Yule–Walker type estimators in GARCH(1,1) models: Asymptotic normality and bootstrap

Gisela Maercker Martin Moser

Abstract

We investigate GARCH(1,1) processes and first prove their stability. Using the representation of the squared GARCH model as an ARMA model we then consider Yule–Walker type estimators for the parameters of the GARCH(1,1) model and derive their asymptotic normality. We use a residual bootstrap to define bootstrap estimators for the Yule–Walker estimates and prove the consistency of this bootstrap method. Some simulation results will demonstrate the small sample behaviour of the bootstrap procedure.

1 Introduction

Many time series exhibit non-constant conditional variance (conditional heteroskedasticity). Nonlinear processes capable of modelling such volatility have come into particular interest in time series analysis, especially in econometrics.

Conditional heteroskedasticity can be modelled by processes of the form

$$X_t = \sigma_t \varepsilon_t, \quad t = 0, 1, 2, \dots \tag{1.1}$$

where the innovations $\{\varepsilon_t\}$ are independent identically distributed (i.i.d.) random variables with mean zero and unit variance, and the volatility σ_t describes the change of (conditional) variance.

In financial time series such as stock returns or foreign exchange rates, volatility clustering has been observed for a long time, i.e. periods of large price changes are followed by periods of small price changes. This phenomenon can be modelled by autoregressive conditional heteroskedastic (ARCH) models introduced by Engle (1982), where the conditional variance σ_t^2 is a linear function of the squared past observations.

Bollerslev (1986) proposed the generalized ARCH or GARCH model by including also lagged values of σ_t^2 in the conditional variance equation. The GARCH(p, q) model is defined by (1.1) with

$$\sigma_t^2 = \alpha_0 + \sum_{j=1}^{p} \alpha_j X_{t-j}^2 + \sum_{i=1}^{q} \beta_i \sigma_{t-i}^2, \quad \alpha_0 > 0,\ \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q \geq 0,\ \alpha_p > 0, \beta_q > 0.$$

GARCH modelling allows a more flexible lag structure than ARCH models and often permits a more parsimonious parametrization. Especially the GARCH(1,1) model has served as an appropriate model in many applications.

Over the past years, many semiparametric and nonparametric approaches to the ARCH model have been studied. For a variety of other extensions and applications of the ARCH model, we refer to the survey article by Bollerslev, Chou and Kroner (1992) and Engle (1995).

In the present work we investigate parameter estimation and bootstrap in GARCH(1,1) models. Rewriting the GARCH(1,1) model it becomes obvious that the squared process can be represented as an ARMA(1,1) (autoregressive moving-average) process. According to this Yule–Walker type estimators for the parameters of the GARCH(1,1) processes can be defined (Bollerslev (1986)). We prove stability of the GARCH(1,1) model and then derive asymptotic normality of the Yule–Walker estimators.

The bootstrap is a method for estimating or approximating the distribution of a statistic and its characteristics based on a resampling of the observed data. Since its introduction by Efron for models with i.i.d. observations, there have been many applications and extensions of the bootstrap principle, also in case of dependent data. Bootstrap methods for e.g. nonparametric ARCH models have been studied by Franke, Kreiss and Mammen (1997), but they do not cover the GARCH case. A wild bootstrap method for quasi-maximum likelihood estimators of GARCH(1,1) models is proposed in Maercker (1998).

We use a residual bootstrap to define bootstrap estimators for the Yule–Walker estimates and prove the consistency of the bootstap procedure. Some simulation results demonstrate the small sample behaviour of the bootstrap method.

The paper is organized as follows. In Section 2 we state sufficient conditions for the stability of GARCH(1,1) models. In Section 3 we define Yule–Walker type estimators for the parameters in GARCH(1,1) processes and derive asymptotic normality of these estimators. In Section 4 we construct bootstrap estimators for the Yule–Walker estimators. We show that this bootstrap method works in the sense that it is consistent. In Section 5 some simulation results for the bootstrap method are presented. All proofs are deferred to the Appendix. Some additional tools and notations about Markov chain theory and mixing which are used for the proofs are also provided in the Appendix.

2 Stability of the GARCH(1,1) model

Assume that we are given observations $X_1, \ldots, X_n$ from the heteroskedastic model

$$X_t = \sigma_t \varepsilon_t, \quad t \in \mathbb{Z}, \tag{2.2}$$

with innovations

$$\{\varepsilon_t\} \text{ i.i.d }, \ E\varepsilon_0 = 0, \ E\varepsilon_0^2 = 1, \tag{2.3}$$

and conditional variance

$$\sigma_t^2 = \omega + \alpha X_{t-1}^2 + \beta \sigma_{t-1}^2, \quad t \in \mathbb{Z}, \quad \text{ where } \omega, \alpha, \beta > 0. \tag{2.4}$$

The process defined by (2.2) and (2.4) is called GARCH(1,1) (generalized autoregressive conditional heteroskedastic) process (Bollerslev (1986)).

Combining (2.4) and (2.2) the conditional variance may be written as

$$\sigma_t^2 = \omega + \sigma_{t-1}^2(\beta + \alpha \varepsilon_{t-1}^2). \tag{2.5}$$

Iterating σ_t in equation (2.4) and (2.5) respectively yields for $h \geq 1$

$$\sigma_t^2 = \omega \sum_{k=0}^{h-1} \beta^k + \alpha \sum_{k=0}^{h-1} \beta^k X_{t-1-k}^2 + \beta^h \sigma_{t-h}^2, \tag{2.6}$$

$$\sigma_t^2 = \omega \sum_{k=0}^{h-1} \prod_{i=1}^{k} \left(\beta + \alpha \varepsilon_{t-i}^2\right) + \sigma_{t-h}^2 \prod_{i=1}^{h} \left(\beta + \alpha \varepsilon_{t-i}^2\right), \tag{2.7}$$

where, as usual, empty products are set equal to one.

We make the following stability assumption on the model.

Assumption S

(S1) $\alpha + \beta < 1$.

(S2) The distribution G of ε_0 has a Lebesgue density g which is positive and continuous.

By Jensen's inequality, (S1) implies

$$E\left[\ln(\beta + \alpha \varepsilon_0^2)\right] < 0. \tag{2.8}$$

As is shown in Nelson (1990), condition (2.8) is necessary and sufficient for the existence of a unique stationary solution of (2.5), which then is given by the infinite series

$$\sigma_t^2 = \omega \sum_{k=0}^{\infty} \prod_{i=1}^{k} (\beta + \alpha \varepsilon_{t-i}^2), \quad t \in \mathbb{Z}. \tag{2.9}$$

In the following we will assume that $\{\sigma_t^2\}$ is the stationary solution of (2.5) and hence may be represented by (2.9). In particular, as $E\varepsilon_t^2 = 1$, we have

$$\mu = EX_t^2 = E\sigma_t^2 = \frac{\omega}{1-(\alpha+\beta)} < \infty. \tag{2.10}$$

Furthermore, $E\sum_{h=0}^{\infty}\beta^h\sigma_{t-h}^2 = \frac{\mu}{1-\beta} < \infty$ implies $\beta^h\sigma_{t-h}^2 \to 0$ as $h \to \infty$ almost surely, thus (2.6) may be extended to

$$\sigma_t^2 = \frac{\omega}{1-\beta} + \alpha\sum_{k=0}^{\infty}\beta^k X_{t-1-k}^2, \quad t \in \mathbb{Z}.$$

Set $\sigma(x,s) = (\omega + \alpha x^2 + \beta s^2)^{\frac{1}{2}}$. Then

$$Y_t := (X_t, \sigma_t)' = \sigma(Y_{t-1})(\varepsilon_t, 1)', \quad t \in \mathbb{Z},$$

is a bivariate Markov process with state space $\mathbb{R} \times [\sqrt{\frac{\omega}{1-\beta}}, \infty)$. With the help of a drift criterion we will establish geometric ergodicity and absolute regularity for this process. Whereas stability of ARCH processes was investigated before by e.g. Guegan and Diebolt (1992), Doukhan (1994), and Borkovec and Klüppelberg (1998), the results do not cover the GARCH case.

For a definition and discussion of geometric ergodicity and absolute regularity we refer to the Appendix.

Lemma 2.1 *Let Assumption S hold. Then the process $\{Y_t\}$ is geometrically ergodic and absolutely regular. Furthermore, there exist constants $c > 0$ and $\rho > 1$ such that the β-mixing coefficients of $\{Y_t\}$ satisfy $\beta_k \le c\rho^{-k}$, $k \in \mathbb{N}$.*

We conclude this section with an ARMA-representation of the squared process $\{X_t^2\}$ (cf. Bollerslev (1986)) which will be used frequently in the sequel. Set

$$\eta_t = X_t^2 - \sigma_t^2 = \sigma_t^2(\varepsilon_t^2 - 1).$$

Then, by (2.4), we have

$$X_t^2 = \omega + (\alpha+\beta)X_{t-1}^2 - \beta\eta_{t-1} + \eta_t. \tag{2.11}$$

Therefore $\{X_t^2\}$ is an ARMA(1,1) process with parameters $\alpha+\beta$ and $-\beta$ and innovations $\{\eta_t\}$. Defining $\mathcal{F}_t$ as the σ-field generated by $\{\varepsilon_s : s \le t\}$ we note

$$E[\eta_t \mid \mathcal{F}_{t-1}] = \sigma_t^2 E[\varepsilon_t^2 - 1] = 0,$$

thus the innovations $\{\eta_t\}$ form a martingale difference sequence.

3 Yule–Walker type estimators – definition and asymptotic normality

Consider the centered squared GARCH process $\{X_t^2 - \mu\}$ in the ARMA-representation

$$\left(X_t^2 - \mu\right) - (\alpha + \beta)\left(X_{t-1}^2 - \mu\right) = \eta_t - \beta\eta_{t-1}, \tag{3.12}$$

$\eta_t = \sigma_t^2(\varepsilon_t^2 - 1)$, derived from (2.11). As is well established in ARMA models the empirical autocovariances of the process can be used to obtain Yule–Walker (YW) type estimators for the parameters μ, $\alpha+\beta$, and β, respectively. Observe that the squared process $\{X_t^2\}$ exhibits autocorrelation whereas the process itself is not correlated over time.

Let us assume that $EX_0^4 < \infty$. Conditions for the existence of moments are given in Bollerslev (1986) and Nelson (1990). As we will need even more stringent conditions for the proof of asymptotic normality of the YW type estimators the discussion of those conditions will be postponed, see Remark 3.2 below.

