E^*[\ddot{W}_0^*(\widehat{\theta}_n)])^{-1}\frac{1}{n}\sum_{t=1}^{n}\left(\ddot{W}_t^*(\bar{\theta}_{n,i,j})\right)_{i,j=1,2,3}\left(\widehat{\theta}_n^* - \widehat{\theta}_n\right) = \frac{1}{n}\sum_{t=1}^{n} L_t^* \,+\, o_{P^*}(n^{-1/2})$$

for some $\bar{\theta}_{n,i,j}$ between $\widehat{\theta}_n$ and $\widehat{\theta}_n^*$ and it remains to prove that

$$\frac{1}{n}\sum_{t=1}^{n}\left(\ddot{W}_t^*(\bar{\theta}_{n,i,j})\right)_{i,j=1,2,3} - E^*[\ddot{W}_0^*(\widehat{\theta}_n)] = o_{P^*}(1).$$

By part (i) of this proof and the bootstrap analogue to (iii) of the proof of Theorem 2.2 in Francq and Zakoïan (2004) we get

$$\frac{1}{n}\sum_{t=1}^{n}\left(\ddot{W}_t^*(\bar{\theta}_{n,i,j})\right)_{i,j=1,2,3} - E^*[\ddot{W}_0^*(\widehat{\theta}_n)] = \frac{1}{n}\sum_{t=1}^{n}\ddot{W}_t^*(\widehat{\theta}_n) - E^*[\ddot{W}_0^*(\widehat{\theta}_n)] \,+\, o_{P^*}(1).$$

Asymptotic negligibility of the r.h.s. follows from $n^{-1}\sum_{t=1}^{n}\ddot{W}_t(\theta_0) - E[\ddot{W}_0(\theta_0)] = o_P(1)$ (see (vi) of the proof of Francq and Zakoïan (2004)), and $E^*[\ddot{W}_0^*(\widehat{\theta}_n)] \xrightarrow{a.s.} E[\ddot{W}_0(\theta_0)]$ if additionally

$$\frac{1}{n}\sum_{t=1}^{n}\ddot{W}_t^*(\widehat{\theta}_n) - \ddot{W}_t(\theta_0) = o_{P^*}(1).$$

This in turn can be deduced from Lemma 4.2 in conjunction with (4.20) to (4.22) in Francq and Zakoïan (2004). □

*Proof of Lemma 4.4.* The proof can be carried out in complete analogy to the verification of Lemma 3.1 if the bootstrap counterparts of the assumptions (A1) and (A2) are satisfied since the Bahadur linearization of the estimator (2.8) is valid by Lemma 4.3. In particular, we have to show an analogue of (6.23) on the bootstrap side. This will essentially follow from convergence of the empirical bootstrap distribution to $P^{X_0}$. To show the latter we can use our Lemma 4.1 which implies that the difference of the empirical bootstrap distribution and the empirical distribution given by $X_1, \ldots, X_n$ is asymptotically negligible. Finally, convergence of the empirical distribution to $P^{X_0}$ follows from the ergodic theorem. Now we check the prerequisites (A1) and (A2).

Obviously, (A1)(i) is satisfied with $\widehat{\theta}_n$ instead of $\theta_0$. The bootstrap innovations $(\varepsilon_t^*)_t$ are i.i.d. and have unit second moment. Moreover, $E^*[\ln(\widehat{\beta}_n + \widehat{\alpha}_n\varepsilon_1^{*2})] < 0$ with probability tending to one which then yields the bootstrap version of (A1)(ii).

Next, we consider the bootstrap analogue of (A2). Clearly, by Lemma 4.1 we obtain a nondegenerate distribution of the $\varepsilon_t^{*2}$s with probability tending to one. Finally, we have to show that $E^*[|\varepsilon_0^*|^4] \le C$ with probability tending to one for a finite constant $C$. In view of (6.31), we get

$$E^*[|\varepsilon_t^*|^4] \,=\, |1 + o_P(1)|n^{-1}\sum_{n=1}^{n} e_t^4 \,=\, |1 + o_P(1)|n^{-1}\sum_{n=1}^{n}\varepsilon_t^4 \,+\, o_P(1),$$

where the latter relation can be verified similarly to (6.27). Now the SLLN implies the desired boundedness of moments. □

