Mixing properties of ARCH and time-varying ARCH processes

Piotr Fryzlewicz Suhasini Subba Rao

Abstract

There exists very few results on mixing for nonstationary processes. However, mixing is often required in statistical inference for nonstationary processes, such as time-varying ARCH (tvARCH) models. In this paper, bounds for the mixing rates of a stochastic process are derived in terms the conditional densities of the process. These bounds are used to obtain the α, 2-mixing and β-mixing rates of the nonstationary time-varying ARCH(p) process and ARCH(∞) process. It is shown that the mixing rate of time-varying ARCH(p) process is geometric, whereas the bounds on the mixing rate of the ARCH(∞) process depends on the rate of decay of the ARCH(∞) parameters. We mention that the methodology given in this paper is applicable to other processes.

Key words: Absolutely regular (β-mixing) ARCH(∞), conditional densities, time-varying ARCH, strong mixing (α-mixing), 2-mixing.

2000 Mathematics Subject Classification: 62M10, 60G99, 60K99.

1 Introduction

Mixing is a measure of dependence between elements of a random sequence that has a wide range of theoretical applications (see Bradley (2007) and below). One of the most popular mixing measures is α-mixing (also called strong mixing), where the α-mixing rate of the nonstationary stochastic process $\{X_t\}$ is defined as a sequence of coefficients $\alpha(k)$ such that

$$\alpha(k) = \sup_{t\in\mathbb{Z}} \sup_{\substack{H\in\sigma(X_t,X_{t-1},\ldots)\\ G\in\sigma(X_{t+k},X_{t+k+1},\ldots)}} |P(G\cap H) - P(G)P(H)|. \tag{1}$$

$\{X_t\}$ is called α-mixing if $\alpha(k) \to 0$ as $k \to \infty$. α-mixing has several applications in statistical inference. For example, if $\{\alpha(k)\}$ decays sufficiently fast to zero as $k \to \infty$, then, amongst other results, it is possible to show asymptotic normality of sums of $\{X_k\}$ (c.f. Davidson (1994), Chapter 24), as well as exponential inequalities for such sums (c.f. Bosq (1998)), asympototic normality of kernel-based nonparametric estimators (c.f. Bosq (1998)) and consistency of change point detection schemes of nonlinear time series (c.f. Fryzlewicz and Subba Rao (2008)). The notion of 2-mixing is related to strong mixing, but is a weaker condition as it measures the dependence between two random variables and not the entire tails. 2-mixing is often used in statistical inference, for example deriving rates in nonparametric regression (see Bosq (1998)). The 2-mixing rate can be used to derive bounds for the covariance between functions of random variables, say $\mathrm{cov}(g(X_t), g(X_{t+k}))$ (see Ibragimov (1962)), which is usually not possible when only the correlation structure of $\{X_k\}$ is known. The 2-mixing rate of $\{X_k\}$ is defined as a sequence $\tilde{\alpha}(k)$ which satisfies

$$\tilde{\alpha}(k) = \sup_{t\in\mathbb{Z}} \sup_{\substack{H\in\sigma(X_t) \\ G\in\sigma(X_{t+k})}} |P(G \cap H) - P(G)P(H)|. \tag{2}$$

It is clear that $\tilde{\alpha}(k) \leq \alpha(k)$. A closely related mixing measure, introduced in Volkonskii and Rozanov (1959) is β-mixing (also called absolutely regular). The β-mixing rate of the stochastic process $\{X_t\}$ is defined as a sequence of coefficients $\beta(k)$ such that

$$\beta(k) = \sup_{t\in\mathbb{Z}} \sup_{\substack{\{H_j\}\in\sigma(X_t,X_{t-1},\ldots) \\ \{G_j\}\in\sigma(X_{t+k},X_{t+k+1},\ldots)}} \sum_i \sum_j |P(G_i \cap H_j) - P(G_i)P(H_j)|, \tag{3}$$

where $\{G_i\}$ and $\{H_j\}$ are finite partitions of the sample space Ω. $\{X_t\}$ is called β-mixing if $\beta(k) \to 0$ as $k \to \infty$. It can be seen that this measure is slightly stronger than α-mixing (since an upper bound for $\beta(k)$ immediately gives a bound for $\alpha(k)$; $\beta(k) \geq \alpha(k)$).

Despite the versatility of mixing, its main drawback is that in general it is difficult to derive bounds for $\alpha(k)$, $\tilde{\alpha}(k)$ and $\beta(k)$. However the mixing bounds of some processes are known. Chanda (1974), Gorodetskii (1977), Athreya and Pantula (1986) and Pham and Tran (1985) show strong mixing of the MA(∞) process. Feigin and Tweedie (1985) and Pham (1986) have shown geometric ergodicity of Bilinear processes (we note that stationary geometrically ergodic Markov chains are geometrically α-mixing, 2-mixing and β-mixing - see, for example, Francq and Zakoïan (2006)). More recently, Tjostheim (1990) and Mokkadem (1990) have shown geometric ergodicity for a general class of Markovian processes. The results in Mokkadem (1990) have been applied in Bousamma (1998) to show geometric ergodicity of stationary ARCH(p) and GARCH(p, q) processes, where p and q are finite integers. Related results on mixing for GARCH(p, q) processes can be found in Carrasco and Chen (2002), Liebscher (2005), Sorokin (2006) and Lindner (2008) (for an excellent review) and Francq and Zakoïan (2006) and Meitz and Saikkonen (2008) (where mixing of 'nonlinear' GARCH(p, q) processes are also considered). Most of these these results are proven by verifying the Meyn-Tweedie conditions (see Feigin and Tweedie (1985) and Meyn and Tweedie (1993)), and, as mentioned above, are derived under the premise that the process is stationary (or asymptotically stationary) and Markovian. Clearly, if a process is nonstationary, then the aforementioned results do not hold. Therefore for nonstationary processes, an alternative method to prove mixing is required.

The main aim of this paper is to derive a bound for (1), (2) and (3) in terms of the densities of the process plus an additional term, which is an extremal probability. These bounds can be applied to various processes. In this paper, we will focus on ARCH-type processes and use the bounds to derive mixing rates for time-varying ARCH(p) (tvARCH) and ARCH(∞) processes. The ARCH family of processes is widely used in finance to model the evolution of returns on financial instruments: we refer the reader to the review article of Giraitis et al. (2005) for a comprehensive overview of mathematical properties of ARCH processes, and a list of further references. It is worth mentioning that Hörmann (2008) and Berkes et al. (2008) have considered a different type of dependence, namely a version of the m-dependence moment measure, for ARCH-type processes. The stationary GARCH(p, q) model tends to be the benchmark financial model. However, in certain situations it may not be the most appropriate model, for example it cannot adequently explain the long memory seen in the data or change according to shifts in the world economy. Therefore, recently attention has been paid to tvARCH models (see, for example, Mikosch and Stărică (2003), Dahlhaus and Subba Rao (2006), Fryzlewicz et al. (2008) and Fryzlewicz and Subba Rao (2008)) and ARCH(∞) models (see Robinson (1991), Giraitis et al. (2000), Giraitis and Robinson (2001) and Subba Rao (2006)). The derivations of the sampling properties of some of the above mentioned papers rely on quite sophisticated assumptions on the dependence structure, in particular their mixing properties.

We will show that due to the p-Markovian nature of the time-varying ARCH(p) process, the α-mixing, 2-mixing and β-mixing bound has the same geometric rate. The story is different for ARCH(∞) processes, where the mixing rates can be different and vary according to the rate of decay of the parameters. An advantage of the approach advocated in this paper is that these methods can readily be used to establish mixing rates of several time series models. This is especially useful in time series analysis, for example, change point detection schemes for nonlinear time series, where strong mixing of the underlying process is often required. The price we pay for the flexibility of our approach is that the assumptions under which we work are slightly stronger than the standard assumptions required to prove geometric mixing of the stationary GARCH process. However, the conditions do not rely on proving irreducibility (which is usually required when showing geometric ergodicity) of the underlying process, which can be difficult to verify.

In Section 2 we derive a bound for the mixing rate of general stochastic processes, in terms of the differences of conditional densities. In Section 3 we derive mixing bounds for time-varying ARCH(p) processes (where p is finite). In Section 4 we derive mixing bounds for ARCH(∞) processes. Proofs which are not in the main body of the paper can be found in the appendix and the accompanying technical report.

2 Some mixing inequalities for general processes

2.1 Notation

For $k > 0$, let $\underline{X}_t^{t-k} = (X_t, \ldots, X_{t-k})$; if $k \leq 0$, then $\underline{X}_t^{t-k} = 0$. Let $\underline{y}_s = (y_s, \ldots, y_0)$. Let $\|\cdot\|$ denote the ℓ_1-norm. Let Ω denote the sample space. The sigma-algebra generated by $X_t, \ldots, X_{t+r}$ is denoted as $\mathcal{F}_{t+r}^t = \sigma(X_t, \ldots, X_{t+r})$.

2.2 Some mixing inequalities

Let us suppose $\{X_t\}$ is an arbitrary stochastic process. In this section we derive some bounds for $\alpha(k)$, $\tilde{\alpha}(k)$ and $\beta(k)$. To do this we will consider bounds for

$$\sup_{H \in \mathcal{F}_t^{t-r_1}, G \in \mathcal{F}_{t+k+r_2}^{t+k}} |P(G \cap H) - P(G)P(H)| \quad \text{and} \quad \sup_{\{H_j\} \in \mathcal{F}_t^{t-r_1}, \{G_i\} \in \mathcal{F}_{t+k+r_2}^{t+k}} \sum_{i,j} |P(G_i \cap H_j) - P(G_i)P(H_j)|,$$

where $r_1, r_2 \geq 0$ and $\{G_i\}$ and $\{H_i\}$ are partitions of Ω. In the proposition below, we give a bound for the mixing rate in terms of conditional densities. Similar bounds for linear processes have been derived in Chanda (1974) and Gorodetskii (1977) (see also Davidson (1994), Chapter

14). However, the bounds in Proposition 2.1 apply to any stochastic process, and it is this generality that allows us to use the result in later sections, where we derive mixing rates for ARCH-type processes.

Proposition 2.1 *Let us suppose that the conditional density of* $\underline{X}_{t+k+r_2}^{t+k}$ *given* $\underline{X}_t^{t-r_1}$ *exists and denote it as* $f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}$*. For* $\underline{\eta} = (\eta_0, \ldots, \eta_{r_1}) \in (\mathbb{R}^+)^{r_1+1}$*, define the set*

$$E = \{\omega;\ \underline{X}_t^{t-r_1}(\omega) \in \mathcal{E}\} \text{ where } \mathcal{E} = \{(\nu_0, \ldots, \nu_{r_1});\ \text{for all } |\nu_j| \leq \eta_j\}. \tag{4}$$