Set $\sigma_\eta^2 = E\eta_t^2$. Recalling $E[\eta_t|\mathcal{F}_{t-1}] = 0$ we note that $E\eta_t X_t^2 = \sigma_\eta^2$, $E\eta_t X_{t+1}^2 = \alpha\sigma_\eta^2$ and $E\eta_t X_s^2 = 0$ for $s < t$. For the derivation of suitable identities involving the covariances $\gamma_h = \mathrm{Cov}(X_0^2, X_h^2)$, $h \geq 0$, we now proceed in the usual way, see also Bollerslev (1986, 1988), by multiplying both sides in (3.12) with X_{t-h}^2, $h = 0,1,\ldots$, and computing expectations. This yields the following identities.

$$\begin{aligned} \gamma_0 - (\alpha+\beta)\gamma_1 &= (1-\beta\alpha)\sigma_\eta^2, & (3.13)\\ \gamma_1 - (\alpha+\beta)\gamma_0 &= -\beta\sigma_\eta^2, & (3.14)\\ \gamma_h - (\alpha+\beta)\gamma_{h-1} &= 0, \qquad h \geq 2. & (3.15) \end{aligned}$$

Elimination of σ_η^2 gives the system

$$\begin{aligned} \alpha + \beta &= \frac{\gamma_2}{\gamma_1}, & (3.16)\\ \beta^{-1} - \alpha &= \frac{(\alpha+\beta)\gamma_1 - \gamma_0}{\gamma_1 - (\alpha+\beta)\gamma_0} = \frac{\gamma_2 - \gamma_0}{\gamma_1 - (\alpha+\beta)\gamma_0}. & (3.17) \end{aligned}$$

Using the empirical moments $\hat{\mu} = \frac{1}{n}\sum_{t=1}^n X_t^2$ and $\hat{\gamma}_h = \frac{1}{n}\sum_{t=1}^{n-h}(X_t^2 - \hat{\mu})(X_{t+h}^2 - \hat{\mu})$ we get the following YW-type estimators

$$\begin{aligned} (\widehat{\alpha+\beta})_n &= \frac{\hat{\gamma}_2}{\hat{\gamma}_1}, & (3.18)\\ (\widehat{\beta^{-1}-\alpha})_n &= \frac{\hat{\gamma}_2 - \hat{\gamma}_0}{\hat{\gamma}_1 - (\widehat{\alpha+\beta})_n\hat{\gamma}_0}, & (3.19)\\ \hat{\omega}_n &= \hat{\mu}\left(1 - (\widehat{\alpha+\beta})_n\right). & (3.20) \end{aligned}$$

In order to derive estimators $\hat{\alpha}_n$ and $\hat{\beta}_n$ of α and β, we set

$$\hat{\beta}_n^{-1} + \hat{\beta}_n = (\widehat{\alpha + \beta})_n + (\widehat{\beta^{-1} - \alpha})_n. \tag{3.21}$$

Denoting the right-hand side of (3.21) by $\hat{c}_n$ we obtain $\hat{\beta}_n^2 - \hat{c}_n\hat{\beta}_n + 1 = 0$. Hence, if $\hat{c}_n \geq 2$, we set

$$\hat{\beta}_n = \hat{c}_n/2 - \sqrt{\hat{c}_n^2/4 - 1}$$

so that $0 < \hat{\beta}_n \leq 1$ and $\hat{\beta}_n < 1$ if $\hat{c}_n > 2$. In practice it might happen that $\hat{c}_n < 2$, then set $\hat{\beta}_n = 0$. But by construction and the ergodic theorem $\hat{c}_n$ is a consistent estimate for $1/\beta + \beta$ and therefore, almost surely, $\hat{c}_n > 2$ for sufficiently large n. Finally we define

$$\hat{\alpha}_n = (\widehat{\alpha + \beta})_n - \hat{\beta}_n.$$

Again, by the ergodic theorem, $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n, \hat{\omega}_n)'$ is a consistent estimator for $\theta = (\alpha, \beta, \omega)'$.

Next, under additional moment assumptions, the asymptotic normality of $\hat{\theta}_n$ will be shown. We may use this to construct confidence sets for θ. However, the normal distribution is only an approximation to the exact distribution of θ_n. An alternative approach which often yields a better approximation is the bootstrap. Bootstrap confidence intervals are obtained by replacing the unknown distribution with its bootstrap estimator. We shall introduce a bootstrap method and study the consistency of this procedure.

Before doing so it should be mentioned that, as well in GARCH as in ARMA models, the YW estimators are less efficient than maximum likelihood (ML) estimators, or, depending on the error distribution, quasi-maximum likelihood (QML) estimators which are commonly used in practice. Indeed, simulation experiments underline this fact. Under normality of the innovations, ML estimators outperform the YW estimators. This is not necessarily true for QML estimators in case of non-normal innovation distributions.

QML estimates are found by an iterative procedure where the YW estimates may be used as initial estimates. In contrast to this in the YW type estimation procedure the observed data are used in a more direct way. This estimation procedure will be imitated by an appropriate bootstrap technique. A bootstrap method for QML estimators is discussed in Maercker (1997, 1998).

For the derivation of asymptotic normality of $\hat{\theta}_n$ we shall make use of a central limit theorem (CLT) for strongly mixing processes.

Theorem 3.1 *Let Assumption S hold and suppose* $E|X_0|^{8+\delta} < \infty$ *for some* $\delta > 0$. *Then*

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{D} \mathcal{N}\left((0,0,0)', \Sigma\right) \tag{3.22}$$

if Σ *is positive definite where* $\Sigma = D_2 D_1 \tilde{\Sigma} D_1' D_2'$,

$$D_1 = \begin{pmatrix} 1 & 0 & 0 \\ -\frac{1-\alpha\beta}{\beta^2}\frac{\gamma_0}{\sigma_\eta^2} & 1 & 0 \\ -\mu & 0 & 1 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 1-\zeta & -\zeta & 0 \\ \zeta & \zeta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \zeta = \frac{1}{2} - \frac{(\beta^{-1}+\beta)}{4\sqrt{(\beta^{-1}+\beta)^2/4-1}},$$

and $\tilde{\Sigma} = E\eta_0^2 Z_0 Z_0'$ *with*

$$\begin{aligned} Z_{t,1} &= \gamma_1^{-1}\left[-\beta X_{t-1}^2 + X_{t-2}^2 - (1-\beta)\mu\right], \\ Z_{t,2} &= (\gamma_1 - (\alpha+\beta)\gamma_0)^{-1}\Big[\beta(1-\beta^2)\eta_{t-1} + \left(\beta^2(\alpha+\beta) + \alpha - \beta - \beta^{-1}\right) X_{t-1} \\ &\quad + X_{t-2}^2 + \left(\beta^{-1} - \alpha\right)(1-\beta)\mu + \omega(1-\beta)\Big], \\ Z_{t,3} &= 1 - \beta. \end{aligned}$$

Remark 3.2 As $\hat{\theta}_n$ is based on the empirical autocovariances $\hat{\gamma}_h$, i.e. on lagged empirical fourth moments of X_t, the need for the rather stringent moment condition $E|X_0|^{8+\delta} < \infty$ in Theorem 3.1 becomes obvious. Under Assumption S expansion (2.9) holds and, by Minkowski's inequality, for any $p > 0$ a sufficient condition for $E|X_0|^{2p} < \infty$ is given by $E\left(\beta + \alpha\varepsilon_0^2\right)^p < 1$. This condition is also necessary, see Nelson (1990). For the case of normally distributed innovations ε_t the restrictions on the parameter space implied by $EX_0^8 < \infty$ and $EX_0^{10} < \infty$ are, among others, illustrated in Figure 3.1 of Bollerslev (1986).

4 The bootstrap procedure

We now discuss a bootstrap method for estimating the distribution of $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$. We use a residual bootstrap to construct bootstrap estimators. It will be shown that the (conditional) distribution of these bootstrap estimators converges in probability to the same asymptotic distribution as given in Theorem 3.1 for the original estimators, that is, the bootstrap procedure is (weakly) consistent.

Given a sample $X_1, \ldots, X_n$, the bootstrap process $\{X_t^*\}$ will be of the form

$$X_t^* = \sigma_t^* \varepsilon_t^*, \quad \sigma_t^{*2} = \hat{\omega} + \hat{\alpha} X_{t-1}^{*2} + \hat{\beta}\sigma_{t-1}^{*2}, \quad t \in \mathbb{Z}, \tag{4.23}$$

$$\{\varepsilon_t^*\} \text{ i.i.d },\ E^*\varepsilon_0^* = 0,\ E\varepsilon_0^{*2} = 1, \tag{4.24}$$

where the distribution G^* of ε_0^* is an estimate of the distribution G of ε_0 and E^* denotes the conditional expectation $E[\cdot|X_1,\ldots,X_n]$. The distribution $\mathcal{L}(\sqrt{n}(\hat{\theta}-\theta))$ will then be approximated by the (conditional) distribution $\mathcal{L}^*(\sqrt{n}(\hat{\theta}^*-\hat{\theta}))$ where $\hat{\theta}^*=(\hat{\alpha}^*,\hat{\beta}^*,\hat{\omega}^*)$ is calculated in the same way as $\hat{\theta}$, with $X_1,\ldots,X_n$ replaced by $X_1^*,\ldots,X_n^*$. For notational simplicity here and later the index n indicating the dependence of the estimators and the bootstrap process on the number of observations will be omitted.

In detail, the construction of $\{X_t^*\}$ consists of the following steps. Compute the Yule-Walker estimate $\hat{\theta}$ as described in Section 2. Set $\hat{\sigma}_0^2=\hat{\mu}$ and define

$$\hat{\sigma}_t^2=\hat{\omega}+\hat{\alpha}X_{t-1}^2+\hat{\beta}\hat{\sigma}_{t-1}^2, \quad t=1,\ldots,n,$$

or equivalently,

$$\hat{\sigma}_t^2=\hat{\omega}\sum_{k=0}^{t-1}\hat{\beta}^k+\hat{\alpha}\sum_{k=0}^{t-1}\hat{\beta}^kX_{t-1-k}^2+\hat{\beta}^t\hat{\mu}, \quad t=1,\ldots,n. \tag{4.25}$$

Calculate empirical residuals

$$\hat{\varepsilon}_t=\frac{X_t}{\hat{\sigma}_t}, \quad t=1,\ldots,n,$$

and let $\hat{G}(x)=\frac{1}{n}\sum_{t=1}^n 1\{\hat{\varepsilon}_t\le x\}$ denote their empirical distribution. In view of Assumption S smooth $\hat{G}$ by convolution and set $\tilde{G}=\hat{G}*\mathcal{N}(0,h^2)$ where $h=n^{-\frac{1}{5}}$. Define the distribution G^* of ε_0^* as the standardized form of $\tilde{G}$, i.e. $G^*(x)=\tilde{G}\left(\sigma_{\tilde{G}}x+\mu_{\tilde{G}}\right)$ where $\mu_{\tilde{G}}=\frac{1}{n}\sum_{t=1}^n\hat{\varepsilon}_t$ and $\sigma_{\tilde{G}}^2=\frac{1}{n}\sum_{t=1}^n\left(\hat{\varepsilon}_t-\mu_{\tilde{G}}\right)^2+h^2$ are the mean and variance of $\tilde{G}$, respectively. As $\hat{\theta}\to\theta$ almost surely we may assume

$$\hat{\omega},\hat{\alpha},\hat{\beta}>0 \text{ and } \hat{\alpha}+\hat{\beta}<1. \tag{4.26}$$

Finally, define the bootstrap GARCH process $\{X_t^*\}$ as the stationary solution of (4.23). In particular, we have as in (2.9)

$$\sigma_t^{*2}=\hat{\omega}\sum_{k=0}^{\infty}\prod_{i=1}^{k}\left(\hat{\beta}+\hat{\alpha}\varepsilon_{t-i}^{*2}\right), \quad t\in\mathbb{Z}. \tag{4.27}$$

Furthermore, by construction and (4.26), $\{X_t^*\}$ fulfills Assumption S. Hence, the conclusions of Lemma 2.1 apply and $Y_t^*=(X_t^*,\sigma_t^*)'$ is a geometrically ergodic and absolutely regular process with exponential decay of the mixing coefficients.