*Proof of Theorem 4.1.* The coupling of the $X_t$ with the $X_t^*$ constructed in the proof of Lemma 4.2 implies a coupling of $T_n$ and $T_n^*$. Denote by $\widetilde{T}_n$ and $\widetilde{T}_n^*$ the versions of $T_n$ and $T_n^*$ based on the coupled variables $\widetilde{X}_t$ and $\widetilde{X}_t^*$, respectively. First, note that the kernels of the $V$-statistics are of the form $h(x,y) = \int g(x,z)g(y,z)\,Q(dz)$ and $h^*(x,y) = \int g^*(x,z)g^*(y,z)\,Q(dz)$, resp., where

$$\begin{aligned} g(x,z) &= (x_1-1)w(z_1-x_2, z_2-x_3) \;-\; E_{\theta_0}\left[\frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)} w(z_1 - Y_0, z_2 - \sigma_0^2(\theta_0))\right]' x_4, \\ g^*(x,z) &= (x_1-1)w(z_1-x_2, z_2-x_3) \;-\; E_{\widehat{\theta}_n}\left[\frac{\dot{\sigma}_1^{*2}(\widehat{\theta}_n)}{\sigma_1^{*2}(\widehat{\theta}_n)} w(z_1 - Y_0^*, z_2 - \sigma_0^{*2}(\widehat{\theta}_n))\right]' x_4. \end{aligned}$$

We obtain by Minkowski's inequality that

$$\begin{aligned} &\left|\sqrt{\widetilde{T}_n} \;-\; \sqrt{\widetilde{T}_n^*}\right| \\ &\quad= \left| \left\{ \int \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n g(\widetilde{X}_t, z) \right)^2 Q(dz) \right\}^{1/2} - \left\{ \int \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n g^*(\widetilde{X}_t^*, z) \right)^2 Q(dz) \right\}^{1/2} \right| \\ &\quad\le \left\{ \int \left( \frac{1}{\sqrt{n}} \sum_{t=1}^n [g(\widetilde{X}_t, z) \;-\; g^*(\widetilde{X}_t^*, z)] \right)^2 Q(dz) \right\}^{1/2}. \end{aligned}$$

It follows from the independence of the vectors $(\widetilde{\varepsilon}_t, \widetilde{\varepsilon}_t^*)'$ that, for $t > s$,

$$E_{\widetilde{P}}\left( g(\widetilde{X}_t, z) \;-\; g^*(\widetilde{X}_t^*, z) \mid \widetilde{X}_s, \widetilde{X}_s^* \right) \;=\; 0 \quad a.s.$$

This implies

$$\begin{aligned} &E_{\widetilde{P}}\left(\sqrt{\widetilde{T}_n} \;-\; \sqrt{\widetilde{T}_n^*}\right)^2 \\ &\quad\le \frac{1}{n} \sum_{s,t=1}^n E_{\widetilde{P}} \left\{ \int [g(\widetilde{X}_s, z) \;-\; g^*(\widetilde{X}_s^*, z)][g(\widetilde{X}_t, z) \;-\; g^*(\widetilde{X}_t^*, z)]\, Q(dz) \right\} \\ &\quad= E_{\widetilde{P}} \int [g(\widetilde{X}_t, z) \;-\; g^*(\widetilde{X}_t^*, z)]^2\, Q(dz) \\ &\quad\le 2\, E_{\widetilde{P}} \int \left\{ (\widetilde{\varepsilon}_1^2 - 1)\, w(z_1 - \widetilde{Y}_0, z_2 - \widetilde{\sigma}_0^2) \;-\; (\widetilde{\varepsilon}_1^{*2} - 1)\, w(z_1 - \widetilde{Y}_0^*, z_2 - \widetilde{\sigma}_0^{*2}) \right\}^2 Q(dz) \\ &\qquad + 2\, E_{\widetilde{P}} \int \left\{ E_{\theta_0}\left[\frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)} w(z_1 - Y_0, z_2 - \sigma_0^2(\theta_0))\right]' \widetilde{L}_1 \;-\; E_{\widehat{\theta}_n}\left[\frac{\dot{\sigma}_1^{*2}(\widehat{\theta}_n)}{\sigma_1^{*2}(\widehat{\theta}_n)} w(z_1 - Y_0^*, z_2 - \sigma_0^{*2}(\widehat{\theta}_n))\right]' \widetilde{L}_1^* \right\}^2 Q(dz). \end{aligned}$$