Then for all $r_1, r_2 \geq 0$ *and* $\underline{\eta}$ *we have*

$$\begin{aligned} &\sup_{H \in \mathcal{F}_t^{t-r_1}, G \in \mathcal{F}_{t+k+r_2}^{t+k}} |P(G \cap H) - P(G)P(H)| \\ \leq\ & 2 \sup_{\underline{x} \in \mathcal{E}} \int_{\mathbb{R}^{r_2+1}} \left| f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) - f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0) \right| d\underline{y} + 4P(E^c), \end{aligned} \tag{5}$$

and

$$\begin{aligned} &\sup_{\{H_j\} \in \mathcal{F}_t^{t-r_1}, \{G_j\} \in \mathcal{F}_{t+k+r_2}^{t+k}} \sum_{i,j} |P(G_i \cap H_j) - P(G_i)P(H_j)| \\ \leq\ & 2 \int_{\mathbb{R}^{r_2+1}} \sup_{\underline{x} \in \mathcal{E}} \left| f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) - f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0) \right| d\underline{y} + 4P(E^c), \end{aligned} \tag{6}$$

where $\{G_i\}$ *and* $\{H_j\}$ *are finite partitions of* Ω. $\underline{X}_t^{t-r_1}$*. Let* $\underline{W}_{t+k-1}^{t+1}$ *be a random vector that is independent of* $\underline{X}_t^{t-r_1}$ *and* $f_{\underline{W}_{t+k-1}^{t+1}}$ *denote the density of* $\underline{W}_{t+k-1}^{t+1}$*, then we have*

$$\begin{aligned} &\sup_{H \in \mathcal{F}_t^{t-r_1}, G \in \mathcal{F}_{t+k+r_2}^{t+k}} |P(G \cap H) - P(G)P(H)| \\ \leq\ & 2 \sum_{s=0}^{r_2} \sup_{\underline{x} \in \mathcal{E}} \int f_{\underline{W}}(\underline{w}) \left\{ \sup_{\underline{y}_{s-1} \in \mathbb{R}^s} \int_{\mathbb{R}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) d\, y_s \right\} d\underline{w} + 4P(E^c) \end{aligned} \tag{7}$$

and

$$\begin{aligned} &\sup_{\{H_j\} \in \mathcal{F}_t^{t-r_1}, \{G_j\} \in \mathcal{F}_{t+k+r_2}^{t+k}} \sum_{i,j} |P(G_i \cap H_j) - P(G_i)P(H_j)| \\ \leq\ & 2 \sum_{s=0}^{r_2} \int f_{\underline{W}}(\underline{w}) \left\{ \sup_{\underline{y}_{s-1} \in \mathbb{R}^s} \int_{\mathbb{R}} \sup_{\underline{x} \in \mathcal{E}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) d\, y_s \right\} d\underline{w} + 4P(E^c) \end{aligned} \tag{8}$$

where $\mathcal{D}_{0,k,t}(y_0|\underline{y}_{-1}, \underline{w}, \underline{x}) = \left|f_{s,k,t}(y_s|\underline{w}, \underline{x}) - f_{s,k,t}(y_s|\underline{w}, 0)\right|$ *and for* $s \geq 1$

$$\mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) = \left|f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) - f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, 0)\right|, \tag{9}$$

with the conditional density of X_{t+k} *given* $(\underline{W}_{t+k-1}^{t+1}, \underline{X}_t^{t-r_1})$ *denoted as* $f_{0,k,t}$ *and the conditional density of* X_{t+k+s} *given* $(\underline{X}_{t+k+s-1}^{t+k}, \underline{W}_{t+k-1}^{t+1}, \underline{X}_t^{t-r_1})$*, denoted as* $f_{s,k,t}$*,* $\underline{x} = (x_0, \ldots, x_{-r_2})$ *and* $\underline{w} = (w_k, \ldots, w_1)$*.*

PROOF. In Appendix A.1. □

Since the above bounds hold for all vectors $\underline{\eta} \in (\mathbb{R}^+)^{r_1+1}$ (note $\underline{\eta}$ defines the set E; see (4)), by choosing the $\underline{\eta}$ which balances the integral and $P(E^c)$, we obtain an upper bound for the mixing rate.

The main application of the inequality in (7) is to processes which are 'driven' by the innovations (for example, linear and ARCH-type processes). If $\underline{W}_{t+k-1}^{t+1}$ is the innovation process, often it can be shown that the conditional density of X_{t+k+s} given $(\underline{X}_{t+k+s-1}^{t+k}, \underline{W}_{t+k-1}^{t+1}, \underline{X}_t^{t-r_1})$ can be written as a function of the innovation density. Deriving the density of X_{t+k+s} given $(\underline{X}_{t+k+s-1}^{t+k}, \underline{W}_{t+k-1}^{t+1}, \underline{X}_t^{t-r_1})$ is not a trivial task, but it is often possible. In the subsequent sections we will apply Proposition 2.1 to obtaining bounds for the mixing rates.

The proof of Proposition 2.1 can be found in the appendix, but we give a flavour of it here. Let

$$H = \{\omega;\, \underline{X}_t^{t-r_1}(\omega) \in \mathcal{H}\}, \quad G = \{\omega;\, \underline{X}_{t+k+r_2}^{t+k}(\omega) \in \mathcal{G}\}. \tag{10}$$

It is straightforward to show that $|P(G\cap H) - P(G)P(H)| \leq |P(G\cap H\cap E) - P(G\cap E)P(H)| + 2P(E^c)$. The advantage of this decomposition is that when we restrict $\underline{X}_t^{t-r_1}$ to the set $\mathcal{E}$ (ie. not large values of $\underline{X}_t^{t-r_1}$), we can obtain a bound for $|P(G\cap H\cap E) - P(G\cap E)P(H)|$. More precisely, by using the inequality

$$\inf_{\underline{x}\in\mathcal{E}} P\big(G\big|\underline{X}_t^{t-r_1} = \underline{x}\big)P(H\cap E) \leq P(G\cap H\cap E) \leq \sup_{\underline{x}\in\mathcal{E}} P\big(G\big|\underline{X}_t^{t-r_1} = \underline{x}\big)P(H\cap E),$$

we can derive upper and lower bounds for $P(G\cap H\cap E) - P(G\cap E)P(H)$ which depend only on E and not H and G, and thus obtain the bounds in Proposition 2.1.

It is worth mentioning that by using (7) one can establish mixing rates for time-varying linear processes (such as the tvMA(∞) process considered in Dahlhaus and Polonik (2006)). Using (7) and similar techniques to those used in Section 4, mixing bounds can be obtained for the tvMA(∞) process.

In the following sections we will derive the mixing rates for ARCH-type processes, where one of the challanging aspects of the proof is establishing a bound for the integral difference in (9).

3 Mixing for the time-varying ARCH(p) process

3.1 The tvARCH process

In Fryzlewicz et al. (2008) we show that the tvARCH process can be used to explain the commonly observed stylised facts in financial time series (such as the empirical long memory). A sequence of random variables $\{X_t\}$ is said to come from a time-varying ARCH(p) if it satisfies the representation

$$X_t = Z_t \left(a_0(t) + \sum_{j=1}^{p} a_j(t) X_{t-j} \right), \tag{11}$$

where $\{Z_t\}$ are independent, identically distributed (iid) positive random variables, with $\mathbb{E}(Z_t) = 1$ and $a_j(\cdot)$ are positive parameters. It is worth comparing (11) with the tvARCH process used in the statistical literature. Unlike the tvARCH process considered in, for example, Dahlhaus and Subba Rao (2006) and Fryzlewicz et al. (2008), we have not placed *any* smoothness conditions on the time varying parameters $\{a_j(\cdot)\}$. The smoothness conditions assumed in Dahlhaus and Subba Rao (2006) and Fryzlewicz et al. (2008) are used in order to do parameter estimation. However, in this paper we are dealing with mixing of the process, which does not require such strong assumptions. The assumptions that we require are stated below.

Assumption 3.1 *(i) For some* $\delta > 0$, $\sup_{t\in\mathbb{Z}} \sum_{j=1}^{p} a_j(t) \leq 1 - \delta$.

(ii) $\inf_{t\in\mathbb{Z}} a_0(t) > 0$ *and* $\sup_{t\in\mathbb{Z}} a_0(t) < \infty$.

(iii) Let f_Z *denote the density of* Z_t. *For all* $a > 0$ *we have* $\int |f_Z(u) - f_Z(u[1+a])|\, du \leq K\, a$, *for some finite* K *independent of* a.

(iv) *Let f_Z denote the density of Z_t. For all $a > 0$ we have $\int \sup_{0 \le \tau \le a} |f_Z(u) - f_Z(u[1+\tau])|\, du \le K\, a$, for some finite K independent of a.*

We note that Assumption 3.1(i,ii) guarantees that the ARCH process has a Volterra expansion as a solution (see Dahlhaus and Subba Rao (2006), Section 5). Assumption 3.1(iii,iv) is a type of Lipschitz condition on the density function and is satisfied by various well known distributions, including the chi-squared distributions. We now consider a class of densities which satisfy Assumption 3.1(iii,iv). Suppose $f_Z : \mathbb{R} \to \mathbb{R}$ is a density function, whose first derivative is bounded, after some finite point m, the derivative f' declines monotonically to zero and satisfies $\int |y f_Z'(y)| dy < \infty$. In this case

$$
\begin{aligned}
& \int_0^\infty \sup_{0 \le \tau \le a} |f_Z(u) - f_Z(u[1+\tau])|\, du \\
\le\ & \int_0^m \sup_{0 \le \tau \le a} |f_Z(u) - f_Z(u[1+\tau])|\, du + \int_m^\infty \sup_{0 \le \tau \le a} |f_Z(u) - f_Z(u[1+\tau])|\, du \\
\le\ & a\big(m^2 \sup_{u \in \mathbb{R}} |f_Z'(u)| + \int_m^\infty u |f_Z'(u)| du\big) \le K a,
\end{aligned}
$$

for some finite K independent of a, hence Assumption 3.1(iii,iv) is satisfied.

We use Assumption 3.1(i,ii,iii) to obtain the strong mixing rate (2-mixing and α-mixing) of the tvARCH(p) process and the slightly stronger conditions Assumption 3.1(i,ii,iv) to obtain the β-mixing rate of the tvARCH(p) process. We mention that in the case that $\{X_t\}$ is a stationary, ergodic time series, Francq and Zakoïan (2006) have shown geometric ergodicity, which they show implies β-mixing, under the weaker condition that the distribution function of $\{Z_t\}$ can have some discontinuities.

3.2 The tvARCH(p) process and the Volterra series expansion

In this section we derive a Volterra series expansion of the tvARCH process (see also Giraitis et al. (2000)). These results allow us to apply Proposition 2.1 to the tvARCH process. We first note that the innovations $\underline{Z}_{t+k-1}^{t+1}$ and $\underline{X}_t^{t-p+1}$ are independent random vectors. Hence comparing with Proposition 2.1, we are interested in obtaining the conditional density of X_{t+k} given $\underline{Z}_{t+k-1}^{t+1}$ and $\underline{X}_t^{t-p+1}$, (denoted $f_{0,k,t}$) and the conditional density of X_{t+k+s} given $\underline{X}_{t+k+s-1}^{t+k}, \underline{Z}_{t+k-1}^{t+1}$ and $\underline{X}_t^{t-p+1}$ (denoted $f_{s,k,t}$). We use these expressions to obtain a bound for $\mathcal{D}_{s,k,t}$ (defined in (9)), which we use to derive a bound for the mixing rate. We now represent $\{X_t\}$ in terms of $\{Z_t\}$.

To do this we define

$$
A_t(z) \;=\; \begin{pmatrix} a_1(t)z_t & a_2(t)z_t & \ldots & a_p(t)z_t \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad A_t = A_t(1) = \begin{pmatrix} a_1(t) & a_2(t) & \ldots & a_p(t) \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}
$$

$$
\underline{b}_t(z) \;=\; (a_0(t)z_t, 0, \ldots, 0)' \qquad \text{and} \qquad \underline{X}_t^{t-p+1} = (X_t, X_{t-1}, \ldots, X_{t-p+1})'. \tag{12}
$$

Using this notation we have the relation $\underline{X}_{t+k}^{t+k-p+1} = A_{t+k}(Z)\underline{X}_{t+k-1}^{t+k-p} + \underline{b}_{t+k}(Z)$. We mention the vector representation of ARCH and GARCH processes has been used in Bougerol and Picard (1992), Basrak et al. (2002) and Straumann and Mikosch (2006) in order to obtain some probabilistic properties for ARCH-type processes. Now iterating the relation k times (to get $\underline{X}_{t+k}^{t+k-p+1}$ in terms of $\underline{X}_t^{t-p+1}$) we have

$$
\underline{X}_{t+k}^{t+k-p+1} = \underline{b}_{t+k}(Z) + \sum_{r=0}^{k-2}\left[\prod_{i=0}^{r-1} A_{t+k-i}(Z)\right]\underline{b}_{t+k-r-1}(Z) + \left[\prod_{i=0}^{k-1} A_{t+k-i}(Z)\right]\underline{X}_t^{t-p+1}, \tag{13}
$$

where we set $[\prod_{i=0}^{-1} A_{t+k-i}(Z)] = I_p$ (I_p denotes the $p \times p$ dimensional identity matrix). We use this expansion below.