The density g^* of the bootstrap innovations is given by

$$g^*(x)=\frac{\sigma_{\tilde{G}}}{nh}\sum_{t=1}^n\varphi\left(\frac{\sigma_{\tilde{G}}x+\mu_{\tilde{G}}-\hat{\varepsilon}_t}{h}\right)$$

where $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$. Hence, g^* may be understood as a standardized kernel estimate of g with kernel φ and bandwidth h. We have chosen $h = n^{-\frac{1}{5}}$ as the rate common in kernel smoothing.

For the consistency proof of the bootstrap proposal we need the fact that, at least on average, $\hat{\sigma}_t^2$ is a good estimate for σ_t^2, as well as the consistency of g^* and the moments of the bootstrap process. Let $\|f\|_\infty = \sup_{x \in \mathbb{R}} |f(x)|$ for $f\colon \mathbb{R} \to \mathbb{R}$.

Lemma 4.1 *Let Assumption S hold and suppose $E|X_0|^{2p} < \infty$ for some $p > 4$.*
(a) For any $r \geq 1$

$$\frac{1}{n} \sum_{t=1}^{n} \left| \hat{\sigma}_t^2 - \sigma_t^2 \right|^r = O_P\left(n^{-\frac{p \wedge r}{2}} \right).$$

(b)

$$\mu_{\tilde{G}} = O_P\left(n^{-\frac{1}{2}} \right), \quad \left| \sigma_{\tilde{G}}^2 - 1 \right| = O_P\left(n^{-\frac{1}{2}} \right).$$

(c) If g is uniformly continuous then

$$\|g^* - g\|_\infty = o_P(1).$$

(d) For $q \in (0, 2p)$, $k_i \in \{0, 2\}$ and $t_i \in \mathbb{Z}$, $i = 1, \ldots, 4$,

$$E^* \left| \varepsilon_0^* \right|^q \to E \left| \varepsilon_0 \right|^q, \quad E^* \prod_{i=1}^{4} \sigma_{t_i}^{*k_i} \to E \prod_{i=1}^{4} \sigma_{t_i}^{k_i}, \quad E^* \prod_{i=1}^{4} X_{t_i}^{*k_i} \to E \prod_{i=1}^{4} X_{t_i}^{k_i},$$

in probability. Furthermore, there is a constant $c > 0$ such that for any subsequence $(k) \subset \mathbb{N}$ there exists a subsequence $(k_\ell) \subset (k)$ such that almost surely

$$\limsup_{\ell \to \infty} E^* \sigma_0^{*q} \leq c, \quad \limsup_{\ell \to \infty} E^* \left| X_0^* \right|^q \leq c.$$

In particular,

$$E^* \sigma_0^{*q} = O_P(1), \quad E^* \left| X_0^* \right|^q = O_P(1).$$

As already observed, the Markov process $\{Y_t^*\}$ is β-mixing with exponential decay of the mixing coefficients. The parameters $\hat{\theta}$ and g^* determining the process $\{Y_t^*\}$ converge in probability to θ and g, respectively, and so even more can be said about the β-mixing coefficients $\beta_n^*(j) = \beta^*(j)$, $j \in \mathbb{N}$, of $\{Y_t^*\}$. This is done in the following theorem where we find it convenient to phrase arguments concerning convergence in probability in terms of almost sure convergence along subsequences.

Theorem 4.2 *Let Assumption S hold. Suppose that* $EX_0^4 < \infty$ *and that* g *is uniformly continuous. Then for any subsequence* $(k) \subset I\!N$ *there exist a subsequence* $(k_\ell) \subset (k)$ *and constants* $C_b > 0$ *and* $\rho_b > 1$ *such that almost surely*

$$\beta^*_{k_\ell}(j) \le C_b \rho_b^{-j} \qquad \text{for all } \ell, j \in I\!N.$$

As a corollary we obtain the consistency of the bootstrap estimators in the sense of, for example, $P^*\left(|\hat{\mu}^* - \mu| > \epsilon\right) = o_P(1)$ for all $\epsilon > 0$. This is not immediate from the ergodic theorem applied to the bootstrap process $\{Y_t^*\}$, as for each sample $X_1, \ldots, X_n$ there is a different process $\{Y_t^*\}$ under consideration.

Corollary 4.3 *Let Assumption S hold and suppose* $E\,|X_0|^{8+\delta} < \infty$ *for some* $\delta > 0$. *Then* $\hat{\mu}^*$, $\hat{\gamma}_h^*$, $h \ge 0$, *and* $\hat{\theta}^*$ *are consistent for* μ, γ_h, $h \ge 0$, *and* θ *in the sense mentioned above.*

Now we are ready for the main result of this section, which states the consistency of the proposed bootstrap procedure.

Theorem 4.4 *Let Assumption S hold. Suppose that* $E\,|X_0|^{8+\delta} < \infty$ *for some* $\delta > 0$ *and that* g *is uniformly continuous. Then*

$$\mathcal{L}^*\left(\sqrt{n}\left(\hat{\theta}^* - \hat{\theta}\right)\right) \xrightarrow{D} \mathcal{N}\left((0,0,0)', \Sigma\right) \quad \text{in prabability} \tag{4.28}$$

if Σ *as defined in Theorem* 3.1 *is positive definite.*

5 Simulations

In order to illustrate the performance of the bootstrap procedures described in the preceding section, we show some results of simulation experiments. We simulate GARCH(1,1) processes of length $n = 1000$ with standard normal error distribution and with parameter $\theta = (\alpha, \beta, \omega)$. The parameter is estimated by the Yule–Walker type estimator. We repeat this procedure to estimate the distribution of the YW estimator. More specifically, the distribution of the standardized estimator is approximated by the estimated density calculated from 2500 Monte Carlo replications. Then the bootstrap approximation of this distribution is calculated. To this aim we calculate the YW estimator $\hat{\theta}_n$ from one simulated GARCH(1,1) process of length $n = 1000$. Based on $\hat{\theta}_n$ we generate 2500 bootstrap processes and calculate the bootstrap YW estimator for each bootstrap sample of length $n = 1000$. The estimated density for the standardized bootstrap estimator calculated from the 2500 bootstrap replications is plotted against the distribution of the original YW estimator.

It should be remarked that the YW estimation procedure is not very stable if (α, β) are chosen near the admissible parameter space with respect to the moment condition.

We show the results for $\theta = (0.1, 0.4, 0.1)$. Figure 1 compares the distribution of $\sqrt{n}(\hat{\alpha}_n - \alpha)$ with the bootstrap distribution of $\sqrt{n}(\hat{\alpha}_n^* - \hat{\alpha}_n)$. Note that the bootstrap procedure is based on only one (randomly chosen) sample of the underlying GARCH process.

Figure 2 and Figure 3 show the results for the parameters β and ω, respectively.

A Appendix

A.1 Proofs

Proof of Lemma 2.1: We will show that $\{Y_t\}$ is ϕ-irreducible with ϕ being the Lebesgue measure restricted to $\mathbb{R} \times [\sqrt{\frac{\omega}{1-\beta}}, \infty)$, aperiodic, and that compact sets are small. Then we shall apply the drift criterion given in Theorem A.4 to obtain geometric ergodicity.

In order to avoid a parameter-dependent state space and with an eye to situations where $\{Y_t\}$ may be started non-stationarily at time $t = 0$, we will formulate the proof for the state space $\mathbb{R} \times \mathbb{R}^+$.

Let $A \subset \mathbb{R} \times [\sqrt{\frac{\omega}{1-\beta}}, \infty)$ be measurable with $\lambda(A) > 0$, λ denoting the Lebesgue measure. Without loss of generality we may assume $A \subset \mathbb{R} \times [\sqrt{\frac{\omega}{1-\beta} + \epsilon}, \infty)$ for some $\epsilon > 0$.

We will show that, given $y \in \mathbb{R} \times \mathbb{R}^+$, there exists an $m \in \mathbb{N}$ such that for all measurable sets $A' \subset [\sqrt{\frac{\omega}{1-\beta} + \epsilon}, \infty)$ with positive Lebesgue measure we have $P(\sigma_m \in A' \mid Y_0 = y) > 0$. This implies $P(Y_m \in A \mid Y_0 = y) > 0$ since $Y_m = \sigma_m(\varepsilon_m, 1)'$ where ε_m is independent of σ_m and has positive Lebesgue density.

Let $\delta_m(x, s) = \omega \sum_{k=0}^{m-1} \beta^k + \beta^{m-1}(\beta s^2 + \alpha x^2)$. Then given $Y_0 = y \in \mathbb{R} \times \mathbb{R}^+$ we can find positive functions $h_i(\varepsilon_{i+1}, \ldots, \varepsilon_{m-1})$, $i = 1, \ldots, m-1$, such that (2.7) takes the form

$$\sigma_m^2 = \delta_m(y) + \sum_{i=1}^{m-1} \varepsilon_i^2 h_i(\varepsilon_{i+1}, \ldots, \varepsilon_{m-1}).$$

Let $A'' = \{r^2 \mid r \in A'\}$. Choosing m large enough, we have $\left|\delta_m(y) - \frac{\omega}{1-\beta}\right| < \epsilon$ and hence

$$P(\sigma_m \in A' \mid Y_0 = y) \;=\; P\left(\sigma_m^2 \in A'' \mid Y_0 = y\right)$$

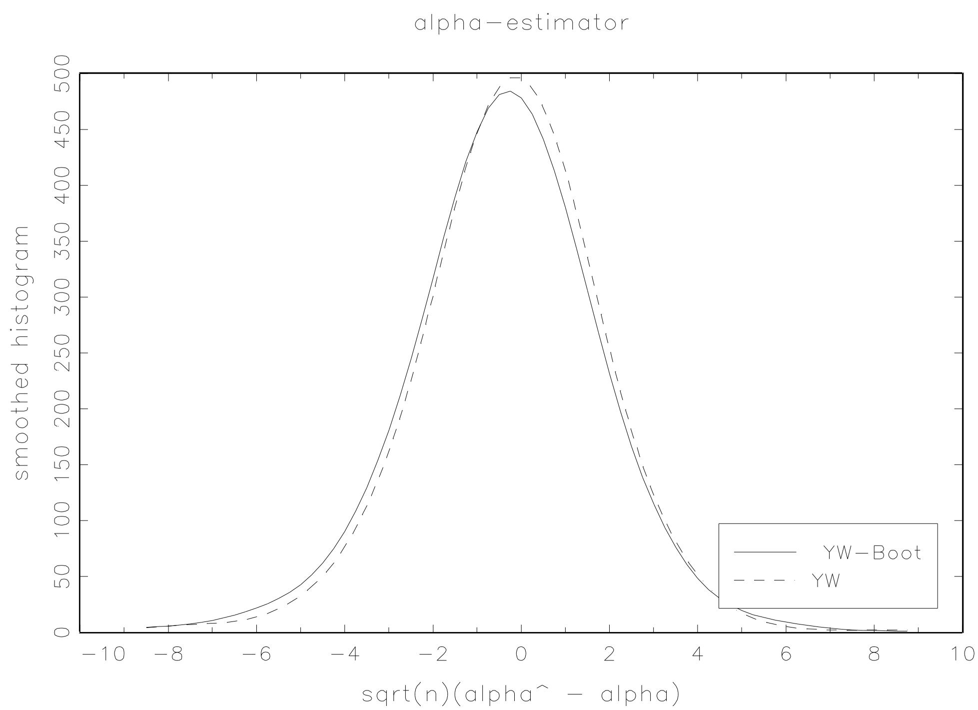

Fig. 1: Distribution of $\sqrt{n}(\hat{\alpha}_n - \alpha)$ with bootstrap approximation and for simulated GARCH(1,1) processes with $(\alpha, \beta, \omega) = (0.1, 0.4, 0.1)$, sample size $n = 1000$. All plots are based on 2500 Monte Carlo replications.

