Bernoulli **17**(1), 2011, 320–346
DOI: 10.3150/10-BEJ270

Mixing properties of ARCH and time-varying ARCH processes

PIOTR FRYZLEWICZ[1] and SUHASINI SUBBA RAO[2]

[1]*Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail: p.fryzlewicz@lse.ac.uk*
[2]*Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA. E-mail: suhasini@stat.tamu.edu*

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

Keywords: 2-mixing; absolutely regular (β-mixing) ARCH(∞); conditional densities; strong mixing (α-mixing); time-varying ARCH

1. Introduction

Mixing is a measure of dependence between elements of a random sequence that has a wide range of theoretical applications (see [7] and below). One of the most popular mixing measures is α-mixing (also called *strong mixing*), where the α-mixing rate of the non-stationary stochastic process $\{X_t\}$ is defined as a sequence of coefficients $\alpha(k)$ such that

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

$\{X_t\}$ 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, among other results, it is possible to show asymptotic normality of sums of $\{X_k\}$ (see [12], Chapter 24), as well as exponential inequalities for such sums (see [4]), asymptotic normality of kernel-based nonparametric estimators (see [4]) and consistency of change point detection schemes of nonlinear time series (see [16]). 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 [4]). 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 [24]), which is usually not possible

1350-7265

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 $\tilde{\alpha}(k) \le \alpha(k)$. A closely related mixing measure, introduced in [39] is β-mixing (also called *absolutely regular mixing*). 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 $\alpha(k)$ due to the fact that $\beta(k)\ge\alpha(k)$).

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 [9], Gorodetskii [20], Athreya and Pantula [1] and Pham and Tran [32] show strong mixing of the MA(∞) process. Feigin and Tweedie [13] and Pham [31] 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, e.g., [14]). More recently, Tjostheim [38] and Mokkadem [30] have shown geometric ergodicity for a general class of Markovian processes. The results in [30] have been applied in [6] 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 [8,25,26,35] (for an excellent review) and [14,27] (where mixing of 'nonlinear' GARCH(p,q) processes is also considered). Most of these these results are proved by verifying the Meyn–Tweedie conditions (see [13] and [28]) and, as mentioned above, are derived under the premise that the process is stationary (or asymptotically stationary) and Markovian. Clearly, if a process is non-stationary, 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 [18] for a comprehensive overview of mathematical properties of ARCH processes and a list of further references. It is worth mentioning that Hörmann [23] and Berkes *et al.* [3] 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 adequately explain the long memory seen in the data or change according to shifts in the world economy. Therefore, attention has recently been paid to tvARCH models (see, e.g., [11,15,16,29]) and ARCH(∞) models (see [17,19,33,37]). The derivations of the sampling properties of some of the aforementioned papers rely on quite sophisticated assumptions on the dependence structure, in particular, on 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 bounds have 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 presented 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, available at http://stats.lse.ac.uk/fryzlewicz/mixing/tvARCH_mixing.pdf.

2. Some mixing inequalities for general processes

2.1. Notation

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

2.2. Some mixing inequalities

Let us suppose that $\{X_t\}$ 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

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

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

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 [9] and [20] (see also [12], 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*

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

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

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

and

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

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

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

and

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

where $\mathbb{E}_{\underline{W}}(g(\underline{W})) = \int g(\underline{w}) f_{\underline{W}}(\underline{w})\,\mathrm{d}\underline{w}$, $f_{\underline{W}}(\underline{w})$ *is the density of* $\underline{w}$, $\mathcal{D}_{0,k,t}(y_0|\underline{y}_{-1}, \underline{w}, \underline{x}) = |f_{s,k,t}(y_s|\underline{w}, \underline{x}) - f_{s,k,t}(y_s|\underline{w}, 0)|$ *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* $f_{0,k,t}$, *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* $f_{s,k,t}$, $\underline{x} = (x_0, \ldots, x_{-r_2})$ *and* $\underline{w} = (w_k, \ldots, w_1)$.

Proof. This can be found in Appendix A.1. □

Since the above bounds hold for all vectors $\underline{\eta} \in (\mathbb{R}^+)^{r_1+1}$ (note that $\underline{\eta}$ 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 (e.g., linear and ARCH-type processes). If $\underline{W}_{t+k-1}^{t+1}$ is the innovation process, it can often 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 obtain bounds for the mixing rates.