Lemma 3.1 *Let us suppose that Assumption 3.1(i) is satisfied. Then for $s \geq 0$ we have*

$$
X_{t+k+s} \;=\; Z_{t+k+s}\left\{\mathcal{P}_{s,k,t}(\underline{Z}) + \mathcal{Q}_{s,k,t}(\underline{Z}, \underline{X})\right\}, \tag{14}
$$

where for $s = 0$ and $n > t$ we have $\mathcal{P}_{0,k,t}(\underline{Z}) = a_0(t+k) + [A_{t+k}\sum_{r=0}^{n-t-2}\prod_{i=1}^{r} A_{t+k-i}(Z)b_{t+k-r-1}(Z)]_1$, $\mathcal{Q}_{0,k,t}(\underline{Z}, \underline{X}) = \left[A_{t+k}\prod_{i=1}^{k-1} A_{t+k-i}(Z)\underline{X}_t^{t-p+1}\right]_1$, ($[\cdot]_1$ denotes the first element of a vector).

For $1 \leq s \leq p$

$$
\begin{aligned}
\mathcal{P}_{s,k,t}(\underline{Z}) \;&=\; a_0(t+k+s) + \sum_{i=1}^{s-1} a_i(t+k+s)X_{t+k+s-i} + \sum_{i=s}^{p} a_i(t+k+s)Z_{k+s-i} \\
&\quad \left\{a_0(t+k+s-i) + [A_{t+k+s-i}\sum_{r=1}^{k+s-i}\{\prod_{d=0}^{r} A_{t+k+s-i-d}(Z)\}b_{t+k+s-i-r}(Z)]_1\right\},
\end{aligned} \tag{15}
$$

$$\mathcal{Q}_{s,k,t}(\underline{Z},\underline{X}) = \Big[\sum_{i=s}^{p} a_i(t+k+s)Z_{k+s-i}A_{t+k+s-i}\{\prod_{d=0}^{k+s-i} A_{t+k+s-i-d}(Z)\underline{X}_t^{t-p+1}\}\Big]_1,$$

and for $s > p$ we have $\mathcal{P}_{s,k,t}(\underline{Z}) = a_0(t+k+s) + \sum_{i=1}^{p} a_i(t+k+s)X_{t+k+s-i}$ and $\mathcal{Q}_{s,k,t}(\underline{Z},\underline{X}) \equiv 0$. We note that $\mathcal{P}_{s,k,t}$ and $\mathcal{Q}_{s,k,t}$ are positive random variables, and for $s \geq 1$, $\mathcal{P}_{s,k,t}$ is a function of $\underline{X}_{t+k+s-1}^{t+k}$ (but this has been suppressed in the notation).

PROOF. In Appendix A.2.

By using (14) we now show that the conditional density of X_{t+k+s} given $\underline{X}_{t+k+s-1}^{t+k}, \underline{Z}_{t+k-1}^{t+1}$ and $\underline{X}_t^{t-p+1}$ is a function of the density of Z_{t+k+s}. It is clear from (14) that Z_{t+k+s} can be expressed as $Z_{t+k+s} = \frac{X_{t+k+s}}{\mathcal{P}_{s,k,t}(\underline{Z})+\mathcal{Q}_{s,k,t}(\underline{Z},\underline{X})}$. Therefore, it is straightforward to show that

$$f_{s,k,t}(y_s|\underline{y}_{s-1},\underline{z},\underline{x}) = \frac{1}{\mathcal{P}_{s,k,t}(\underline{z})+\mathcal{Q}_{s,k,t}(\underline{z},\underline{x})} f_Z\left(\frac{y_s}{\mathcal{P}_{s,k,t}(\underline{z})+\mathcal{Q}_{s,k,t}(\underline{z},\underline{x})}\right). \tag{16}$$

3.3 Strong mixing of the tvARCH(p) process

The aim in this section is to prove geometric mixing of the tvARCH(p) process without appealing to geometric ergodicity. Naturally, the results in this section also apply to stationary ARCH(p) processes.

In the following lemma we use Proposition 2.1 to obtain bounds for the mixing rates. It is worth mentioning that the techniques used in the proof below can be applied to other Markov processes.

Lemma 3.2 *Suppose $\{X_t\}$ is a tvARCH process which satisfies (11). Then for any $\underline{\eta} = (\eta_0, \ldots, \eta_{-p+1}) \in (\mathbb{R}^+)^p$ we have*

$$\begin{aligned}
&\sup_{G\in\mathcal{F}_{\infty}^{t+k}, H\in\mathcal{F}_t^{-\infty}} |P(G\cap H) - P(G)P(H)| \\
\leq\ & 2\sum_{s=0}^{p-1}\sup_{\underline{x}\in\mathcal{E}}\int\prod_{i=1}^{k-1} f_Z(z_i) \sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\left\{\int_{\mathbb{R}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1},\underline{z},\underline{x})dy_s\right\}d\underline{z} + 4\sum_{j=0}^{p-1} P(|X_{t-j}| \geq \eta_{-j+1}), \quad (17)
\end{aligned}$$

and

$$\sup_{\{H_j\}\in\mathcal{F}_t^{-\infty},\{G_j\}\in\underline{F}_{\infty}^{t+k}} \sum_{i,j} |P(G_i\cap H_j) - P(G_i)P(H_j)|$$
$$\leq \quad 2\sum_{s=0}^{p-1}\sup_{\underline{x}\in\mathcal{E}}\int\prod_{i=1}^{k-1} f_Z(z_i)\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\left\{\int_{\mathbb{R}}\sup_{\underline{x}\in\mathcal{E}}\mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1},\underline{z},\underline{x})dy_s\right\}d\underline{z} + 4\sum_{j=0}^{p-1} P(|X_{t-j}|\geq \eta_{-j+1}), \quad (18)$$

where $\underline{z} = (z_1,\ldots,z_{k-1})$ *and* $\{G_i\}$ *and* $\{H_j\}$ *are partitions of* Ω.

PROOF In Appendix A.2. □

To obtain a mixing rate for the tvARCH(p) process we need to bound the integral in (17), then obtain the set E which minimises (17). We will start by bounding $\mathcal{D}_{s,k,t}$, which, we recall, is based on the conditional density $f_{s,k,t}$ (defined in (16)).

Lemma 3.3 *Let* $\mathcal{D}_{s,k,t}$ *and* $\mathcal{Q}_{s,k,t}$ *be defined as in (9) and (15) respectively.*

(i) Suppose Assumption 3.1(i,ii,iii) holds, then for all $\underline{x}\in(\mathbb{R}^+)^p$ *we have*

$$\sum_{s=0}^{p-1}\int\prod_{i=1}^{k-1} f_Z(z_i)\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\left\{\int \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1},\underline{z},\underline{x})dy_s\right\}d\underline{z} \leq K\frac{\mathbb{E}[\mathcal{Q}_{s,k,t}(\underline{Z},\underline{x})]}{\inf_{t\in\mathbb{Z}} a_0(t)} \leq K(1-\tilde{\delta})^k\|\underline{x}\|, \quad (19)$$

where K *is a finite constant and* $0<\tilde{\delta}\leq\delta<1$ *(δ is defined in Assumption 3.1(i)).*

(ii) Suppose Assumption 3.1(i,ii,vi) holds, then for any set $\mathcal{E}$ *(defined as in (4)) we have*

$$\sum_{s=0}^{p-1}\int\prod_{i=1}^{k-1} f_Z(z_i)\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\left\{\int\sup_{\underline{x}\in\mathcal{E}}\mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1},\underline{z},\underline{x})dy_s\right\}d\underline{z} \leq \sup_{\underline{x}\in\mathcal{E}} K(1-\tilde{\delta})^k\|\underline{x}\|. \quad (20)$$

PROOF. In Appendix A.2. □

We now use the lemmas above to show geometric mixing of the tvARCH process.

Theorem 3.1 *(i) Suppose Assumption 3.1(i,ii,iii) holds, then*

$$\sup_{\substack{G\in\sigma(\underline{X}_{\infty}^{t+k})\\ H\in\sigma(\underline{X}_t^{-\infty})}} |P(G\cap H) - P(G)P(H)| \leq K\alpha^k,$$

(ii) Suppose Assumption 3.1(i,ii,iv) holds, then

$$\sup_{\substack{\{H_j\}\in\sigma(\underline{X}_t^{-\infty})\\ \{G_j\}\in\sigma(\underline{X}_{\infty}^{t+k})}} \sum_i \sum_j |P(G_i \cap H_j) - P(G_i)P(H_j)| \leq K\alpha^k,$$

for any $\sqrt{1-\delta} < \alpha < 1$, and where K is a finite constant independent of t and k.

PROOF. We use (17) to prove the (i). (19) gives a bound for the integral difference in (17), therefore all that remains is to bound the probabilities in (17). To do this we first use Markov's inequality, to give $\sum_{j=0}^{p-1} P(|X_{t-j}| \geq \eta_{-j}) \leq \sum_{j=0}^{p-1} \mathbb{E}|X_{t-j}|\eta_{-j}^{-1}$. By using the Volterra expansion of X_t (see Dahlhaus and Subba Rao (2006), Section 5) it can be shown that $\sup_{t\in\mathbb{Z}} \mathbb{E}|X_t| \leq (\sup_{t\in\mathbb{Z}} a_0(t))/\sup_{t\in\mathbb{Z}}(1-\sum_{j=1}^{p} a_j(t)))$. Using these bounds and substituting (19) into (17) gives for every $\underline{\eta} \in (\mathbb{R}^+)^p$ the bound

$$\sup_{\substack{G\in\sigma(\underline{X}_{\infty}^{t+k})\\ H\in\sigma(\underline{X}_t^{-\infty})}} |P(G\cap H) - P(G)P(H)| \leq 2\frac{K(1-\tilde{\delta})^k \sum_{j=0}^{p-1}\eta_{-j}}{\inf_{t\in\mathbb{Z}} a_0(t)} + 4K\sum_{j=0}^{p-1}\frac{1}{\eta_{-j}}.$$

We observe the right hand side of the above is minimised when $\eta_{-j} = (1-\tilde{\delta})^{k/2}$ (for $0 \leq j \leq (p-1)$), which gives the bound

$$\sup_{\substack{H\in\sigma(\underline{X}_t^{-\infty})\\ G\in\sigma(\underline{X}_{\infty}^{t+k})}} |P(G\cap H) - P(G)P(H)| \leq K\sqrt{(1-\tilde{\delta})^k}.$$

Since the above is true for any $0 < \tilde{\delta} < \delta$, (ii) is true for any α which satisfies $\sqrt{1-\delta} < \alpha < 1$, thus giving the result.

To prove (ii) we use an identical argument but use the bound in (20) instead of (19), we omit the details. $\square$

Remark 3.1 *We observe that K and α defined in the above theorem are independent of t, therefore under Assumption 3.1(i,ii,iii) we have $\alpha(k) \leq K\alpha^k$ (α-mixing, defined in (1)) and under Assumption 3.1(i,ii,iv) $\beta(k) \leq K\alpha^k$ (β-mixing, defined in (3)) for all $\sqrt{1-\delta} < \alpha < 1$.*

Moreover, since $\sigma(X_{t+k}) \subset \sigma(X_{t+k},\ldots,X_{t+p-1})$ and $\sigma(X_t) \subset \sigma(X_t,\ldots,X_{t-p+1})$ the 2-mixing rate is also geometric with $\tilde{\alpha}(k) \leq K\alpha^k$ ($\tilde{\alpha}(k)$ defined in (2)).

4 Mixing for ARCH(∞) processes

In this section we derive mixing rates for the ARCH(∞) process, we first define the process and state the assumptions that we will use.