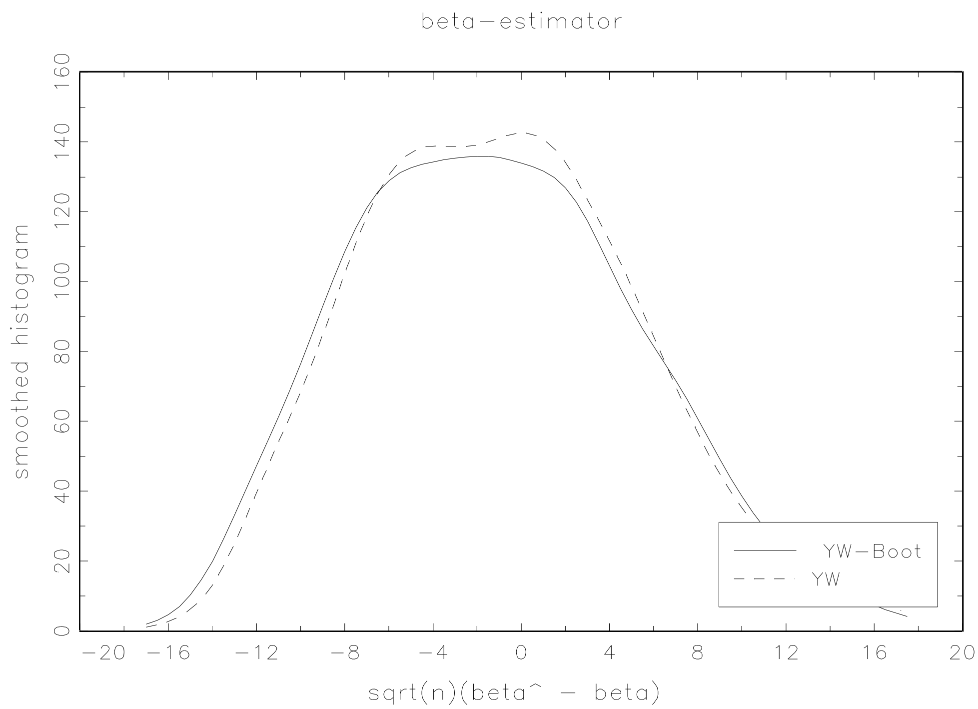

Fig. 2: Distribution of $\sqrt{n}(\hat{\beta}_n - \beta)$ with bootstrap approximation. Parameter as in Figure 1.

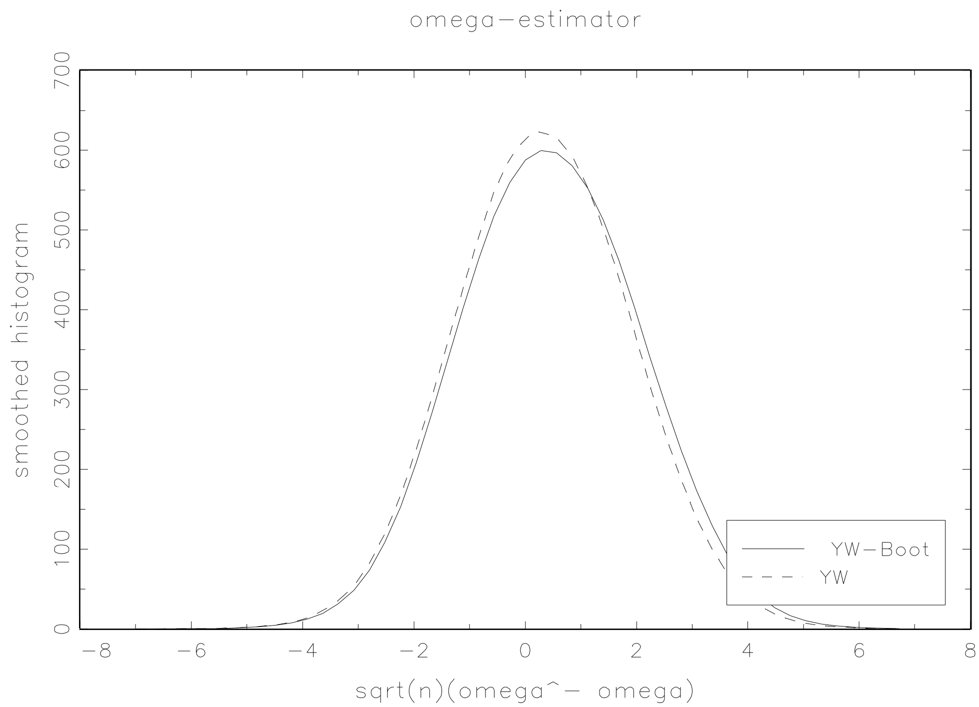

Fig. 3: Distribution of $\sqrt{n}(\hat{\omega}_n - \omega)$ with bootstrap approximation. Parameter as in Figure 1.

$$
\begin{aligned}
&= \int \dots \int 1\left\{\sum_{i=1}^{m-1} z_i^2 h_i(z_{i+1},\dots,z_{m-1}) \in A'' - \delta_m(y)\right\} \prod_{i=1}^{m-1} g(z_i)\, dz_1 \dots dz_{m-1} \\
&> 0
\end{aligned}
$$

as $g > 0$ and $A'' - \delta_m(y) \subset I\!R^+$. Thus $\{Y_t\}$ is ϕ-irreducible with ϕ being the Lebesgue measure restricted to $I\!R \times [\sqrt{\frac{\omega}{1-\beta}}, \infty)$.

Using Lemma $A.3$, the aperiodicity follows similarly. Choose the compact set $A = [0,1] \times [\sqrt{\frac{\omega}{1-\beta}}, \sqrt{\frac{\omega}{1-\beta}} + 1]$, say, and $m_1 \in I\!N$ such that

$$
\max_{y \in A} \left| \delta_m(y) - \frac{\omega}{1-\beta} \right| < \epsilon, \quad m \geq m_1.
$$

Then, with the same arguments as above, for all $B \subset A$ with $\lambda(B) > 0$ we obtain

$$
P\left(Y_m \in B \mid Y_0 = y\right) > 0 \quad \text{and} \quad P\left(Y_{m+1} \in B \mid Y_0 = y\right) > 0 \quad \text{for all } y \in B.
$$

In order to prove that compact sets are small, consider the 2-step transition probability and let $C = [-M, M] \times [0, M]$ for some $M > 0$. We obtain for any Borel set $A \subset I\!R \times I\!R^+$ and any $y \in C$

$$
\begin{aligned}
P\left(Y_2 \in A \mid Y_0 = y\right) &= \iint 1\left\{\sqrt{\omega + (\alpha z_1^2 + \beta)\sigma^2(y)}(z_2, 1)' \in A\right\} g(z_1) g(z_2)\, dz_1 dz_2 \\
&\geq \iint 1\left\{u(z_2, 1)' \in A\right\} g_C(u) g(z_2)\, du dz_2 \qquad (1.29)
\end{aligned}
$$

where we substituted $u = \sqrt{\omega + (\alpha z_1^2 + \beta)\sigma^2(y)}$, used $u \geq \sqrt{\omega}$, $\alpha\sigma^2(y) \leq M^2 + \omega$ and, if $z_1 = 0$, $u \leq M + 2\omega$, and put

$$
g_C(u) = 1_{[M+2\omega,\infty)}(u) \sqrt{\tfrac{\omega}{(M^2+\omega)u}} \inf_{y \in C} g\left(\sqrt{\tfrac{u^2 - \omega - \beta\sigma^2(y)}{\alpha\sigma^2(y)}}\right).
$$

As g is positive and continuous, g_C is positive on $[M + 2\omega, \infty)$, and

$$
\nu_C(\cdot) = \iint 1\{u(z,1)' \in \cdot\} g_C(u) g(z)\, du dz,
$$

defines a non-trivial measure on $I\!R \times I\!R^+$ with

$$
P\left(Y_2 \in A \mid Y_0 = y\right) \geq \nu_C(A) \quad \text{for all } y \in C, \text{ and all Borel sets } A \subset I\!R \times I\!R^+.
$$

Thus we have shown that C is small.

Now we are ready to apply the drift criterion given in Theorem A.4. Set $d = \frac{1}{2}\left(\frac{1}{\beta}+\frac{1}{1-\alpha}\right)-1$ and define the test function $V(y)=1+dx^2+s^2$ for $y=(x,s)'\in I\!R\times I\!R^+$. Then we have $d>0$ and, as $E\varepsilon_1^2=1$,

$$
\begin{aligned}
\Delta V(y) &= E\left[1+dX_1^2+\sigma_1^2 \mid Y_0=y\right]-(1+dx^2+s^2)\\
&= (1+d)(\omega+\alpha x^2+\beta s^2)-(dx^2+s^2)\\
&= (1-\alpha)x^2\left(\frac{\alpha}{1-\alpha}-d\right)+\beta s^2\left(1+d-\frac{1}{\beta}\right)+(1+d)\omega\\
&= -\left[(1-\alpha)x^2+\beta s^2\right]\frac{1}{2}\left(\frac{1}{\beta}-\frac{1}{1-\alpha}\right)+(1+d)\omega.
\end{aligned}
$$

Now $\alpha+\beta<1$ implies $\beta<1-\alpha$ and $\frac{1}{\beta}-\frac{1}{1-\alpha}>0$. Defining the constants $\delta = \frac{1}{4}\frac{\beta}{\max(1,d)}\left(\frac{1}{\beta}-\frac{1}{1-\alpha}\right) = \frac{1-(\alpha+\beta)}{4\max(1,d)(1-\alpha)}$ and $b=(1+d)\omega+2\delta$, and the compact set $C = [-\sqrt{\frac{b}{\delta}},\sqrt{\frac{b}{\delta}}]\times[0,\sqrt{\frac{b}{\delta}}]$ we thus have $\delta\in(0,\frac{1}{2})$ and

$$
\begin{aligned}
\Delta V(y) &\le -2\delta V(y)+(1+d)\omega+2\delta\\
&\le -\delta V(y)+b1_C(y) \quad \text{for all } y\in I\!R\times I\!R^+.
\end{aligned}
\tag{1.30}
$$

Thus the drift criterion holds. As $EV(Y_0)=1+(d+1)\mu<\infty$, this concludes the proof of the lemma. ■

Proof of Theorem 3.1: In order to avoid cumbersome notation, we will not distinguish between, say, $\sum_{t=1}^n x_t$ and $\sum_{t=h+1}^n x_t$. So, for instance, we will write $\hat\gamma_h = \frac{1}{n}\sum X_t^2X_{t-h}^2-\hat\mu^2$ neglecting terms of order $O_P(n^{-1})$.