The assertion of the Theorem now follows from Lemma 4.2 and

$$\int \left( E_{\widehat{\theta}_n}\left[\frac{\dot{\sigma}_1^{*2}(\widehat{\theta}_n)}{\sigma_1^{*2}(\widehat{\theta}_n)} w(z_1 - Y_0^*, z_2 - \sigma_0^{*2}(\widehat{\theta}_n))\right] \;-\; E_{\theta_0}\left[\frac{\dot{\sigma}_1^2(\theta_0)}{\sigma_1^2(\theta_0)} w(z_1 - Y_0, z_2 - \sigma_0^2(\theta_0))\right] \right)^2 Q(dz) \;\stackrel{a.s.}{\longrightarrow}\; 0,$$

where the latter follows from (A3) and the proof of Lemma 4.2. □

*Proof of Lemma 4.5.* We split up

$$
\begin{aligned}
n^{-1}\widehat{T}_n^* & \\
= \quad & \int \left\{ \frac{1}{n} \sum_{t=1}^{n} \left( \frac{Y_t^{*2}}{(\widehat{\sigma}_t^*)^2} - 1 \right) w(z_1 - Y_{t-1}^*, z_2 - \widehat{\sigma}_{t-1}^{*2}) \right\}^2 Q(dz_1, dz_2) \\
\leq \quad & 3 \int \left\{ \frac{1}{n} \sum_{t=1}^{n} (\varepsilon_t^{*2} - 1) w(z_1 - Y_{t-1}^*, z_2 - \sigma_{t-1}^{*2}) \right\}^2 Q(dz_1, dz_2) \\
& + 3 \int \left\{ \frac{1}{n} \sum_{t=1}^{n} \left( \frac{Y_t^{*2}}{(\widehat{\sigma}_t^*)^2} - \varepsilon_t^{*2} \right) w(z_1 - Y_{t-1}^*, z_2 - \widehat{\sigma}_{t-1}^{*2}) \right\}^2 Q(dz_1, dz_2) \\
& + 3 \int \left\{ \frac{1}{n} \sum_{t=1}^{n} (\varepsilon_t^{*2} - 1) \left( w(z_1 - Y_{t-1}^*, z_2 - \widehat{\sigma}_{t-1}^{*2}) - w(z_1 - Y_{t-1}^*, z_2 - \sigma_{t-1}^{*2}) \right) \right\}^2 Q(dz_1, dz_2) \\
=: \quad & R_{n,1}^* + R_{n,2}^* + R_{n,3}^*.
\end{aligned}
$$

The sum in the integrand of $R_{n,1}^*$ is a sum of martingale differences. Therefore, and since $\varepsilon_t^*$ is independent of $(Y_{t-1}^*, \sigma_{t-1}^{*2})$ we have

$$
E^* R_{n,1}^* = \frac{3}{n} E^*[(\varepsilon_1^{*2} - 1)^2] \int w^2(z_1 - Y_{t-1}^*, z_2 - \sigma_{t-1}^{*2}) \, Q(dz_1, dz_2) = O_P(n^{-1}).
$$

Moreover, note that

$$
R_{n,2}^* \leq 6 \int \left\{ \frac{1}{n} \sum_{t=1}^{n} \left( \frac{Y_t^{*2}}{\sigma_t^{*2}(\widehat{\theta}_n^*)} - \varepsilon_t^{*2} \right) w(z_1 - Y_{t-1}^*, z_2 - \sigma_t^{*2}(\widehat{\theta}_n^*)) \right\}^2 Q(dz_1, dz_2) + o_P(1).
$$

We abbreviate the first summand on the r.h.s. by $\bar{R}_{n,2}^*$. By the Cauchy-Schwarz inequality, $\sigma_t^{*2}(\widehat{\theta}_n^*) \geq \widehat{\omega}^* \geq u_1 > 0$ and since $\varepsilon_t^*$ is independent of $(\sigma_t^{*2}(\widehat{\theta}_n^*), \sigma_t^{*2})$ we obtain

$$
E^* \bar{R}_{n,2}^* \leq 12 \, E^* \left[ \varepsilon_1^{*4} \frac{|\sigma_t^{*2}(\widehat{\theta}_n^*) - \sigma_1^{*2}|^2}{\sigma_t^{*4}(\widehat{\theta}_n^*)} \right] \|w\|_\infty^2 \xrightarrow{P} 0.
$$

Similarly, $R_{n,3}^* = \bar{R}_{n,3}^* + o_P(1)$ with

$$
\bar{R}_{n,3}^* = 6 \int \left\{ \frac{1}{n} \sum_{t=1}^{n} (\varepsilon_t^{*2} - 1) \left( w(z_1 - Y_{t-1}^*, z_2 - \sigma_t^{*2}(\widehat{\theta}_n^*)) - w(z_1 - Y_{t-1}^*, z_2 - \sigma_{t-1}^{*2}) \right) \right\}^2 Q(dz_1, dz_2).
$$