4.1 The ARCH(∞) process

The ARCH(∞) process has many interesting features, which are useful in several applications. For example, under certain conditions on the coefficients, the ARCH(∞) process can exhibit 'near long memory' behaviour (see Giraitis et al. (2000)). The ARCH(∞) process satisfies the representation

$$X_t = Z_t \left(a_0 + \sum_{j=1}^{\infty} a_j X_{t-j} \right), \tag{21}$$

where Z_t are iid postive random variables with $\mathbb{E}(Z_t) = 1$ and a_j are positive parameters. The GARCH(p,q) model also has an ARCH(∞) representation, where the a_j decay geometrically with j. Giraitis and Robinson (2001), Robinson and Zaffaroni (2006) and Subba Rao (2006) consider parameter estimation for the ARCH(∞) process.

We will use Assumption 3.1 and the assumptions below.

Assumption 4.1 *(i) We have $\sum_{j=1}^{\infty} a_j < 1-\delta$ and $a_0 > 0$.*

(ii) For some $\nu > 1$, $\mathbb{E}|X_t|^{\nu} < \infty$ (we note that this is fulfilled if $[\mathbb{E}|Z_0^{\nu}|]^{1/\nu} \sum_{j=1}^{\infty} a_j < 1$).

Giraitis et al. (2000) have shown that under Assumption 4.1(i), the ARCH(∞) process has a stationary solution and a finite mean (that is $\sup_{t\in\mathbb{Z}} \mathbb{E}(X_t) < \infty$). It is worth mentioning that since the ARCH(∞) process has a stationary solution the shift t, plays no role when obtaining mixing bounds, ie. $\sup_{G\in\sigma(X_{k+t}),H\in\sigma(X_t)} |P(G\cap H) - P(G)P(H)| = \sup_{G\in\sigma(X_k),H\in\sigma(X_0)} |P(G\cap H) - P(G)P(H)|$. Furthermore, the conditional density of X_{t+k} given $\underline{Z}_{t+k-1}^{t+1}$ and $\underline{X}_t^{-\infty}$ is not a function of t, hence in the section below we let $f_{0,k}$ denote the conditional density of X_{t+k} given ($\underline{Z}_{t+k-1}^{t+1}$ and $\underline{X}_t^{-\infty}$) and for $s \geq 1$, let $f_{s,k}$ denote the conditional density of X_{t+k+s} given ($\underline{X}_{t+k+s-1}^{t+k}, \underline{Z}_{t+k-1}^{t}$ and $\underline{X}_t^{-\infty}$).

4.2 The ARCH(∞) process and the Volterra series expansion

We now write X_k in terms of $\underline{Z}_{k-1}^1$ and $\underline{X} = (X_0, X_{-1}, \ldots)$ and use this to derive the conditional densities $f_{0,k}$ and $f_{s,k}$. It can be seen from the result below (equation (22)) that in general the ARCH(∞) process is not Markovian.

Lemma 4.1 *Suppose $\{X_t\}$ satisfies (21). Then*

$$X_k = \mathcal{P}_{0,k}(\underline{Z})Z_k + \mathcal{Q}_{0,k}(\underline{Z}, \underline{X})Z_k, \tag{22}$$

where

$$\begin{aligned}
\mathcal{P}_{0,k}(\underline{Z}) &= \left[a_0 + \sum_{m=1}^{k} \sum_{k=j_m>\ldots>j_1>0} \left(\prod_{i=1}^{m-1} a_{j_{i+1}-j_i}\right)\left(\prod_{i=1}^{m-1} Z_{j_i}\right)\right] \\
\mathcal{Q}_{0,k}(\underline{Z}, \underline{X}) &= \sum_{r=1}^{k} \left\{\sum_{m=1}^{k} \sum_{k=j_m>\ldots>j_1=r} \left(\prod_{i=1}^{m-1} a_{j_{i+1}-j_i}\right)\left(\prod_{i=1}^{m-1} Z_{j_i}\right)\right\} d_r(\underline{X}).
\end{aligned} \tag{23}$$

Furthermore, setting $\mathcal{Q}_{0,k} = 0$, for $k \geq 1$ we have that $\mathcal{Q}_{0,k}(\underline{Z}, \underline{X})$ satisfies the recursion $\mathcal{Q}_{0,k}(\underline{Z}, \underline{X}) = \sum_{j=1}^{k} a_j \mathcal{Q}_{0,k-j}(\underline{Z}, \underline{X}) Z_{k-j} + d_k(\underline{X})$, where $d_k(\underline{X}) = \sum_{j=0}^{\infty} a_{k+j} X_{-j}$ (for $k \geq 1$).

PROOF. In Appendix A.3 of the Technical report. $\square$

We will use the result above to derive the 2-mixing rate. To derive α and β mixing we require the density of X_{k+s} given $\underline{X}_{k+s-1}^k$, $\underline{Z}_{k-1}^1$ and $\underline{X}_0^{-\infty}$, which uses the following lemma.

Lemma 4.2 *Suppose $\{X_t\}$ satisfies (21). Then for $s \geq 1$ we have*

$$\begin{aligned}
X_{k+s} &= Z_{k+s}\left\{\mathcal{P}_{s,k}(\underline{Z}) + \mathcal{Q}_{s,k}(\underline{Z}, \underline{X})\right\}, \qquad (24) \\
\textit{where } \mathcal{P}_{s,k}(\underline{Z}) &= a_0 + \sum_{j=1}^{s} a_j X_{k+s-j} + \sum_{j=s+1}^{\infty} a_j Z_{k+s-j} \mathcal{P}_{0,k+s-j}(\underline{Z}) \\
\mathcal{Q}_{s,k}(\underline{Z}, \underline{X}) &= \sum_{j=s+1}^{k+s} a_j Z_{k+s-j} \mathcal{Q}_{0,k+s-j}(\underline{Z}, \underline{X}) + d_{k+s}(\underline{X}). \qquad (25)
\end{aligned}$$

PROOF. In Appendix A.3 of the Technical report.

Using (22) and (24) for all $s \geq 0$ we have that $Z_{k+s} = \frac{X_{k+s}}{\mathcal{P}_{s,k}(\underline{Z}) + \mathcal{Q}_{s,k}(\underline{Z},\underline{X})}$, which leads to the conditional densities

$$f_{s,k}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x}) = \frac{1}{\mathcal{P}_{s,k}(\underline{z}) + \mathcal{Q}_{s,k}(\underline{z}, \underline{x})} f_Z\left(\frac{y_s}{\mathcal{P}_{s,k}(\underline{z}) + \mathcal{Q}_{s,k}(\underline{z}, \underline{x})}\right). \tag{26}$$

In the proofs below $\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x})$ plays a prominent role. By using the recursion in Lemma 4.1 and (25), setting $\underline{x} = \underline{X}_0^{-\infty}$ and noting that $\mathbb{E}(\mathcal{Q}_{s,k}(\underline{Z}, \underline{x})) = \mathcal{Q}_{s,k}(\underline{1}_{k-1}, \underline{x})$ we obtain the recursion $\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x}) = \sum_{j=1}^{k} a_{j+s}\mathcal{Q}_{0,k-j}(\underline{1}_{k-j-1}, \underline{x}) + d_{k+s}(\underline{x})$. We use this to obtain a solution for $\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x})$ in terms of $\{d_k(\underline{x})\}_k$ in the lemma below.

Lemma 4.3 *Suppose $\{X_t\}$ satisfies (21) and Assumption 4.1 are fulfilled. Then, there exists $\{\psi_j\}$ such that for all $|z| \leq 1$ we have $(1 - \sum_{j=1}^{\infty} a_j z^j)^{-1} = \sum_{j=0}^{\infty} \psi_j z^j$. Furthermore, if $\sum_j |j^\alpha a_j| < \infty$, then Hannan and Kavaliers (1986) have shown that $\sum_j |j^\alpha \psi_j| < \infty$. For $k \leq 0$, set $d_k(\underline{x}) = 0$ and $\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x}) = 0$, then for $k \geq 1$, $\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x})$ has the solution*

$$\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x}) = \sum_{j=0}^{\infty} \psi_j d_{k-j}(\underline{x}) = \sum_{j=0}^{k-1} \psi_j d_{k-j}(\underline{x}) = \sum_{j=0}^{k-1} \psi_j \left\{\sum_{i=0}^{\infty} a_{k-j+i} x_{-i}\right\}, \tag{27}$$

where $\underline{x} = (x_0, x_{-1}, \ldots)$.

PROOF. In Appendix A.3 of the Technical report. □

4.3 Mixing for ARCH(∞)-type processes

In this section we show that the mixing rates are not necessarily geometric and depend on the rate of decay of the coefficients $\{a_j\}$ (we illustrate this in the following example). Furthermore for ARCH(∞) processes the strong mixing rate and 2-mixing rate can be different.

Example 4.1 *Let us consider the ARCH(∞) process, $\{X_t\}$, defined in (21). Giraitis et al. (2000) have shown that if $a_j \sim j^{-(1+\delta)}$ (for some $\delta > 0$) and $[\mathbb{E}(Z_t^2)]^{1/2} \sum_{j=1}^{\infty} a_j < 1$, then $|cov(X_0, X_k)| \sim k^{-(1+\delta)}$. That is, the absolute sum of the covariances is finite, but 'only just' if δ is small. If $Z_t < 1$, it is straightforward to see that X_t is a bounded random variable and by using Ibragimov's inequality (see Hall and Heyde (1980)) we have*

$$|cov(X_0, X_k)| \leq C \sup_{A \in \sigma(X_0), B \in \sigma(X_k)} |P(A \cap B) - P(A)P(B)|,$$

for some $C < \infty$. Noting that $|cov(X_0, X_k)| = O(k^{-(1+\delta)})$ this gives a lower bound of $O(k^{-(1+\delta)})$ on the 2-mixing rate. $\square$

To obtain the mixing rates we will use Proposition 2.1, this result requires bounds on $\mathcal{D}_{s,k} = |f_{s,k}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x}) - f_{s,k}(y_s|\underline{y}_{s-1}, \underline{z}, 0)|$ and its integral.

Lemma 4.4 *Suppose $\{X_t\}$ satisfies (21), f_Z is the density of Z_t and let $\mathcal{D}_{s,k}$ and $\mathcal{Q}_{0,k}(\cdot)$ be defined as in (9) and (23). Suppose Assumptions 3.1(iii) and 4.1 are fulfilled, then*

$$\int \prod_{i=1}^{k-1} f_Z(z_i)\{\int |f_{0,k}(y|\underline{z}, \underline{x}) - f_{0,k}(y|\underline{z}, 0)|dy\}d\underline{z} \leq \frac{\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x})}{a_0} = \sum_{j=0}^{k-1} |\psi_j| \left\{\sum_{i=0}^{\infty} a_{k-j+i}x_{-i}\right\} \tag{28}$$

and for $s \geq 1$

$$\int \prod_{i=1}^{k-1} f_Z(z_i)\{\sup_{\underline{y}_{s-1}\in\mathbb{R}^s} \int \mathcal{D}_{s,k}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})dy_s\}d\underline{z} \leq \frac{1}{a_0}\{\sum_{j=s+1}^{k+s} a_j \sum_{l=0}^{k+s-j} |\psi_l| \sum_{i=0}^{\infty} a_{k+s-j-l+i}x_{-i} + \sum_{i=0}^{\infty} a_{k+s+i}x_{-}$$

Suppose Assumptions 3.1(iv) and 4.1 are fulfilled, and $\mathcal{E}$ is defined as in (4) then

$$\begin{aligned}&\int \prod_{i=1}^{k-1} f_Z(z_i)\{\sup_{\underline{y}_{s-1}\in\mathbb{R}^s} \int \sup_{\underline{x}\in\mathcal{E}} \mathcal{D}_{s,k}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})dy_s\}d\underline{z} \\ \leq\ & \frac{1}{a_0}\{\sum_{j=s+1}^{k+s} a_j \sum_{l=0}^{k+s-j} |\psi_l| \sum_{i=0}^{\infty} a_{k+s-j-l+i}\eta_{-i} + \sum_{i=0}^{\infty} a_{k+s+i}\eta_{-i}\},\end{aligned} \tag{30}$$

where $\underline{x} = (x_0, x_{-1}, \ldots)$ is a positive vector.