In a first step we will prove joint asymptotic normality of

$$
\begin{aligned}
(\widehat{\alpha+\beta})_n &= \frac{\hat\gamma_2}{\hat\gamma_1},\\
(\widetilde{\beta^{-1}-\alpha})_n &= \frac{\hat\gamma_2-\hat\gamma_0}{\hat\gamma_1-(\alpha+\beta)\hat\gamma_0},\\
\tilde\omega_n &= \hat\mu\left(1-(\alpha+\beta)\right).
\end{aligned}
$$

By (3.12) we have

$$
\sqrt{n}\,(\tilde\omega_n-\omega) = \sqrt{n}\,(\hat\mu-\mu)\left(1-(\alpha+\beta)\right) = (1-\beta)n^{-\frac{1}{2}}\sum\eta_t. \tag{1.31}
$$

Furthermore

$$
\sqrt{n}\left((\widehat{\alpha+\beta})_n-(\alpha+\beta)\right) = \hat\gamma_1^{-1}\sqrt{n}\left(\hat\gamma_2-(\alpha+\beta)\hat\gamma_1\right)
$$

$$
\begin{aligned}
&= \hat{\gamma}_1^{-1} \frac{1}{\sqrt{n}} \sum X_{t-2}^2 \left[\left(X_t^2 - \hat{\mu} \right) - (\alpha+\beta) \left(X_{t-1}^2 - \hat{\mu} \right) \right] \\
&= \hat{\gamma}_1^{-1} \frac{1}{\sqrt{n}} \sum X_{t-2}^2 \left(\eta_t - \beta \eta_{t-1} + (1-(\alpha+\beta))(\mu - \hat{\mu}) \right) \\
&= \hat{\gamma}_1^{-1} \frac{1}{\sqrt{n}} \sum \eta_t \left(-\beta X_{t-1}^2 + X_{t-2}^2 - (1-\beta)\hat{\mu} \right)
\end{aligned}
$$

using (1.31). In a similar way we get

$$
\begin{aligned}
\hat{\gamma}_1 - (\alpha+\beta)\hat{\gamma}_0 &= \frac{1}{n} \sum \eta_t \left(-\beta X_t^2 + X_{t-1}^2 - (1-\beta)\hat{\mu} \right), \\
\hat{\gamma}_2 - \hat{\gamma}_0 &= \frac{1}{n} \sum \left[(\alpha+\beta) X_{t-1}^2 + \eta_t - \beta \eta_{t-1} \right] \left(X_{t-2}^2 - X_t^2 \right) \\
&= \frac{1}{n} \sum \eta_t \left(\beta X_{t+1}^2 - X_t^2 - \beta X_{t-1}^2 + X_{t-2}^2 \right)
\end{aligned}
$$

and therefore

$$
\begin{aligned}
&\hat{\gamma}_2 - \hat{\gamma}_0 - \left(\beta^{-1} - \alpha \right) \left(\hat{\gamma}_1 - (\alpha+\beta)\hat{\gamma}_0 \right) \\
&\quad = \frac{1}{n} \sum \eta_t \left(\beta X_{t+1}^2 - \alpha\beta X_t^2 - \left(\beta + \beta^{-1} - \alpha \right) X_{t-1}^2 + X_{t-2}^2 + \left(\beta^{-1} - \alpha \right) (1-\beta)\hat{\mu} \right).
\end{aligned}
$$

Observing that

$$
X_{t+1}^2 - \alpha X_t^2 = \omega + \beta X_t^2 + \eta_{t+1} - \beta \eta_t = \omega(1+\beta) + \beta(\alpha+\beta) X_{t-1}^2 - \beta^2 \eta_{t-1} + \eta_{t+1}
$$

we conclude that

$$
\begin{aligned}
\sqrt{n} \left((\widetilde{\beta^{-1} - \alpha})_n - \left(\beta^{-1} - \alpha \right) \right) &= \left(\hat{\gamma}_1 - (\alpha+\beta)\hat{\gamma}_0 \right)^{-1} \frac{1}{\sqrt{n}} \sum \eta_t \Big[\beta(1-\beta^2)\eta_{t-1} + \\
&+ \left(\beta^2(\alpha+\beta) + \alpha - \beta - \beta^{-1} \right) X_{t-1}^2 + X_{t-2}^2 + \left(\beta^{-1} - \alpha \right)(1-\beta)\hat{\mu} + \omega(1+\beta) \Big].
\end{aligned}
$$

By the ergodic theorem we have $\hat{\mu} \to \mu$ and $\hat{\gamma}_h \to \gamma_h$ almost surely. In order to prove

$$
\sqrt{n} \left((\widehat{\alpha+\beta})_n - (\alpha+\beta), (\widetilde{\beta^{-1} - \alpha})_n - \left(\beta^{-1} - \alpha \right), \tilde{\omega}_n - \omega \right) \xrightarrow{D} \mathcal{N}\left(0, \tilde{\Sigma} \right) \tag{1.32}
$$

it is therefore sufficient to show

$$
\frac{1}{\sqrt{n}} \sum \eta_t c' Z_t \xrightarrow{D} \mathcal{N}\left(0, c' \tilde{\Sigma} c \right), \quad c \in \mathbb{R}^3, \tag{1.33}
$$

using the Cramer-Wold device. As Z_t is $\mathcal{F}_{t-1}$-measurable and $E[\eta_t \mid \mathcal{F}_{t-1}] = 0$ we observe that $\{\eta_t Z_t(c)\}$ is a martingale difference sequence and hence

$$\sum_{t=-\infty}^{\infty} \operatorname{Cov}(\eta_t c' Z_t, \eta_0 c' Z_0) = c' E \eta_0^2 Z_0^2 c.$$

An application of the CLT for strongly mixing sequences, see Appendix A.2, now gives (1.33).

For the next step we note

$$\begin{aligned}
(\widehat{\beta^{-1} - \alpha})_n - (\widetilde{\beta^{-1} - \alpha})_n &= \frac{(\widehat{\beta^{-1} - \alpha})_n \hat{\gamma}_0}{\hat{\gamma}_1 - (\alpha+\beta)\hat{\gamma}_0} \left((\widehat{\alpha+\beta})_n - (\alpha+\beta) \right), \\
\hat{\omega}_n - \tilde{\omega}_n &= -\hat{\mu} \left((\widehat{\alpha+\beta})_n - (\alpha+\beta) \right)
\end{aligned}$$

Hence (1.32), the relationship $\gamma_1 - (\alpha+\beta)\gamma_0 = -\beta\sigma_\eta^2$ and the ergodic theorem imply

$$\sqrt{n} \left((\widehat{\alpha+\beta})_n - (\alpha+\beta), (\widehat{\beta^{-1} - \alpha})_n - \left(\beta^{-1} - \alpha \right), \hat{\omega}_n - \omega \right) \xrightarrow{D} \mathcal{N}\left(0, D_1 \tilde{\Sigma} D_1' \right).$$

An application of the delta method, using the function

$$T(x,y,z) = \begin{pmatrix} x - \frac{x+y}{2} + \sqrt{(x+y)^2/4 - 1} \\ \frac{x+y}{2} - \sqrt{(x+y)^2/4 - 1} \\ z \end{pmatrix} \tag{1.34}$$

with derivative $\nabla T(\alpha+\beta, \beta^{-1} - \alpha, \omega) = D_2$, concludes the proof of the theorem. ∎

Proof of Lemma 4.1: (a) ¿From (2.6) and (4.25) we have

$$\hat{\sigma}_t^2 - \sigma_t^2 = \sum_{k=0}^{t-1} \left[\left(\hat{\omega} \hat{\beta}^k - \omega \beta^k \right) + \left(\hat{\alpha} \hat{\beta}^k - \alpha \beta^k \right) X_{t-1-k}^2 \right] + \hat{\beta}^t \hat{\mu} - \beta^t \sigma_0^2.$$

Choose $b \in (\beta, 1)$ and set $B_n = \{\hat{\beta} \le b\}$. If $1 < r \le p$, application of the Hölder inequality with r and $q = (1 - \frac{1}{r})^{-1}$ gives

$$\left(\sum_{k=0}^{t-1} b^k X_{t-k-1}^2 \right)^r \le \left(\sum_{k=0}^{t-1} b^{\frac{kq}{2}} \right)^{\frac{r}{q}} \left(\sum_{k=0}^{t-1} b^{\frac{kr}{2}} X_{t-k-1}^{2r} \right)$$

and hence

$$E \frac{1}{n} \sum_{t=1}^{n} \left(\sum_{k=0}^{t-1} b^k X_{t-k-1}^2 \right)^r \le \left(1 - b^{\frac{q}{2}} \right)^{-\frac{r}{q}} \left(1 - b^{\frac{r}{2}} \right)^{-1} E X_0^{2r},$$

implying $\frac{1}{n}\sum_{t=1}^{n}\left(\sum_{k=0}^{t-1} b^k X_{t-k-1}^2\right)^r = O_P(1)$. Similarly, $\frac{1}{n}\sum_{t=1}^{n}\left(\sum_{k=0}^{t-1} k b^k X_{t-k-1}^2\right)^r = O_P(1)$. Part (a) follows now from Theorem 3.1, $P(B_n) \to 1$ and $\left|\hat{\beta}^k - \beta^k\right| \leq \left|\hat{\beta} - \beta\right| k b^{k-1}$ on B_n. The case $r = 1$ is even simpler. If $r > p$ write

$$\left(\sum_{k=0}^{t-1} b^k X_{t-k-1}^2\right)^r \leq \left(\max_{0\leq s\leq t-1} X_s^2\right)^{r-p} (1-b)^{-(r-p)} \left(\sum_{k=0}^{t-1} b^k X_{t-k-1}^2\right)^p$$

and use the fact that $n^{-\frac{1}{2}} \max_{0\leq s\leq n} X_s^2 = o_P(1)$ by the stationarity of $\{X_t^2\}$ and as $EX_t^4 < \infty$.