Finally, by (A3) and again by the Cauchy-Schwarz inequality,

$$
\begin{aligned}
E^* \bar{R}_{n,3}^* \quad & \leq \quad 6 \, E^*[(\varepsilon_1^{*2} - 1)^2] \, E^* \left[ \int \left( w(z_1 - Y_0^*, z_2 - \sigma_t^{*2}(\widehat{\theta}_n^*)) - w(z_1 - Y_0^*, z_2 - \sigma_0^{*2}) \right)^2 Q(dz_1, dz_2) \right] \\
& \xrightarrow{P} \quad 0,
\end{aligned}
$$

which completes the proof. □

*Proof of Corollary 4.2.* We prove only (i) since (ii) follows directly from Proposition 3.2 and Lemma 4.5. Since

$$
\left| P(\widehat{T}_n > t_\gamma^*) - \gamma \right| \leq \sup_{t \in \mathbb{R}} \left| P(\widehat{T}_n \leq t) - P^*(\widehat{T}_n^* \leq t) \right| + \left| P^*(\widehat{T}_n^* > t_\gamma^*) - \gamma \right|,
$$

part (i) of the corollary follows from Proposition 3.1, Lemma 4.4, and Theorem 4.1 if the limit distribution of $\widehat{T}_n$ is continuous; see also Lemma 2.11 of van der Vaart (1998). The limit variable $Z = \sum_k \lambda_k Z_k^2$ has indeed a continuous distribution if at least one of the $\lambda_k$s is nonzero which in turn follows from

$$E[h(X_1, X_1)] > 0. \tag{6.50}$$

□

*Acknowledgment* . The authors are grateful to Fabian Tinkl, Friedrich-Alexander Universität Erlangen-Nürnberg, for providing an R-code for QMLE based on GARCH(1,1) processes. This research was partly funded by the German Research Foundation DFG, project NE 606/2-2.

## References

Anderson, T. W. and Darling, D. A. (1954). A test of goodness of fit. *Journal of the American Statistical Association* **49**, 765–769.

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

Berkes, I., Horváth, L., and Kokoszka, P. (2004). A weighted goodness-of-fit test for GARCH(1,1) specification. *Lithuanian Mathematical Journal* **44**, 3–22.

Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9**, 1196–1217.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

Bradley, R. C. (2007). *Introduction to Strong Mixing Conditions, Volume I.* Kendrick Press.

Dehling, H. and Mikosch, T. (1994). Random quadratic forms and the bootstrap for $U$-statistics. *Journal of Multivariate Analysis* **51**, 392–413.

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

Escanciano, J. C. (2006). Goodness-of-fit tests for linear and nonlinear time series models. *Journal of the American Statistical Association* **101**, 531–541.

Escanciano, J. C. (2007a). Model checks using residual marked empirical processes. *Statistica Sinica* **17**, 115–138.

Escanciano, J. C. (2007b). Joint and marginal diagnostic tests for conditional mean and variance specifications. *CAEPR Working Paper* **2007-009**.

Escanciano, J. C. (2008). Joint and marginal specification tests for conditional mean and variance models. *Journal of Econometrics* **143**, 74–87.

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

Francq, C. and Zakoïan, J.-M. (2010). *GARCH Models: Structure, Statistical Inference and Financial Applications.* Chichester, West Sussex: Wiley.

Freedman, D. A. (1975). On tail probabilities for martingales. *The Annals of Probability* **3**, 100–118.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* **48**, 1779–1801.

Kristensen, D. and Linton, O. (2006). A closed-form estimator for the GARCH(1,1)-model. *Econometric Theory* **22**, 323–337.

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

Leucht, A. and Neumann, M. H. (2009). Consistency of general bootstrap methods for degenerate $U$- and $V$-type statistics. *Journal of Multivariate Analysis* **100**, 1622–1633.

Leucht, A. and Neumann, M. H. (2013a). Degenerate $U$- and $V$-statistics under ergodicity: Asymptotics, bootstrap and applications in statistics. *Annals of the Institute of Statistical Mathematics* **65**, 349–386.

Leucht, A. and Neumann, M. H. (2013b). Dependent wild bootstrap for degenerate $U$- and $V$-statistics. *Journal of Multivariate Analysis* **117**, 257–280.

McLeish, D. L. (1974). Dependent central limit theorems and invariance principles. *The Annals of Probability* **2**, 620–628.

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

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* **59**, 347–370.

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Sentana, E. (1995). Quadratic ARCH models. *Review of Economic Studies* **62**, 639–661.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models*. Lecture Notes in Statistics **181**, Berlin, Heidelberg: Springer.

Tinkl, F. (2013). Quasi-maximum likelihood estimation in generalized polynomial autoregressive conditional heteroscedasticity models. *IWQW Discussion Paper Series* 03/2013.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge: Cambridge University Press.