PROOF. In Appendix A.3 of the Technical report. $\square$

We require the following simple lemma to prove the theorem below.

Lemma 4.5 *Let us suppose that $\{c_i\}$, $\{d_i\}$ and $\{\eta_{-i}\}$ are positive sequences, then*

$$\inf_{\underline{\eta}} \{\sum_{i=0}^{\infty}(c_i\eta_{-i} + d_i\eta_{-i}^{-\nu})\} = (\nu^{\frac{1}{1+\nu}} + \nu^{-\frac{\nu}{\nu+1}})\sum_{i=0}^{\infty} c_i^{\frac{\nu}{\nu+1}} d_i^{\frac{1}{\nu+1}}. \tag{31}$$

PROOF. In Appendix A.3 of the Technical report. $\square$

In the following theorem we obtain α-mixing and β-mixing bounds for the ARCH(∞) process.

Theorem 4.1 *Suppose $\{X_t\}$ satisfies (21).*

(a) Suppose Assumptions 3.1(iii) and 4.1 hold. Then, we have

$$\begin{aligned}
& \sup_{G\in\mathcal{F}_{\infty}^{k},H\in\mathcal{F}_{0}^{-\infty}} |P(G\cap H) - P(G)P(H)| \\
\leq\ & K(\nu)\sum_{i=0}^{\infty}\left[\frac{1}{a_0}\sum_{s=0}^{\infty}\sum_{j=s+1}^{k+s} a_j \sum_{l=0}^{k+s-j}|\psi_l| a_{k+s-j-l+i} + \frac{1}{a_0}\sum_{s=0}^{\infty} a_{k+s+i}\right]^{\frac{\nu}{\nu+1}} [\mathbb{E}|X_0|^{\nu}]^{\frac{1}{\nu+1}} \quad (32)
\end{aligned}$$

where $K(\nu) = 3(\nu^{\frac{1}{1+\nu}} + \nu^{-\frac{\nu}{\nu+1}})$.

(i) If the parameters of the ARCH(∞) process satisfy $|a_j| \sim j^{-\delta}$ and the $|\psi_j| \sim j^{-\delta}$ (defined in Lemma 4.3), then we have

$$\sup_{G\in\mathcal{F}_{\infty}^{k},H\in\mathcal{F}_{0}^{-\infty}} |P(G\cap H) - P(G)P(H)| \leq K\cdot\left[k(k+1)^{-\tilde{\delta}+3} + (k+1)^{-\tilde{\delta}+2}\right],$$

where $\tilde{\delta} = \delta\times(\frac{\nu}{\nu+1})$.

(ii) If the parameters of the ARCH(∞) process satisfy $|a_j| \sim \delta^j$ and $\psi_j \sim \delta^j$, where $0<\delta<1$ (an example is the GARCH(p,q) process), then we have

$$\sup_{G\in\mathcal{F}_{\infty}^{k},H\in\mathcal{F}_{-\infty}^{0}} |P(G\cap H) - P(G)P(H)| \ \leq\ C\cdot k\cdot\delta^{k/2}$$

where C is a finite constant.

(b) Suppose Assumptions 3.1(iv) and 4.1 hold. Then, we have

$$\begin{aligned}
& \sup_{\{G_i\}\in\mathcal{F}_{\infty}^{k},\{H_j\}\in\mathcal{F}_{0}^{-\infty}} \sum_i\sum_j |P(G_i\cap H_j) - P(G_i)P(H_j)| \\
\leq\ & K(\nu)\sum_{i=0}^{\infty}\left[\frac{1}{a_0}\sum_{s=0}^{\infty}\sum_{j=s+1}^{k+s} a_j \sum_{l=0}^{k+s-j}|\psi_l| a_{k+s-j-l+i} + \frac{1}{a_0}\sum_{s=0}^{\infty} a_{k+s+i}\right]^{\frac{\nu}{\nu+1}} [\mathbb{E}|X_0|^{\nu}]^{\frac{1}{\nu+1}}, \quad (33)
\end{aligned}$$

where $\{G_i\}$ and $\{H_j\}$ are partitions of Ω. We mention that the bounds for the α-mixing rates for different orders of $\{a_j\}$ and $\{\psi_j\}$ derived in (i) also hold for the β-mixing rate.

PROOF. We first prove (a). We use that

$$\sup_{G\in\mathcal{F}_{\infty}^{k},H\in\mathcal{F}_{0}^{-\infty}}|P(G\cap H)-P(G)P(H)| = \lim_{n\to\infty}\sup_{G\in\mathcal{F}_{k+n}^{k}} H\in\mathcal{F}_{0}^{-\infty}|P(G\cap H)-P(G)P(H)|,$$

and find a bound for each n. By using (5) to bound $\sup_{G\in\mathcal{F}_{k+n}^{k},H\in\mathcal{F}_{0}^{-\infty}}|P(G\cap H)-P(G)P(H)|$ we see that for all sets $\mathcal{E}$ (as defined in (4)) we have

$$\begin{aligned}
&\sup_{G\in\mathcal{F}_{k+n}^{k},H\in\underline{F}_{0}^{-\infty}}|P(G\cap H)-P(G)P(H)| \qquad (34)\\
\leq\ & 2\sup_{\underline{x}\in\mathcal{E}}\sum_{s=0}^{n}\int_{\mathbb{R}^k}\prod_{i=1}^{k-1}f_Z(z_i)\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\Big\{\int\mathcal{D}_{s,k}(y_s|\underline{y}_{s-1},\underline{z},\underline{x})dy_s\Big\}d\underline{z}+4P(X_0>\eta_0 \text{ or } \ldots X_{-n}>\eta_{-n}).
\end{aligned}$$

To bound the integral in (34) we use (29) to obtain

$$\begin{aligned}
&\sup_{\underline{x}\in\mathcal{E}}\sum_{s=0}^{n}\int_{\mathbb{R}^k}\prod_{i=1}^{k-1}f_Z(z_i)\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\int_{\mathbb{R}}\mathcal{D}_{s,k}(y_s|\underline{y}_{s-1},\underline{z},\underline{x})dy_s d\underline{z}\\
=\ & \frac{1}{a_0}\sum_{s=0}^{n}\Big\{\sum_{j=s+1}^{k+s}a_j\sum_{l=0}^{k+s-j}|\psi_l|\sum_{i=0}^{\infty}a_{k+s-j-l+i}\eta_{-i}+\sum_{i=0}^{\infty}a_{k+s+i}\eta_{-i}\Big\}. \qquad (35)
\end{aligned}$$

Now by using Markov's inequality we have that $P(X_0>\eta_0 \text{ or } ,\ldots,X_{-n}\geq\eta_{-n})\leq\sum_{i=0}^{n}\frac{\mathbb{E}(|X_i|^{\nu})}{\eta_{-i}^{\nu}}$. Substituting (35) and the above into (34) and letting $n\to\infty$ gives

$$\begin{aligned}
&\sup_{G\in\mathcal{F}_{\infty}^{k},H\in\mathcal{F}_{0}^{-\infty}}|P(G\cap H)-P(G)P(H)|\\
\leq\ & \inf_{\eta}\Big[\frac{2}{a_0}\sum_{s=0}^{\infty}\Big\{\sum_{j=s+1}^{k+s}a_j\sum_{l=0}^{k+s-j}|\psi_l|\sum_{i=0}^{\infty}a_{k+s-j-l+i}\eta_{-i}+\sum_{i=0}^{\infty}a_{k+s+i}\eta_{-i}\Big\}+4\mathbb{E}|X_0|^{\nu}\sum_{i=0}^{\infty}\eta_{-i}^{-\nu}\Big] \qquad (36)
\end{aligned}$$

where $\eta=(\eta_0,\eta_{-1},\ldots)$.

Now we use (31) to minimise (36), which gives us (33). The proof of (i) can be found in the appendix. It is straightforward to prove (ii) using (31).

The proof of (b) is very similar to the proof of (a) but uses (30) rather than (29), we omit the details. □

Remark 4.1 *Under the assumptions in Theorem 4.1(a) we have a bound for the α-mixing rate, that is $\alpha(k)\leq\zeta(k)$, where $\zeta(k)=K\big[\frac{1}{a_0}\sum_{s=0}^{\infty}\sum_{j=s+1}^{k+s}a_j\sum_{l=0}^{k+s-j}|\psi_l|a_{k+s-j-l+i}+\frac{1}{a_0}\sum_{s=0}^{\infty}a_{k+s+i}\big]^{\frac{\nu}{\nu+1}}$. Under the assumptions in Theorem 4.1(a) the β-mixing coefficient is bounded by $\beta(k)\leq\zeta(k)$.*

In the following theorem we consider a bound for the 2-mixing rate of an ARCH(∞) process.

Theorem 4.2 *Suppose $\{X_t\}$ satisfies (21) and Assumptions 3.1(iii) and 4.1 holds. Then we have*

$$\sup_{G\in\sigma(X_k),H\in\mathcal{F}_0^{-\infty}} |P(G\cap H)-P(G)P(H)| \leq K(\nu)\sum_{i=0}^{\infty}\left[\frac{1}{a_0}\sum_{j=0}^{k-1}a_j|\psi_j|a_{k-j+i}\right]^{\frac{\nu}{\nu+1}}[\mathbb{E}|X_0|^{\nu}]^{\frac{1}{\nu+1}} \quad (37)$$

where $K(\nu)=3(\nu^{\frac{1}{1+\nu}}+\nu^{-\frac{\nu}{\nu+1}})$.

If the parameters of the ARCH(∞) process satisfy $a_j\sim j^{-\delta}$ and $|\psi_j|\sim j^{-\delta}$ (ψ_j defined in Lemma 4.3), then we have

$$\sup_{G\in\sigma(X_k),H\in\mathcal{F}_0^{-\infty}} |P(G\cap H)-P(G)P(H)| \leq K\cdot k(k+1)^{-\tilde{\delta}+1} \quad (38)$$

where $\tilde{\delta}=\delta\times(\frac{\nu}{\nu+1})$.

PROOF. We use a similar proof to the proof of Theorem 4.1. The integral difference is replaced with the bound in (28) and again we use Markov's inequality, together they give the bound

$$\sup_{G\in\sigma(X_k),H\in\mathcal{F}_0^{-\infty}} |P(G\cap H)-P(G)P(H)| \leq \inf_{\underline{\eta}}\left[2\frac{1}{a_0}\sum_{j=0}^{k-1}|\psi_j|\left\{\sum_{i=0}^{\infty}a_{k-j+i}\eta_{-i}\right\}+4\mathbb{E}|X_0|^{\nu}\sum_{i=0}^{\infty}\frac{1}{\eta_{-i}^{\nu}}\right]. \quad (39)$$

Finally to obtain (37) and (38) we use (39) and a proof similar to Theorem 4.1(i) hence we omit the details. $\square$

Remark 4.2 *Comparing (38) and Theorem 4.1(i) we see that the 2-mixing bound is of a smaller order than the strong mixing bound.*

In fact, it could well be that the 2-mixing bound is of a smaller order than Theorem 4.2(i). This is because Theorem 4.2(i) gives a bound for $\sup_{G\in\sigma(X_k),H\in\sigma(X_0,X_{-1},\ldots)}|P(G\cap H)-P(G)P(H)|$ whereas the 2-mixing bound restricts the sigma-algebra of the left tail to $\sigma(X_0)$. However, we have not been able to show this and this is a problem that requires further consideration.

Acknowledgments

We would like to thank Piotr Kokoszka, Mika Meitz and Joseph Tadjuidje for several useful discussions. We also thank the Associate Editor and two anonymous referees for suggestions and comments which greatly improved many aspects of the paper. SSR is partially supported by an NSF grant under DMS-0806096 and the Deutsche Forschungsgemeinschaft under DA 187/15-1.

A Proofs

A.1 Proof of Proposition 2.1

We will use the following three lemmas to prove Proposition 2.1.