(b) It suffices to show $\frac{1}{n}\sum_{t=1}^{n} |\hat{\varepsilon}_t - \varepsilon_t| = O_P\left(n^{-\frac{1}{2}}\right)$. From

$$|\hat{\varepsilon}_t - \varepsilon_t| \leq |X_t| \left|\frac{1}{\hat{\sigma}_t} - \frac{1}{\sigma_t}\right| \leq |X_t| \frac{|\sigma_t - \hat{\sigma}_t|}{\sqrt{\omega\hat{\omega}}} \leq |X_t| \frac{|\sigma_t^2 - \hat{\sigma}_t^2|}{\sqrt{\omega\hat{\omega}}\left(\sqrt{\omega} + \sqrt{\hat{\omega}}\right)}$$

and $\frac{1}{n}\sum_{t=1}^{n} X_t^2 = \hat{\mu}$ we conclude

$$\left|\frac{1}{n}\sum_{t=1}^{n} (\hat{\varepsilon}_t - \varepsilon_t)\right| \leq \frac{\sqrt{\hat{\mu}}}{\sqrt{\omega\hat{\omega}}\left(\sqrt{\omega} + \sqrt{\hat{\omega}}\right)} \left[\frac{1}{n}\sum_{t=1}^{n} \left(\sigma_t^2 - \hat{\sigma}_t^2\right)^2\right]^{\frac{1}{2}}.$$

The assertion now follows from part (a) and the ergodic theorem. In the same way we have $\frac{1}{n}\sum_{t=1}^{n} |\hat{\varepsilon}_t^2 - \varepsilon_t^2| = O_P\left(n^{-\frac{1}{2}}\right)$ and hence $\left|\sigma_{\tilde{G}}^2 - 1\right| = O_P\left(n^{-\frac{1}{2}}\right)$.

(c) We first note that $g(x) \to 0$ for $|x| \to \infty$ as g is a uniformly continuous density function. Because of this and part (b) it is therefore sufficient to show $\|\tilde{g} - g\|_\infty = o_P(1)$ where $\tilde{g}(x) = \frac{1}{nh}\sum_{t=1}^{n} \varphi\left(\frac{x-\hat{\varepsilon}_t}{h}\right)$ is the density of $\tilde{G}$. As

$$\frac{1}{nh}\sum_{t=1}^{n} \left|\varphi\left(\frac{x-\hat{\varepsilon}_t}{h}\right) - \varphi\left(\frac{x-\varepsilon_t}{h}\right)\right| \leq \|\varphi'\|_\infty \frac{1}{h^2}\frac{1}{n}\sum_{t=1}^{n} |\hat{\varepsilon}_t - \varepsilon_t| = O_P\left(n^{-\frac{1}{10}}\right)$$

by the proof of part (b), $\tilde{g}$ may be replaced by $\breve{g}(x) = \frac{1}{nh}\sum_{t=1}^{n} \varphi\left(\frac{x-\varepsilon_t}{h}\right)$. The result now follows from $\sup_{|x|\leq n} |\breve{g}(x) - g(x)| \to 0$ almost surely (Bosq (1996), Theorem 2.2), $\sup_{x\geq n} |\breve{g}(x)| \leq \frac{1}{h}\varphi\left(\frac{n}{2h}\right)$ on $A_n = \left\{\sup_{1\leq t\leq n} |\varepsilon_t| > \frac{n}{2}\right\}$ and $P(A_n) \to 0$.

(d) In view of (b) and $h \to 0$ the first assertion will follow from $\frac{1}{n}\sum_{t=1}^{n} (|\hat{\varepsilon}_t|^q - |\varepsilon_t|^q) = o_P(1)$, $0 < q < 2p$. Set $\frac{1}{r} = 1 - \frac{q}{2p}$. Then, as $\min(\hat{\sigma}_t^2, \sigma_t^2) \geq \min(\hat{\omega}, \omega)$,

$$\left|\frac{1}{\hat{\sigma}_t^q} - \frac{1}{\sigma_t^q}\right| \leq (q+1) \min(\hat{\omega}, \omega)^{-\frac{q+1}{2}} |\hat{\sigma}_t - \sigma_t| \leq (q+1) \frac{\min(\hat{\omega}, \omega)^{-\frac{q+1}{2}}}{\hat{\omega}^{\frac{1}{2}} + \omega^{\frac{1}{2}}} \left|\hat{\sigma}_t^2 - \sigma_t^2\right|$$

and therefore

$$\begin{aligned}\frac{1}{n}\sum_{t=1}^{n}\left||\hat{\varepsilon}_t|^q-|\varepsilon_t|^q\right| &= \frac{1}{n}\sum_{t=1}^{n}|X_t|^q\left|\frac{1}{\hat{\sigma}_t^q}-\frac{1}{\sigma_t^q}\right| \\ &\le O_P(1)\frac{1}{n}\sum_{t=1}^{n}|X_t|^q\left|\hat{\sigma}_t^2-\sigma_t^2\right| \\ &\le O_P(1)\left(\frac{1}{n}\sum_{t=1}^{n}|X_t|^{2p}\right)^{\frac{q}{2p}}\left(\frac{1}{n}\sum_{t=1}^{n}\left|\hat{\sigma}_t^2-\sigma_t^2\right|^r\right)^{\frac{1}{r}} \\ &= o_P(1)\end{aligned}$$

by part (a) and the ergodic theorem.

With regard to the conditional variances, we first prove the boundedness of $E^*\sigma_0^{*q}$ along subsequences of arbitrary subsequences. Recall from Remark 3.2 that $E(\beta+\alpha\varepsilon_0^2)^{\frac{q}{2}}<1$ and choose $\kappa\in(E(\beta+\alpha\varepsilon_0^2)^{\frac{q}{2}},1)$. If $(k)\subset I\!N$ is any subsequence, choose $(k_\ell)\subset(k)$ in such a way that $E^*\left(\hat{\beta}+\hat{\alpha}\varepsilon_0^{*2}\right)^{\frac{q}{2}}$ converges to $E(\beta+\alpha\varepsilon_0^2)^{\frac{q}{2}}$ almost surely. Then, by Minkowski's inequality and (2.9),

$$\limsup_{\ell\to\infty}\left(E^*\sigma_0^{*q}\right)^{\frac{1}{q}}\le\limsup_{\ell\to\infty}\left(\sum_{k=0}^{\infty}\left(E^*\left(\hat{\beta}+\hat{\alpha}\varepsilon_0^{*2}\right)^{\frac{q}{2}}\right)^{\frac{2k}{q}}\right)^{\frac{1}{2}}\le\left(\frac{1}{1-\kappa^{\frac{2}{q}}}\right)^{\frac{1}{2}}$$

almost surely.

For the mixed moments of conditional variances we only discuss the case $k_1=k_2=k_3=k_4=2$, the other cases being similar. Consider

$$E^*\prod_{i=1}^{4}\sigma_{t_i}^{*2}=\hat{\omega}^4\sum_{k_1=0}^{\infty}\sum_{k_2=0}^{\infty}\sum_{k_3=0}^{\infty}\sum_{k_4=0}^{\infty}E^*\prod_{j_1=1}^{k_1}\prod_{j_2=1}^{k_2}\prod_{j_3=1}^{k_3}\prod_{j_4=1}^{k_4}\prod_{i=1}^{4}\left(\hat{\beta}+\hat{\alpha}\varepsilon_{t_i-j_i}^{*2}\right).$$

Again, we may assume $E^*\left(\hat{\beta}+\hat{\alpha}\varepsilon_0^{*2}\right)^4\le\kappa<1$. Hence, by independence,

$$E^*\prod_{j_1=1}^{k_1}\prod_{j_2=1}^{k_2}\prod_{j_3=1}^{k_3}\prod_{j_4=1}^{k_4}\prod_{i=1}^{4}\left(\hat{\beta}+\hat{\alpha}\varepsilon_{t_i-j_i}^{*2}\right)\le\kappa^{k_1+k_2+k_3+k_4}$$

as, for given $s\in\mathbb{Z}$, the factor $\left(\hat{\beta}+\hat{\alpha}\varepsilon_s^{*2}\right)$ appears at most four times in each product. The assertion now follows by dominated convergence.

The assertions concerning X_t^* then follow from $E^*|X_0^*|^q=E^*|\varepsilon_0^*|^q\,E^*\sigma_0^{*q}$. ∎

Proof of Theorem 4.2: The proof is an adaptation of the proof of Theorem 3.1 in Franke et al. (1998), referred to as (F) from now on, to our situation. As in (F) the proof consists of two steps. First, the identification of the parameters on which the mixing coefficients of the process $\{Y_t\}$ depend, and second, the convergence of these parameters for the bootstrap processes. The first step is formulated in the following lemma which corresponds to Theorem 2.1 and Corollary 2.1 in (F).

Lemma A.1 *Let Assumption S hold. Then there exist constants $C_Y > 0$ and $\rho_Y > 1$ such that*

$$\beta(n) \leq C_Y \rho_Y^{-j} \qquad \text{for all } j \in I\!N$$

where $\beta(j)$, $j \in I\!N$, are the β-mixing coefficients of the process $\{Y_t\}$. The constants C_Y and ρ_Y depend only on the parameters $K, \rho, \epsilon, n_0, \gamma, \kappa, d$ specified in conditions (A1) and (A2) below.

(A1) K is a compact set, $\rho > 1$, $d, \epsilon > 0$, $A < \infty$ such that

$$E\left[\|Y_t\| \mid Y_{t-1} = y\right] \leq 1_{K^c}(y)\left(\rho^{-1}\|y\| - \epsilon\right) + A1_K(y)$$

where $\|(x,s)\| = dx^2 + s^2$.
(A2) We have $n_0 \in I\!N$, $\gamma \in (0,1)$, and there is a probability measure ϕ such that

$$\inf_{y \in K} P^{n_0}(y, B) \geq \gamma\phi(B)$$

for all measurable sets B. Furthermore $\kappa > 0$ and

$$\inf_{y \in K} P(y, K) \geq \kappa. \tag{1.35}$$

Condition (A1), a reformulation of the drift condition, and condition (A2), whose first part concerns the 'smallness' of K, are as in (F), with the only exception that the absolute value $|X_t|$ of the one-dimensional process there has been replaced by the norm $\|Y_t\| = dX_t^2 + \sigma_t^2$. Both conditions are fulfilled in our setting as the following arguments will show.

Recall

$$\Delta V(y) \leq -\delta V(y) + b1_C(y), \quad y \in I\!R \times I\!R^+, \tag{1.36}$$

¿from drift equation (1.30) where

$$\begin{aligned} V(y) &= 1 + \|y\| = 1 + dx^2 + s^2, \quad y = (x,s)', \\ d &= \frac{1}{2}\left(\frac{1}{\beta} + \frac{1}{1-\alpha}\right) - 1, \\ \delta &= \frac{1-(\alpha+\beta)}{4\max(1,d)(1-\alpha)}, \\ C &= \left[-\sqrt{\frac{b}{\delta}}, \sqrt{\frac{b}{\delta}}\right] \times \left[0, \sqrt{\frac{b}{\delta}}\right], \\ b &= (1+d)\omega + 2\delta. \end{aligned}$$

As (1.36) is equivalent to

$$E\left[\|Y_t\| \mid Y_{t-1} = y\right] \leq (1-\delta)\|y\| - \delta + b1_C(y),$$

(A1) holds with $K = C$, $\rho = (1-\delta)^{-1}$, $\epsilon = \delta$ and $A = b$.

Furthermore, by (1.29), the first part of (A2) holds with $n_0 = 2$, $\gamma = \nu_K(I\!R \times I\!R^+)$ and $\phi = \gamma^{-1}\nu_K$ where

$$\begin{aligned} \nu_K(\cdot) &= \iint 1\{u(z,1)' \in \cdot\} g_K(u) g(z)\, du dz, \\ g_K(u) &= 1_{[M+2\omega,\infty)}(u)\sqrt{\tfrac{\omega}{(M^2+\omega)u}}\ \inf_{y\in C} g\left(\sqrt{\tfrac{u^2-\omega-\beta\sigma^2(y)}{\alpha\sigma^2(y)}}\right), \\ M &= \sqrt{\frac{b}{\delta}}. \end{aligned}$$

For the second part of (A2) consider $y = (x,s) \in K$, that is $0 \leq x^2, s^2 \leq \frac{b}{\delta}$. Note that $\sigma^2(y) \leq \frac{b}{\delta}$ because of $\sigma^2(y) = \omega + \alpha x^2 + \beta s^2 \leq \frac{b}{\delta}\left(\frac{\omega\delta}{b} + (\alpha+\beta)\right)$ and $\frac{\omega}{b} \leq \frac{1}{1+d} \leq 2(1-\alpha)$. Hence we have

$$\begin{aligned} P\left(Y_1 \in K \mid Y_0 = y\right) &= \int 1\left\{\sigma(y)(z,1)' \in K\right\} g(z)\, dz \\ &\geq \int 1\{|z| \leq 1\} g(z)\, dz \end{aligned}$$

and (1.35) holds with $\kappa = \int 1\{|z| \leq 1\} g(z)\, dz$.

Proof of Lemma A.1: The proof is completely analogous to the proof of Theorem 2.1 and Corollary 2.1 in (F), again the only difference being the replacement of $|X_t|$ by $\|Y_t\| = dX_t^2 + \sigma_t^2$. In particular, Lemma 2.1 in (F) concerning the return times $\tau_K = \inf\{t \geq 1 \mid Y_t \in K\}$ here takes the following form:

Lemma A.2 *Suppose (A1) is fulfilled. Then*
(i) $E_y\rho^{\tau_K} \leq \varepsilon^{-1}\|y\|$ *for all* $y \notin K$,
(ii) $E_y\rho^{\tau_K} \leq \rho(1+\varepsilon^{-1}A)$ *for all* $y \in K$.

With these changes, the arguments in (F) carry over to our situation without further modification. ■

Proof of Theorem 4.2 (continuation): The proof is completed by showing that, along suitable subsequences, the constants $K, \rho, \epsilon, n_0, \gamma, \kappa, d$ specified in conditions (A1) and (A2) may be chosen in such a way that, outside some null set, they are valid simultaneously for all bootstrap processes. By Theorem 3.1 and Lemma 4.1 we may choose a subseqence such that $\hat{\theta} \to \theta$, $\|\hat{g} - g\|_\infty \to 0$ and (4.26) is satisfied almost surely. By a slight abuse of notation we will denote this subsequence again by (n). Hence outside some null set the bootstrap processes $\{Y_t^*\}$ based on the observations $X_1, \ldots, X_n$ are aperiodic, ϕ-irreducible Markov processes. The argument is completed by showing that the bootstrap constants $K^*, \rho^*, \epsilon^*, n_0^*, \gamma^*, \kappa^*, d^*$ converge against $K, \rho, \epsilon, n_0, \gamma, \kappa, d$ when both are defined as in the discussion after the formulation of (A1) and (A2). This is obvious for $\rho^* = (1-\delta^*)^{-1}$, $\epsilon^* = \delta^*$ and d^*. Furthermore $K^* = [-M^*, M^*] \times [0, M^*]$ converges in an obvious way to K and we may set $n_0^* = n_0 = 2$. Finally, uniform convergence of $g^* = \hat{g}$ implies convergence of κ^* and pointwise convergence of g_K^* in $\mathbb{R} \setminus \{M + 2\omega\}$, and the latter implies convergence of the integrals γ^*. ■

Proof of Corollary 4.3: We only prove the assertion for $\hat{\gamma}_0^*$. The proof for $\hat{\mu}^*$ and $\hat{\gamma}_h^*$, $h > 0$, is similar, and the result extends to $\hat{\theta}^*$ by continuity. Defining $\frac{1}{r} = \frac{\delta/2}{8+\delta/2}$ and $\frac{1}{p} = \frac{1}{q} = \frac{4}{8+\delta/2}$ we obtain from the covariance inequality in Appendix A.2

$$\mathrm{Var}\hat{\gamma}_0^* \leq \frac{2}{n}\sum_{k=0}^{n-1}\left|\mathrm{Cov}\left(X_0^{*4}, X_k^{*4}\right)\right| \leq \frac{8}{n}\sum_{k=0}^{n-1}\beta_n^*(k)^{\frac{\delta/2}{8+\delta/2}}\left(E^* X_0^{*(8+\delta/2)}\right)^{\frac{8}{8+\delta/2}}.$$

As $E^* X_0^{*(8+\delta/2)} = O_P(1)$ and $E^*\hat{\gamma}_0^* - \gamma_0 = o_P(1)$ by Lemma 4.1, Theorem 4.2 implies $E^*\left|\hat{\gamma}_0^* - \gamma_0\right| = o_P(1)$ and hence $P^*\left(\left|\hat{\gamma}_0^* - \gamma_0\right| > \epsilon\right) = o_P(1)$ for all $\epsilon > 0$. ■

Proof of Theorem 4.4: By Corollary 4.3 the bootstrap estimators are consistent. It is therefore sufficient to prove

$$\frac{1}{\sqrt{n}}\sum \eta_t^* Z_t^* \xrightarrow{D} \mathcal{N}\left(0, \tilde{\Sigma}\right) \quad \text{in probability,} \tag{1.37}$$

cp. (1.33), where η_t^* and Z_t^* are defined as η_t and Z_t in the proof of Theorem 3.1, with $X_1, \ldots, X_n$ replaced by $X_1^*, \ldots, X_n^*$. As $\{\frac{1}{\sqrt{n}}\sum \eta_t^* Z_t^*\}$ is P^*-tight by the martingale property and Lemma 4.1, the corresponding characteristic functions are uniformly equicontinuous and (1.37) is equivalent to

$$\frac{1}{\sqrt{n}}\sum \eta_t^* c' Z_t^* \xrightarrow{D} \mathcal{N}\left(0, c'\tilde{\Sigma}c\right) \quad \text{in probability,} \quad \text{for all } c \in \mathbb{Q}.$$

We will apply Theorem A1 of Politis et al. (1997) which states a central limit theorem for a triangular array of mixing sequences. We have to verify the three conditions of this theorem. The moment condition is fulfilled by the moment assumptions on X_t. The condition for the mixing conditions is fulfilled by Theorem 4.2. The convergence of the bootstrap variance follows from Lemma 4.1. This concludes the proof of Theorem 4.4. ■

A.2 Geometric ergodicity

Here we give a short introduction into some Markov chain terminology and a criterion for geometric ergodicity. For details and proofs we refer to the book of Meyn and Tweedie (1993).

Let $\mathbf{X} = \{X_t\}_{t\geq 0}$ be a time-homogeneous Markov chain with state space $(\mathcal{X}, \mathcal{B}(\mathcal{X}))$, where $\mathcal{B}(\mathcal{X})$ is a countably generated σ-algebra on $\mathcal{X}$. In our examples in the previous sections we have $\mathcal{X} = I\!\!R \times I\!\!R^+$ and $\mathcal{B}(\mathcal{X})$ the Borel-σ-algebra on $I\!\!R \times I\!\!R^+$. Let $P = \{P(x, A) : x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X})\}$ denote the transition probability kernel and ν the initial distribution. Define the n-step transition probabilities inductively by $P^1(x, A) = P(x, A)$ and, for $n \geq 2$, by

$$P^n(x, A) = \int_{\mathcal{X}} P(x, dy) P^{n-1}(y, A), \quad x \in \mathcal{X}, A \in \mathcal{B}(\mathcal{X}).$$

Let P_ν be the corresponding probability measure such that, under P_ν, $\mathbf{X}$ is a Markov chain with transition probability P and initial distribution ν. Write P_x for P_ν if the process is started in x, i.e. if $\nu = \delta_x$ is the Dirac measure in x. In particular we have $P_x(X_n \in A) = P(X_n \in A \mid X_0 = x) = P^n(x, A)$ for $n \in I\!\!N$, $A \in \mathcal{B}(\mathcal{X})$.

Define for any set $A \in \mathcal{B}(\mathcal{X})$ the *occupation time* η_A, which counts the numbers of visits to A, by $\eta_A = \sum_{t=1}^{\infty} 1\{X_t \in A\}$, and the *first return time* on A, τ_A, by $\tau_A = \min\{t \geq 1 : X_t \in A\}$ so that $P_x(\tau_A < \infty\} = P_x(\mathbf{X} \text{ ever enters } A)$. We call $\mathbf{X}$ ϕ-*irreducible* if there exists a measure ϕ on $\mathcal{B}(\mathcal{X})$ such that, whenever $\phi(A) > 0$, we have $P_x(\tau_A < \infty) > 0$ for all $x \in \mathcal{X}$. An irreducibility measure ψ is called maximal if for any irreducibility measure ϕ we have that $\psi(A) = 0$ implies $\phi(A) = 0$, $A \in \mathcal{B}(\mathcal{X})$. If $\mathbf{X}$ is ϕ-irreducible then there exists a maximal irreducibility measure ψ, see Meyn and Tweedie (1993), Proposition 4.2.2. The Markov chain is called ψ-irreducible if it is ϕ-irreducible for some ϕ and the measure ψ is a maximal irreducibility measure.

The set $A \in \mathcal{B}(\mathcal{X})$ is called *recurrent* if $E_x[\eta_A] = \infty$ for all $x \in A$. The chain $\mathbf{X}$ is called *recurrent* if it is ψ-irreducible and every set $A \in \mathcal{B}(\mathcal{X})$ with $\psi(A) > 0$ is recurrent. The set A is called *Harris recurrent* if $P_x\{\mathbf{X} \in A \text{ infinitely often}\} = P_x(\eta_A = \infty) = 1$, $x \in A$. This is equivalent to $P_x(\tau_A < \infty) = 1$, $x \in A$. The chain $\mathbf{X}$ is called *Harris recurrent* if it is ψ-irreducible and every set $A \in \mathcal{B}(\mathcal{X})$ with $\psi(A) > 0$ is Harris recurrent. Note that if a set is Harris recurrent, then it is recurrent.