Lemma A.1 *Let $G \in \mathcal{F}_{t+k+r_2}^{t+k} = \sigma(\underline{X}_{t+k+r_2}^{t+k})$ and $H, E \in \mathcal{F}_t^{t-r_1} = \sigma(\underline{X}_t^{t-r_1})$ (where E is defined in (4)), and use the notation in Proposition 2.1. Then we have*

$$\begin{aligned} & |P(G \cap H \cap E) - P(G \cap E)P(H)| \qquad (40) \\ \leq \;& 2P(H) \sup_{\underline{x} \in \mathcal{E}} \left| P(G | \underline{X}_t^{t-r_1} = \underline{x}) - P(G | \underline{X}_t^{t-r_1} = 0) \right| + \inf_{\underline{x} \in \mathcal{E}} P(G | \underline{X}_t^{t-r_1} = \underline{x}) \Big\{ P(H)P(E^c) + P(H \cap E^c) \Big\} \end{aligned}$$

PROOF. To prove the result we first observe that

$$\begin{aligned} P(G \cap H \cap E) &= P\big(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G}, \underline{X}_t^{t-r_1} \in (\mathcal{H} \cap \mathcal{E})\big) = \int_{\mathcal{H} \cap \mathcal{E}} \int_{\mathcal{G}} dP(\underline{X}_t^{t-r_1} \leq \underline{y}, \underline{X}_{t+k+r_2}^{t+k} \leq \underline{x}) \\ &= \int_{\mathcal{H} \cap \mathcal{E}} \left\{ \int_{\mathcal{G}} dP(\underline{X}_{t+k+r_2}^{t+k} \leq \underline{y} | \underline{X}_t^{t-r_1} = \underline{x}) \right\} dP(\underline{X}_t^{t-r_1} \leq \underline{x}) \\ &= \int_{\mathcal{H} \cap \mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_t^{t-r_1} = \underline{x}) dP(\underline{X}_t^{t-r_1} \leq \underline{x}). \qquad (41) \end{aligned}$$

Therefore, by using the above and that $P(H \cap E) \leq P(H)$ we obtain the following inequalities

$$\begin{aligned} \inf_{\underline{x} \in \mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_t^{t-r_1} = \underline{x}) P(H \cap E) &\leq P(G \cap H \cap E) \leq \sup_{\underline{x} \in \mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_t^{t-r_1} = \underline{x}) P(H) \qquad (42) \\ \inf_{\underline{x} \in \mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_t^{t-r_1} = \underline{x}) P(E) &\leq P(G \cap E) \leq \sup_{\underline{x} \in \mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_t^{t-r_1} = \underline{x}) P(E). \qquad (43) \end{aligned}$$

Subtracting (42) from (43) and using $P(H \cap E) = P(H) - P(H \cap E^c)$ give the inequalities

$$\begin{aligned} P(G \cap H \cap E) - P(G \cap E)P(H) &\leq \sup_{\underline{x} \in \mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_t^{t-r_1} = \underline{x}) P(H) \\ &\quad - \inf_{\underline{x} \in \mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_t^{t-r_1} = \underline{x}) P(H) + P(E^c)P(H) \qquad (44) \\ P(G \cap H \cap E) - P(G \cap E)P(H) &\geq \inf_{\underline{x} \in \mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_t^{t-r_1} = \underline{x}) P(H) \\ &\quad - \sup_{\underline{x} \in \mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_t^{t-r_1} = \underline{x}) P(H) - P(E^c \cap H). \qquad (45) \end{aligned}$$

Combining (44) and (45) we obtain

$$\begin{aligned}
& |P(G \cap H \cap E) - P(G \cap E)P(H)| \\
\leq\ & P(H)\Big| \sup_{\underline{x} \in \mathcal{E}} P(G|\underline{X}_t^{t-r_1} = \underline{x}) - \inf_{\underline{x} \in \mathcal{E}} P(G|\underline{X}_t^{t-r_1} = \underline{x})\Big| + \inf_{\underline{x} \in \mathcal{E}} P(G|\underline{X}_t^{t-r_1} = \underline{x})\Big\{P(H)P(E^c) + P(H \cap E^c
\end{aligned}$$

Using the triangle inequality we have

$$\left| \sup_{\underline{x} \in \mathcal{E}} P(G|\underline{X}_t^{t-r_1} = \underline{x}) - \inf_{\underline{x} \in \mathcal{E}} P(G|\underline{X}_t^{t-r_1} = \underline{x}) \right| \leq 2 \sup_{\underline{x} \in \mathcal{E}} \left| P(G|\underline{X}_t^{t-r_1} = \underline{x}) - P(G|\underline{X}_t^{t-r_1} = 0) \right|.$$

Substituting the above into (46) gives (40), as required. □

We now obtain a bound for the first term on the right hand side of (40).

Lemma A.2 *Let $f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}$ denote the density of $\underline{X}_{t+k+r_2}^{t+k}$ given $\underline{X}_t^{t-r_1}$ and $\mathcal{G}$ and $\mathcal{H}$ be defined as in (10), then*

$$\left|P(G|\underline{X}_t^{t-r_1} = \underline{x}) - P(G|\underline{X}_t^{t-r_1} = 0)\right| \leq \int_{\mathcal{G}} \mathcal{D}_{0,k,t}(\underline{y}|\underline{x})d\underline{y}. \tag{47}$$

Let $\underline{W}_{t+k-1}^{t+1}$ be a random vector which is independent of $\underline{X}_t^{t-r_1}$ and let $f_{\underline{W}}$ denote the density of $\underline{W}_{t+k-1}^{t+1}$. If $G \in \sigma(X_{t+k})$ then

$$\int_{\mathcal{G}} \left|f_{X_{t+k}|\underline{X}_t^{t-r_1}}(y|\underline{x}) - f_{X_{t+k}|\underline{X}_t^{t-r_1}}(y|0)\right|dy \leq \int_{\mathbb{R}^{k-1}} f_{\underline{W}}(\underline{w})\left\{\int_{\mathcal{G}} \mathcal{D}_{0,k,t}(y|\underline{w}, \underline{x})dy\right\}d\underline{w} \tag{48}$$

and if $G \in \sigma(\underline{X}_{t+k+r_2}^{t+k})$ then

$$\int_{\mathcal{G}} \left|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) - f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\right|d\underline{y} \leq \sum_{s=0}^{r_2} \int_{\mathbb{R}^{k-1}} f_{\underline{W}}(\underline{w})\left\{\sup_{\underline{y}_{s-1}} \int_{\mathbf{G}_s} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x})dy_s\right\}d\underline{w}, (49$$

where $\mathcal{G} = \mathbf{G}_1 \otimes \ldots \otimes \mathbf{G}_n$, and $\mathbf{G}_j \subset \mathbb{R}$.

PROOF. The proof of (47) is clear, hence we omit the details.

To prove (48) we first note that by independence of $\underline{W}_{t+k-1}^{t+1}$ and $\underline{X}_t^{t-r_2}$ we have that $f_{\underline{W}|\underline{X}_t^{t-r_1}}(\underline{w}|\underline{x}) = f_{\underline{W}}(\underline{w})$, where $f_{\underline{W}|\underline{X}_t^{t-r_1}}$ is the conditional density of $\underline{W}_{t+k-1}^{t+1}$ given $\underline{X}_t^{t-r_1}$. Therefore we have

$$f_{X_{t+k}|\underline{X}_t^{t-r_1}}(y|\underline{x}) = \int_{\mathbb{R}^{k-1}} f_{X_{t+k}|\underline{W},\underline{X}_t^{t-r_1}}(y|\underline{w}, \underline{x})f_{\underline{W}}(\underline{w})d\underline{w} = \int_{\mathbb{R}^{k-1}} f_{0,k,t}(y|\underline{w}, \underline{x})f_{\underline{W}}(\underline{w})d\underline{w}.$$

Now substituting the above into $\int_{\mathcal{G}} \left| f_{X_{t+k}|\underline{X}_t^{t-r_1}}(y|\underline{x}) - f_{X_{t+k}|\underline{X}_t^{t-r_1}}(y|0) \right| dy$ gives (48).

To prove (49) we note by using the same argument to prove (48) we have

$$f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) = \int_{\mathbb{R}^{k-1}} f_{\underline{W}}(\underline{w}) \prod_{s=0}^{r_2} f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) d\underline{w}. \tag{50}$$

Now repeatedly subtracting and adding $f_{s,k,t}$ from the above gives

$$f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) - f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0) = \sum_{s=0}^{r_2} \int_{\mathbb{R}^{k-1}} f_{\underline{W}}(\underline{w}) \left\{ \prod_{a=0}^{s-1} f_{a,k,t}(y_a|\underline{y}_{a-1}, \underline{w}, \underline{x}) \right\} \times \left\{ \prod_{b=s+1}^{r_2} f_{b,k,t}(y_b|\underline{y}_{b-1}, \underline{w}, 0) \right\} \left\{ f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) - f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, 0) \right\} d\underline{w}. \tag{51}$$

Therefore taking the integral of the above over $\mathcal{G}$ gives

$$\int_{\mathcal{G}} \left| f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) - f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0) \right| d\underline{y} \leq \sum_{s=0}^{r_2} \int_{\mathbb{R}^k} f_{\underline{W}}(\underline{w}) \left\{ \left[\prod_{a=0}^{s-1} \int_{\mathcal{G}_a} f_{a,k,t}(y_a|\underline{y}_{a-1}, \underline{w}, \underline{x}) dy_a \times \prod_{b=s+1}^{r_2} \int_{\mathcal{G}_b} f_{b,k,t}(y_b|\underline{y}_{b-1}, \underline{w}, \underline{x}) dy_b \right] \times \sup_{\underline{y}_{s-1}} \int_{\mathcal{G}_s} \left| f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) - f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, 0) \right| dy_s \right\} d\underline{w}. \tag{52}$$

To obtain (49) we observe that since $\mathcal{G}_j \subset \mathbb{R}$ and $\int_{\mathbb{R}} f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) dy_s = 1$ we have $\left(\prod_{a=0}^{s-1} \int_{\mathcal{G}_a} f_{a,k,t}(y_a|\underline{y}_{a-1}, \underline{w}, \underline{x}) dy_a \right) \left(\prod_{b=s+1}^{r_2} \int_{\mathcal{G}_b} f_{b,k,t}(y_b|\underline{y}_{b-1}, \underline{w}, \underline{x}) dy_b \right) \leq 1$. Finally substituting the above upper bound into (52) gives (49). □

The following lemma will be used to show β-mixing and uses the above lemmas.

Lemma A.3 *Suppose that $\{G_i\} \in \mathcal{F}_{t+k+r_2}^{t+k}$, $\{H_j\} \in \mathcal{F}_t^{t-r_1}$ and $\{G_i\}$ and $\{H_j\}$ are partitions of Ω. Then we have*

$$\begin{aligned} & \sum_{i,j} |P(G_i \cap H_j \cap E) - P(G_i \cap E)P(H_j)| \\ \leq \; & 2\sum_i \sup_{\underline{x} \in \mathcal{E}} |P(G_i|\underline{X}_t^{t-r_1} = \underline{x}) - P(G_i|\underline{X}_t^{t-r_1} = 0)| + 2P(E^c) \\ \textit{and} \quad & \sum_{i,j} |P(G_i \cap H_j \cap E^c) - P(G_i \cap E^c)P(H_j)| \leq 2P(E^c). \end{aligned} \tag{53}$$

PROOF. Substituting the inequality in (40) into $\sum_{i,j}|P(G_i\cap H_j\cap E)-P(G_i\cap E)P(H_j)|$ gives

$$\begin{aligned}
&\sum_{i,j}|P(G_i\cap H_j\cap E)-P(G_i\cap E)P(H_j)|\\
\leq\ & 2\sum_j P(H_j)\sum_i \sup_{\underline{x}\in\mathcal{E}}\big|P(G_i|\underline{X}_t^{t-r_1}=\underline{x})-P(G_i|\underline{X}_t^{t-r_1}=0)\big|+\\
&\sum_{i,j}\inf_{\underline{x}\in\mathcal{E}}P(G_i|\underline{X}_t^{t-r_1}=\underline{x})\big\{P(H_j)P(E^c)+P(H_j\cap E^c)\big\}. \qquad (54)
\end{aligned}$$

The sets $\{H_j\}$ are partitions of Ω, hence $\sum_i P(H_j)=1$ and $\sum_i P(H_j\cap E^c)\leq 1$. Using these observations together with (54) gives (53).