If the chain $\mathbf{X}$ is recurrent then there exists an invariant measure π, i.e. a σ-finite measure π on $\mathcal{B}(\mathcal{X})$ with the property

$$\pi(A) = \int_{\mathcal{X}} \pi(dx) P(x, A), \quad A \in \mathcal{B}(\mathcal{X}).$$

This invariant measure is not necessarily finite. If it is finite, then it can be normalized to an invariant probability measure. Suppose that $\mathbf{X}$ is ψ-irreducible, and admits an invariant probability measure π. Then $\mathbf{X}$ is called a *positive* chain. If $\mathbf{X}$ is Harris recurrent and positive, then $\mathbf{X}$ is called a *positive Harris (recurrent)* chain. Observe that invariant probability measures define stationary processes if we choose them as initial distribution for X_0. Moreover, if a limiting distribution of $P_\nu(X_n \in \cdot)$ exists, it is an invariant probability measure. Hence invariant probability measures also define the long term behavior of the chain. Before we can state the existence of such limits, we need to introduce *small sets*.

A set $C \in \mathcal{B}(\mathcal{X})$ is called a *small set* if there exists an $m > 0$, and a non-trivial measure ν_m on $\mathcal{B}(\mathcal{X})$, such that for all $x \in C, B \in \mathcal{B}(\mathcal{X})$

$$P^m(x, B) \geq \nu_m(B).$$

Moreover, we define the one-step "mean drift": The drift operator Δ is defined for any non-negative measurable function V by

$$\begin{aligned} \Delta V(x) &:= \int_{\mathcal{X}} P(x, dy) V(y) - V(x) \\ &= E_x[V(X_1) - V(X_0)], \quad x \in \mathcal{X}. \end{aligned}$$

Henceforth we concentrate on aperiodic ψ-irreducible Markov chains. For a formal definition of periodic respectively aperiodic chains see Meyn and Tweedie (1993). A useful criterion to check aperiodicity is given in the following Lemma, cf. Tong (1990).

Lemma A.3 *If* $\mathbf{X}$ *is* ψ*-irreducible, a necessary and sufficient condition for* $\mathbf{X}$ *to be aperiodic is that there exists an* $A \in \mathcal{B}(\mathcal{X})$ *with* $\psi(A) > 0$ *and the property: For all* $B \subseteq A$ *with* $B \in \mathcal{B}(\mathcal{X})$ *and* $\psi(B) > 0$ *there exists a positive integer* n *such that*

$$P^n(x, B) > 0 \quad \text{and} \quad P^{n+1}(x, B) > 0, \quad x \in B.$$

PROOF: Tong (1990), Proposition A1.2. ■

If $\mathbf{X}$ is a positive Harris chain, then $\mathbf{X}$ is called *geometrically ergodic* if there is some $\rho > 1$ and some function $W : \mathcal{X} \to (0, \infty)$ such that $E_\pi W(X_0) < \infty$ and

$$\|P^n(x, \cdot) - \pi\| \leq W(x) \rho^{-n}, \quad x \in \mathcal{X},$$

where $\|\cdot\|$ denotes the total variation norm.

Now we are ready to state the following drift criterion for geometric ergodicity, cf. Meyn and Tweedie (1993), Theorem 15.0.1.

Theorem A.4 *Suppose that the chain* $\mathbf{X}$ *is* ψ*-irreducible and aperiodic. If there exists a small set* C*, constants* $b < \infty, \beta > 0$*, and a measurable function* $V : \mathcal{X} \to [1, \infty)$ *satisfying*

$$\Delta V(x) \leq -\beta V(x) + b 1_C(x), \quad x \in \mathcal{X}, \tag{1.38}$$

then the chain $\mathbf{X}$ *is positive recurrent with invariant probability measure* π *and there exist constants* $\rho > 1$*,* $M < \infty$ *such that*

$$\|P^n(x, \cdot) - \pi\| \leq M V(x) \rho^{-n}, \quad x \in \mathcal{X}. \tag{1.39}$$

PROOF: Meyn and Tweedie (1993). ■

Finally we state the ergodic theorem. It shows that the strong law of large numbers also holds for positive Harris recurrent Markov chains, cf. Meyn and Tweedie (1993), Theorem 17.1.7.

Theorem A.5 (ergodic theorem) *The following are equivalent when an invariant probability measure* π *exists for* $\mathbf{X}$*:*

(i) $\mathbf{X}$ *is positive Harris recurrent.*

(ii) For each $f \in L_1(\mathcal{X}, \mathcal{B}(\mathcal{X}), \pi)$*,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t=1}^{n} f(X_t) = \int f d\pi \qquad \textit{almost surely}$$

for any initial distribution.

PROOF: Meyn and Tweedie (1993). ■

A.3 Mixing

We give definitions of α- and β-mixing and state some results for mixing sequences which are used in the previous sections. For details we refer to Doukhan (1994) and the references therein.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $\mathcal{B}, \mathcal{C}$ two σ-subfields of $\mathcal{A}$. Then the α*-mixing* coefficient of $\mathcal{B}$ and $\mathcal{C}$ is defined as

$$\alpha(\mathcal{B}, \mathcal{C}) = \sup_{B\in\mathcal{B},\, C\in\mathcal{C}} |P(B\cap C) - P(B)P(C)|$$

and the β-*mixing* coefficient of $\mathcal{B}$ and $\mathcal{C}$ as

$$\beta(\mathcal{B},\mathcal{C}) = E\left[\sup_{C\in\mathcal{C}} |P(C \mid \mathcal{B}) - P(C)|\right].$$

The mixing coefficients may be used to obtain covariance inequalities. If X and Y are measurable random variables with respect to $\mathcal{B}$ and $\mathcal{C}$, respectively, then

$$|\mathrm{Cov}(X,Y)| \le 8\alpha^{\frac{1}{r}}(\mathcal{B},\mathcal{C})\,(E|X|^p)^{\frac{1}{p}}\,(E|Y|^q)^{\frac{1}{q}} \quad \text{for any } p,q,r \ge 1 \text{ and } \frac{1}{r}+\frac{1}{p}+\frac{1}{q}=1,$$

see Doukhan (1994), p. 9.

Let $\mathbf{X} = \{X_t\}_{t\in\mathbb{Z}}$ be a sequence of random variables taking values in $\mathbb{R}^d$ and $\mathcal{F}_n^m$ be the σ-algebra generated by $\{X_t : n \le t \le m\}$, $-\infty \le n \le m \le \infty$. Then the process $\mathbf{X}$ is called α-*mixing* (or *strongly mixing*) if

$$\alpha_k := \sup_{t\in\mathbb{Z}} \alpha\left(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+k}^\infty\right) \stackrel{k\to\infty}{\longrightarrow} 0$$

and β-*mixing* (or *absolutely regular*) if

$$\beta_k := \sup_{t\in\mathbb{Z}} \beta\left(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+k}^\infty\right) \stackrel{k\to\infty}{\longrightarrow} 0.$$

As α-*mixing* and β-*mixing* coefficients are related by $2\alpha(\mathcal{B},\mathcal{C}) \le \beta(\mathcal{B},\mathcal{C})$, see Doukhan (1994), p.4, absolute regularity implies strong mixing.

If $\mathbf{X}$ is a time-homogeneous Markov process with marginal distributions ν_t, then

$$\beta_k = \sup_{t\in\mathbb{Z}} \int \nu_t(dx) \left\| P^k(x,\cdot) - \nu_{t+k} \right\|,$$

see Doukhan (1994), p. 88. In particular, if $\mathbf{X}$ is stationary and geometrically ergodic, then there exist constants $c > 0$ and $\rho > 1$ such that

$$\beta_k \le c\rho^{-k}, \quad k \in \mathbb{N}.$$

Finally we state the central limit theorem for strongly mixing sequences (Ibragimov (1962)).

Theorem A.6 (CLT for strongly mixing sequences) *Let $\{X_t\}$ be a real valued, centered, stationary, strongly mixing sequence, and put $S_n = \sum_{t=1}^n X_t$, $\sigma_n = ES_n^2$. If, for some $\delta > 0$,*

$$E\,|X_1|^{2+\delta} < \infty \quad \textit{and} \quad \sum_{k=1}^\infty \alpha_k^{\delta/(2+\delta)} < \infty,$$

then

$$\frac{\sigma_n^2}{n} \longrightarrow \sigma^2 = E|X_1|^2 + 2\sum_{k=2}^{\infty} EX_1X_k.$$

If, in addition, $\sigma^2 > 0$, *then*

$$\frac{S_n}{\sqrt{n}\sigma} \xrightarrow{D} \mathcal{N}(0,1).$$

Acknowledgements

We would like to thank Jens-Peter Kreiss for many helpful comments. The second author acknowledges the support of the DFG-project MA 1026/6-1.

References

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

Bollerslev, T. (1988) On the correlation structure for the generalized autoregressive conditional heteroskedastic process. *Journal of Time Series Analysis* 9, 121–131.

Bollerslev, T., R. Y. Chou & K. F. Kroner (1992) ARCH modelling in finance: a review of the theory and empirical evidence. *Journal of Econometrics* 52, 5–59.

Borkovec, M. & C. Klüppelberg (1998) The tail of the stationary distribution of an autoregressive process with ARCH(1) errors. Submitted for publication.

Bosq, D. (1996) *Nonparametric statistics for stochastic processes. Lecture Notes in Statistics* 110. New York: Springer.

Doukhan, P. (1994) *Mixing: Properties and Examples. Lecture Notes in Statistics* 85. New York: Springer.

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

Engle, R. F. (Ed.) (1995) *ARCH, Selected Readings*. Oxford: Oxford University Press.

Franke, J., J.-P. Kreiss & E. Mammen (1997) Bootstrap of kernel smoothing in nonlinear time series. Submitted for publication.

Franke, J., J.-P. Kreiss, E. Mammen & M. H. Neumann (1998) Properties of the nonparametric autoregressive bootstrap. Submitted for publication.

Guegan, D. & J. Diebolt (1992) Probabilistic properties of the β-ARCH model. Prepublication 92-4, University Paris-Nord.

Ibragimov, I. A. (1962) Some limit theorems for stationary processes. *Theory of Probability and its Applications* 7, 349–382.

Maercker, G. (1997) *Statistical Inference in Conditional Heteroskedastic Autoregressive Models*. Aachen: Shaker. (Dissertation, Technical University of Braunschweig)

Maercker, G. (1998) Bootstrapping GARCH(1,1) models. In: Refenes, A.-P. et al. (Ed.) *Decision Technologies for Computational Finance, Proceedings of the fifth international conference Computational Finance*. Dordrecht: Kluwer, 207–218.

Meyn, S. P. & R. L. Tweedie (1993) *Markov Chains and Stochastic Stability*. London: Springer.

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

Politis, D. N., J. P. Romano & M. Wolf (1997) Subsampling for heteroskedastic time series. *Journal of Econometrics* 81, 281-317.

Tong, H. (1990) *Nonlinear time series: A dynamical approach*. New York: Oxford University Press.

Gisela Maercker
Commerzbank AG
Neue Mainzer Straße 32-36
D–60261 Frankfurt am Main
Germany
Telephone +49 69 136-42368
Email gisela_maercker@commerzbank.com

Martin Moser
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
D–69120 Heidelberg
Germany
Telephone +49 6221 548984
Email moser@statlab.uni-heidelberg.de