(53) immediately follows from the fact that $\{H_j\}$ and $\{G_i\}$ are disjoint sets gives (53). $\Box$

Using the above three lemmas we can now prove Proposition 2.1.

PROOF of Proposition 2.1 equation (5) It is straightforward to show that

$$\big|P(G\cap H)-P(G)P(H)\big|\leq\big|P(G\cap H\cap E)-P(G\cap E)P(H)\big|+\big|P(G\cap H\cap E^c)-P(G\cap E^c)P(H)\big|.$$

Now by substituting (47) into Lemma A.1 and using the above gives

$$\begin{aligned}
&\big|P(G\cap H)-P(G)P(H)\big|\leq 2\sup_{\underline{x}\in\mathcal{E}}\int_{\mathcal{G}}\big|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x})-f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\big|d\underline{y}+\\
&\inf_{\underline{x}\in\mathcal{E}}P(G|\underline{X}_t^{t-r_1}=\underline{x})\big\{P(H)P(E^c)+P(H\cap E^c)\big\}+P(G\cap H\cap E^c)+P(G\cap E^c)P(H).
\end{aligned}$$

Finally by using that $\mathcal{G}\subset\mathbb{R}^{r_2+1}$, $P(G\cap H\cap E^c)\leq P(E^c)$, $P(G\cap E^c)P(H)\leq P(E^c)$ and $\inf_{\underline{x}\in\mathcal{E}}P(G|\underline{X}_t^{t-r_1}=\underline{x})\leq 1$ we obtain (5). $\Box$

PROOF of Proposition 2.1 equation (6) It is worth noting that the proof of (6) is similar to the proof of (5). Using (53) and the same arguments as in the proof of (5) we have

$$\begin{aligned}
&\sum_{i,j}|P(G_i\cap H_j)-P(G_i)P(H_j)|\leq 2\sum_i\sup_{\underline{x}\in\mathcal{E}}\int_{\mathcal{G}_i}\Big|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x})-f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\Big|d\underline{y}+4P(E^c)\\
&\leq 2\sum_i\int_{\mathcal{G}_i}\sup_{\underline{x}\in\mathcal{E}}\Big|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x})-f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\Big|d\underline{y}+4P(E^c) \qquad (55)\\
&\leq 2\int_{\mathbb{R}^{r_2+1}}\sup_{\underline{x}\in\mathcal{E}}\Big|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x})-f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\Big|d\underline{y}+4P(E^c),
\end{aligned}$$

where $H_j=\{\omega;\ \underline{X}_t^{t-r_1}(\omega)\in\mathcal{H}_j\}$ and $G_i=\{\omega;\ \underline{X}_{t+k+r_2}^{t+k}(\omega)\in\mathcal{G}_i\}$, which gives (6). $\Box$

PROOF of Proposition 2.1 equation (7). To prove (7) we note that $\int_{\mathbf{G}_s} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x})dy_s \leq \int_{\mathbb{R}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x})dy_s$. Now by substituting this inequality into (49) and what results into (5) gives (7). $\square$

PROOF of Proposition 2.1 equation (8). To prove (8) we use (49) and that for all positive functions f, $\sum_i \int_{\mathcal{G}_{s,i}} f(u)du \leq \int_{\mathbb{R}} f(u)du$ we have

$$\begin{aligned} &\sum_i \int_{\mathcal{G}_i} \sup_{\underline{x}\in\mathcal{E}} \big|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) - f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\big|d\underline{y} \\ \leq\ & \sum_i \sum_{s=0}^{r_2} \int_{\mathbb{R}^k} f_{\underline{W}}(\underline{w}) \sup_{\underline{y}_{s-1}\in(\mathbb{R}^+)^s} \Big\{ \int_{\mathbf{G}_{s,i}} \sup_{\underline{x}\in\mathcal{E}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x})dy_s \Big\} d\underline{w} \\ \leq\ & \sum_{s=0}^{r_2} \int_{\mathbb{R}^k} f_{\underline{W}}(\underline{w}) \sup_{\underline{y}_{s-1}\in(\mathbb{R}^+)^s} \Big\{ \int_{\mathbb{R}} \sup_{\underline{x}\in\mathcal{E}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x})dy_s \Big\} d\underline{w} \end{aligned}$$

where $\mathcal{G}_s = \mathbf{G}_{s,1} \otimes \ldots \otimes \mathbf{G}_{s,n}$ and $\mathbf{G}_{s,i} \subset \mathbb{R}$. Finally substituting the above into the right hand side of (6) gives (8). $\square$

A.2 Proofs in Section 3

PROOF of Lemma 3.2 We first note that since $\{X_t\}$ satisfies a tvARCH(p) representation ($p < \infty$) it is p-Markovian, hence for any $r_2 > p$ the sigma-algebras generated by $\underline{X}_{t+k+r_2}^{t+k}$ and $\underline{Z}_{t+k+r_2}^{t+k+p}, \underline{X}_{t+k}^{t+k+p-1}$ are the same. Moreover, by using that for all $\tau > t$, Z_τ is independent of X_τ we have

$$\sup_{G\in\mathcal{F}_\infty^{t+k}, H\in\mathcal{F}_t^{-\infty}} |P(G\cap H) - P(G)P(H)| = \sup_{G\in\mathcal{F}_{t+k+p-1}^{t+k}, H\in\mathcal{F}_t^{t-p+1}} |P(G\cap H) - P(G)P(H)|. \quad (56)$$

Now by using the above, Proposition 2.1, equation (7), and that $\underline{Z}_{t+k-1}^{t+1}$ and $\underline{X}_t^{t-p+1}$ are independent, for any set $\mathcal{E}$ (defined as in (4)) we have

$$\begin{aligned} &\sup_{G\in\mathcal{F}_{t+k+p-1}^{t+k}, H\in\mathcal{F}_t^{t-p+1}} |P(G\cap H) - P(G)P(H)| \\ \leq\ & 2\sup_{\underline{x}\in\mathcal{E}} \sum_{s=0}^{p-1} \int \prod_{i=1}^{k-1} f_Z(z_i) \sup_{\underline{y}_{s-1}\in\mathbb{R}^s} \Big\{ \int \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})dy_s \Big\} d\underline{z} + 4P(X_t > \eta_0 \text{ or } \ldots X_{t-p+1} > \eta_{-p+1}). \end{aligned} \quad (5$$

Finally using that $P(X_t > \eta_0 \text{ or } X_{t-1} > \eta_{-1} \ldots X_{t-p+1} > \eta_{-p+1}) \leq \sum_{j=0}^{p-1} P(X_{t-j} > \eta_{-j})$ gives (17).

The proof of (18) uses a similar proof as that given above, but uses (8) instead of (7), we omit the details. $\square$

PROOF of Lemma 3.1 We first prove (14) with $s = 0$. Suppose $k \geq 1$ and focusing on the first element of $\underline{X}_{t+k}^{t+k-p+1}$ in (13) and factoring out Z_{t+k} gives

$$X_{t+k} = Z_{t+k}\big\{a_0(t+k) + [A_{t+k}\sum_{r=0}^{k-2}\prod_{i=1}^{r} A_{t+k-i}(Z) b_{t+k-r-1}(Z)]_1 + [A_{t+k}\{\prod_{i=1}^{k-1} A_{t+k-i}(Z)\}\underline{X}_t^{t-p+1}]_1\big\},$$

which is (14) (with $s = 0$). To prove (14) for $1 \leq s \leq p$, we notice by using the tvARCH(p) representation in (11) and (14) for $s = 0$ gives

$$\begin{aligned} X_{t+k+s} &= Z_{t+k+s}\left\{a_0(t+k+s) + \sum_{i=1}^{s-1} a_i(t+k+s)X_{t+k+s-i} + \sum_{i=s}^{p} a_i(t+k+s)X_{t+k+s-i}\right\} \\ &= Z_{t+k+s}\left\{\mathcal{P}_{s,k,t}(\underline{Z}) + \mathcal{Q}_{s,k,t}(\underline{Z},\underline{X})\right\}, \end{aligned}$$

where $\mathcal{P}_{s,k,t}$ and $\mathcal{Q}_{s,k,t}$ are defined in (15). Hence this gives (14). Since $a_j(\cdot)$ and Z_t are positive, it is clear that $\mathcal{P}_{s,k,t}$ and $\mathcal{Q}_{s,k,t}$ are positive random variables. $\square$

We require the following simple lemma to prove Lemmas 3.3 and 4.4.

Lemma A.4 *Suppose that Assumption 3.1(iii) is satisfied, then for any positive A and B we have*

$$\int_{\mathbb{R}} \left|\frac{1}{A+B} f_Z(\frac{y}{A+B}) - \frac{1}{A} f_Z(\frac{y}{A})\right| dy \leq K(\frac{B}{A} + \frac{B}{A+B}). \tag{58}$$

Suppose that Assumption 3.1(iv) is satisfied, then for any positive A, positive continuous function $B : \mathbb{R}^{r_2+1} \to \mathbb{R}$ and set E (defined as in (4)) we have

$$\int_{\mathbb{R}} \sup_{\underline{x}\in\mathcal{E}} \left|\frac{1}{A+B(\underline{x})} f_Z(\frac{y}{A+B(\underline{x})}) - \frac{1}{A} f_Z(\frac{y}{A})\right| dy \leq K \sup_{\underline{x}\in\mathcal{E}}(\frac{B(\underline{x})}{A} + \frac{B(\underline{x})}{A+B(\underline{x})}). \tag{59}$$

PROOF. To prove (58) we observe that

$$\int_{\mathbb{R}} \left|\frac{1}{A+B} f_Z(\frac{y}{A+B}) - \frac{1}{A} f_Z(\frac{y}{A})\right| dy = I + II$$

where $I = \int_{\mathbb{R}} \frac{1}{A+B}|f_Z(\frac{y}{A+B}) - f_Z(\frac{y}{A})|dy$ and $II = \int_{\mathbb{R}} (\frac{1}{A+B} - \frac{1}{A})f_Z(\frac{y}{A})$.

To bound I, we note that by changing variables with $u = y/(A+B)$ and under Assumption 3.1(iii) we get

$$I \leq \int_{\mathbb{R}} \big|f_Z(u) - f_Z\big(u(1+\frac{B}{A})\big)\big|du \leq K\frac{B}{A}.$$

It is straightforward to show $II \leq \frac{B}{A+B}$. Hence the bounds for I and II give (58).

The proof of (59) is the same as above, but uses Assumption 3.1(iv) instead of Assumption 3.1(iii), we omit the details. □

PROOF of Lemma 3.3. We first show that

$$\sup_{\underline{y}_{s-1}\in\mathbb{R}^s} \int \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})dy_s \leq \frac{K}{\inf_{t\in\mathbb{Z}} a_0(t)}\mathcal{Q}_{s,k,t}(\underline{z}, \underline{x}) \tag{60}$$

and use this to prove (19). We note that when $\underline{x} = 0$, $\mathcal{Q}_{s,k,t}(\underline{z}, 0) = 0$ and $f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, 0) = \mathcal{P}_{s,k,t}(\underline{z})^{-1}f_Z(\frac{y_s}{\mathcal{P}_{t+k+s,t+k}(\underline{z})})$. Therefore using (16) gives

$$\mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x}) = \left|\frac{1}{\mathcal{P}_{s,k,t}(\underline{z}) + \mathcal{Q}_{s,k,t}(\underline{z}, \underline{x})}f_Z\big(\frac{y_s}{\mathcal{P}_{s,k,t}(\underline{z}) + \mathcal{Q}_{s,k,t}(\underline{z}, \underline{x})}\big) - \frac{1}{\mathcal{P}_{s,t,k}(\underline{x})}f_Z\big(\frac{y_s}{\mathcal{P}_{s,k,t}(\underline{z})}\big)\right|.$$

Now recalling that $\mathcal{P}_{s,k,t}$ and $\mathcal{Q}_{s,k,t}$ are both positive and setting $A = \mathcal{P}_{s,k,t}(\underline{z})$ and $B = \mathcal{Q}_{s,k,t}(\underline{z}, \underline{x})$ and using (58) we have

$$\int_{\mathbb{R}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})dy_s \leq K\left(\frac{\mathcal{Q}_{s,k,t}(\underline{z}, \underline{x})}{\mathcal{P}_{s,k,t}(\underline{z})} + \frac{\mathcal{Q}_{s,k,t}(\underline{z}, \underline{x})}{\mathcal{P}_{s,k,t}(\underline{z}) + \mathcal{Q}_{s,k,t}(\underline{z}, \underline{x})}\right).$$

Finally, since $\mathcal{P}_{s,k,t}(\underline{z}) > \inf_{t\in\mathbb{Z}} a_0(t)$ we have $\int_{\mathbb{R}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})dy_s \leq K\frac{\mathcal{Q}_{s,k,t}(\underline{z},\underline{x})}{\inf_{t\in\mathbb{Z}} a_0(t)}$, thus giving (60). By using (60) we now prove (19). Substituting (60) into the integral on the left hand side of (19) and using that $\mathbb{E}[\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})] = \mathcal{Q}_{s,k,t}(\underline{1}_{k-1}, \underline{x})$, using this and substituting (60) into (17) gives

$$\int\prod_{i=1}^{k-1} f_Z(z_i) \sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\{\int_{\mathbb{R}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})dy_s\}d\underline{z} \leq K\frac{\mathbb{E}[\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})]}{\inf_{t\in\mathbb{Z}} a_0(t)} = K\frac{\mathcal{Q}_{s,k,t}(\underline{1}_{k-1}, \underline{x})}{\inf_{t\in\mathbb{Z}} a_0(t)}. \tag{61}$$

We now find a bound for $\mathcal{Q}_{s,k,t}$. By definition of $\mathcal{Q}_{s,k,t}$ in (15) and using the matrix norm inequality $[A\underline{x}]_1 \leq K\|A\|_{spec}\|\underline{x}\|$ ($\|\cdot\|_{spec}$ is the spectral norm) we have

$$\begin{aligned}
\mathcal{Q}_{s,k,t}(\underline{1}_{k-1}, \underline{x}) &= \sum_{i=s+1}^{p} a_i(t+k+s)\big[A_{t+k+s-i}\sum_{r=1}^{k+s-i}\{\prod_{d=0}^{k+s-i} A_{t+k+s-i-d}\}\underline{x}\big]_1 \\
&\leq \frac{K}{\inf_{t\in\mathbb{Z}} a_0(t)}\sum_{i=s}^{p} a_i(t+k+s)\big\|A_{t+k+s-i}\{\prod_{d=0}^{k-1} A_{t+k+s-i-d}\}\big\|_{spec}\|\underline{x}\|.
\end{aligned}$$

To bound the above, we note that by Assumption 3.1(i), $\sup_{t\in\mathbb{Z}}\sum_{j=1}^{p} a_j(t) \leq (1-\delta)$, therefore there exists a $\tilde{\delta}$, where $0 < \tilde{\delta} < \delta < 1$, such that for all t we have $\|A_{t+k+s-i}\{\prod_{d=0}^{k-1} A_{t+k+s-i-d}\}\|_{spec} \leq K(1-\tilde{\delta})^{k+1}$, for some finite K. Altogether this gives

$$\begin{aligned}
\mathcal{Q}_{s,k,t}(\underline{1}_{k-1}, \underline{x}) &\leq \frac{K}{\inf_{t\in\mathbb{Z}} a_0(t)}\sum_{i=s}^{p} a_i(t+k+s)\big\|A_{t+k+s-i}\{\prod_{d=0}^{k+s-i} A_{t+k+s-i-d}\}\big\|_{spec}\|\underline{x}\| \\
&\leq \frac{K}{\inf_{t\in\mathbb{Z}} a_0(t)}\sum_{i=s}^{p} a_i(t+k+s)(1-\tilde{\delta})^{k+s-i}\|\underline{x}\|. \qquad (62)
\end{aligned}$$

Substituting the above into (61) gives (19).

We now prove (20). We use the same proof to show (60), but apply (58) instead of (59) to obtain

$$\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\int \sup_{\underline{x}\in\mathcal{E}}\mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})dy_s \leq \frac{K}{\inf_{t\in\mathbb{Z}} a_0(t)}\sup_{\underline{x}\in\mathcal{E}}\mathcal{Q}_{s,k,t}(\underline{z}, \underline{x}).$$

By substituting the above into (18) and using the same proof to prove (19) we obtain

$$\sum_{s=0}^{p-1}\int\prod_{i=1}^{k-1} f_Z(z_i)\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\{\int_{\mathbb{R}}\sup_{\underline{x}\in\mathcal{E}}\mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})dy_s\}d\underline{z} \leq K\frac{\mathbb{E}[\sup_{\underline{x}\in\mathcal{E}}\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})]}{\inf_{t\in\mathbb{Z}} a_0(t)}. \qquad (63)$$

Since $\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})$ is a positive function and $\sup_{\underline{x}\in\mathcal{E}}\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x}) = \mathcal{Q}_{s,k,t}(\underline{Z}, \underline{\eta})$ we have $\mathbb{E}[\sup_{\underline{x}\in\mathcal{E}}\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})]$ $\sup_{\underline{x}\in\mathcal{E}}\mathbb{E}[\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})] = \sup_{\underline{x}\in\mathcal{E}}\mathcal{Q}_{s,k,t}(\underline{1}_{k-1}, \underline{x})$, hence by using (62) we have

$$\frac{\mathbb{E}[\sup_{\underline{x}\in\mathcal{E}}\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})]}{\inf_{t\in\mathbb{Z}} a_0(t)} \leq \frac{K(1-\tilde{\delta})^k\|\underline{x}\|}{\inf_{t\in\mathbb{Z}} a_0(t)}.$$

Substituting the above into (63) gives (20). □

References

K. B. Athreya and S. G. Pantula. Mixing properties of Harris chains and autoregressive processes. *J. Appl. Probab.*, 23:880–892, 1986.

B. Basrak, R.A. Davis, and T Mikosch. Regular variation of GARCH processes. *Stochastic Processes and their Applications*, 99:95–115, 2002.

I. Berkes, S. Hörmann, and J. Schauer. Asymptotic results for the empirical process of stationary sequences. *Stochastic Processes and their Applications*, 00:000–000, 2008. Forthcoming.

D. Bosq. *Nonparametric Statistics for Stochastic Processes.* Springer, New York, 1998.

P. Bougerol and N. Picard. Stationarity of GARCH processes and some nonnegative time series. *J. Econometrics*, 52:115–127, 1992.

F. Bousamma. *Ergodicité, mélange et estimation dans les modèles GARCH.* PhD thesis, Paris 7, 1998.

R. C. Bradley. *Introduction to Strong Mixing Conditions Volumes 1,2 and 3.* Kendrick Press, 2007.

M. Carrasco and X. Chen. Mixing and moment properties of various garch and stochastic volatility models. *Econometric Theory*, 18:17–39, 2002.

K. C. Chanda. Strong mixing properties of linear stochastic processes. *J. Appl. Prob.*, 11: 401–408, 1974.

R. Dahlhaus and W. Polonik. Nonparametric quasi-maximum likelihood estimation for gaussian locally stationary processes. *Ann. Statist.*, 34:2790–2842, 2006.

R. Dahlhaus and S. Subba Rao. Statistical inference of time varying ARCH processes. *Ann. Statistics*, 34:1074–1114, 2006.

J Davidson. *Stochastic Limit Theory.* Oxford University Press, Oxford, 1994.

P.D. Feigin and R. L. Tweedie. Random coefficient autoregressive processes. a markov chain analysis of stationarity and finiteness of moments. *Journal of Time Series Analysis*, 6:1–14, 1985.

C. Francq and J.-M. Zakoïan. Mixing properties of a general class of garch$(1,1)$ models without moment assumptions. *Econometric Theory*, 22:815–834, 2006.

P. Fryzlewicz, T. Sapatinas, and S. Subba Rao. Normalised least squares estimation in time-varying ARCH models. *Ann. Statistics*, 36:742–786, 2008.

P. Fryzlewicz and S. Subba Rao. BaSTA: consistent multiscale multiple change-point detection for piecewise-stationary ARCH processes. 2008.

L. Giraitis, P. Kokoskza, and R. Leipus. Stationary ARCH models: Dependence structure and central limit theorem. *Econometric Theory*, 16:3–22, 2000.

L. Giraitis, R. Leipus, and D Surgailis. Recent advances in ARCH modelling. In A. Kirman and G. Teyssiere, editors, *Long Memory in Economics*, pages 3–39. Berlin, 2005.

L. Giraitis and P.M. Robinson. Whittle estimation of ARCH models. *Econometric Theory*, 17: 608–631, 2001.

V.V. Gorodetskii. On the strong mixing propery for linear sequences. *Theory of Probability and its Applications*, 22:411–413, 1977.

P Hall and C.C. Heyde. *Martingale Limit Theory and its Application*. Academic Press, New York, 1980.

E. J. Hannan and L. Kavaliers. Regression, autoregressive models. *J. Time Series Anal.*, 7: 27–49, 1986.

S Hörmann. Augmented GARCH sequences: Dependence structure and asymptotics. *Bernoulli*, 14:543–561, 2008.

I. A. Ibragimov. Some limit theorems for stationary processes. *Theory of Probability and its Applications*, 7:349–82, 1962.

E. Liebscher. Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *J. Time Series Analysis*, 26:669–689, 2005.

A. Lindner. *Handbook of financial time series*, chapter Stationarity, Mixing, Distributional Properties and Moments of GARCH(p,q)- Processes. Springer Verlag, Berlin, 2008.

M. Meitz and P. Saikkonen. Ergodicity, mixing, and existence of moments of a class of markov models with applications to garch and acd models. *Econometric Theory*, 24:1291–1320, 2008.

S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability.* Springer-Verlag, 1993.

T. Mikosch and C. Stărică. Long-range dependence effects and arch modelling. In P. Doukhan, G. Oppenheim, and M.S. Taqqu, editors, *Theory and Applications of Long Range Dependence*, pages 439–459. Birkhäuser, Boston, 2003.

A. Mokkadem. Propertiés de mélange des processus autoregressifs polnomiaux. *Annales de l'I H. P.*, 26:219–260, 1990.

D. T. Pham. The mixing propery of bilinear and generalised random coefficient autorregressive models. *Stochastic Processes and their Applications*, 23:291–300, 1986.

D. T. Pham and T .T Tran. Some mixing properties of time series models. *Stochastic processes and their applications*, 19:297–303, 1985.

P. M. Robinson and P. Zaffaroni. Pseudo-maximum likelihood estimation of ARCH(∞) models. *Ann. Statist.*, 34:1049–1074, 2006.

P.M. Robinson. Testing for strong serial correlation and dynamic conditional heteroskedasity in multiple regression. *Journal of Econometrics*, 47:67–78, 1991.

A. A. Sorokin. Uniform bound for strong mixing coefficient and maximum of residual empirical process of ARCH sequences (in Russian). *ArXiv (0610747)*, 2006.

D. Straumann and T. Mikosch. Quasi-maximum likelihood estimation in conditionally hetroscedastic time series: a stochastic recurrence equation approach. *Ann. Statist.*, pages 2449–2495, 2006.

S. Subba Rao. A note on uniform convergence of an ARCH(∞) estimator. *Sankhya*, 68:600–620, 2006.

D. Tjostheim. Nonlinear time series and markov chains. *Adv. in Appl. Probab*, 22:587–611, 1990.

V. A. Volkonskii and Yu. A. Rozanov. Some limit theorems for random functions I. *Theor. Probabl Appl.*, 4:178–197, 1959.