The proof of Proposition 2.1 can be found in the Appendix, but we give a brief outline of it here. Let

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

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 $\mathcal{E}$ (i.e., 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

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

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 [10]). Using (7) and techniques similar 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 challenging 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 [15], it is shown that the tvARCH process can be used to explain the commonly observed stylized facts in financial time series (such as the empirical long memory). A sequence of random variables $\{X_t\}$ is said to come from the squares of a time-varying ARCH(p) process 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 (i.i.d.) positive random variables, where $\mathbb{E}(Z_t) = 1$ and $a_j(\cdot)$ are positive parameters. It is worth comparing (11) with the squared tvARCH process used in the statistical literature. Unlike the squared tvARCH process considered in, for example, [11] and [15], we have not placed *any* smoothness conditions on the time-varying parameters $\{a_j(\cdot)\}$. The smoothness conditions assumed in [11] and [15] are used in order to carry out 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. From now on, with a slight abuse of terminology, we will call the squared tvARCH process simply the tvARCH process.

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

(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* $\int |f_Z(u) - f_Z(u[1+a])|\,\mathrm{d}u \le Ka$ *for some finite* K *independent of* a.

(iv) *Let* f_Z *denote the density of* Z_t. *For all* $a > 0$, *we have* $\int \sup_{0\le\tau\le a} |f_Z(u) - f_Z(u[1+\tau])|\,\mathrm{d}u \le Ka$ *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 [11], 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 that f'_Z is bounded, that after some finite point m the derivative f' declines monotonically to zero and satisfies $\int |yf'_Z(y)|\,\mathrm{d}y < \infty$. In this case,

$$\begin{aligned}
&\int_0^\infty \sup_{0\le\tau\le a} |f_Z(u) - f_Z(u[1+\tau])|\,\mathrm{d}u \\
&\quad\le \int_0^m \sup_{0\le\tau\le a} |f_Z(u) - f_Z(u[1+\tau])|\,\mathrm{d}u + \int_m^\infty \sup_{0\le\tau\le a} |f_Z(u) - f_Z(u[1+\tau])|\,\mathrm{d}u \\
&\quad\le a\left(m^2 \sup_{u\in\mathbb{R}} |f'_Z(u)| + \int_m^\infty u|f'_Z(u)|\,\mathrm{d}u\right) \le Ka
\end{aligned}$$

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

We use Assumption 3.1(i)–(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) and (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, [14] 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 [17]). 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

$$A_t = \begin{pmatrix} a_1(t)Z_t & a_2(t)Z_t & \dots & a_p(t)Z_t \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}, \qquad \tilde{A}_t = \begin{pmatrix} a_1(t) & a_2(t) & \dots & a_p(t) \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

$$\underline{b}_t = (a_0(t)Z_t, 0, \dots, 0)' \quad \text{and} \quad \underline{X}_t^{t-p+1} = (X_t, X_{t-1}, \dots, X_{t-p+1})'.$$

Using this notation, we have the relation $\underline{X}_{t+k}^{t+k-p+1} = A_{t+k}\underline{X}_{t+k-1}^{t+k-p} + \underline{b}_{t+k}$. We note that the vector representation of ARCH and GARCH processes has been used in [2,5,36] 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

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

where we set $[\prod_{i=0}^{-1} A_{t+k-i}] = I_p$ (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. For $s \geq 0$, we then have*

$$X_{t+k+s} = Z_{t+k+s}\{\mathcal{P}_{s,k,t}(\underline{Z}) + \mathcal{Q}_{s,k,t}(\underline{X})\}, \tag{12}$$

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

For $1 \le s \le p$,

$$\begin{aligned}\mathcal{P}_{s,k,t}(\underline{Z}) = a_0(t+k+s) &+ \sum_{i=1}^{s-1} a_i(t+k+s) X_{t+k+s-i} \\ &+ \sum_{i=s}^{p} a_i(t+k+s) Z_{k+s-i} \\ &\quad \times \Biggl\{ a_0(t+k+s-i) \\ &\quad + \Biggl[\tilde{A}_{t+k+s-i} \sum_{r=1}^{k+s-i} \Biggl\{ \prod_{d=0}^{r} A_{t+k+s-i-d} \Biggr\} b_{t+k+s-i-r} \Biggr]_1 \Biggr\}, \end{aligned} \tag{13}$$

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

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* $\mathcal{P}_{s,k,t}(\cdot)$ *and* $\mathcal{Q}_{s,k,t}(\cdot)$ *are positive random variables and for* $s \ge 1$, $\mathcal{P}_{s,k,t}(\cdot)$ *is a function of* $\underline{X}_{t+k+s-1}^{t+k}$ (*but this has been suppressed in the notation*).

Proof. This is found in Appendix A.2. □

By using (12), 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 (12) 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

$$f_{s,k,t}(y_s | \underline{y}_{s-1}, \underline{z}, \underline{x}) = \frac{1}{\mathcal{P}_{s,k,t}(\underline{z}) + \mathcal{Q}_{s,k,t}(\underline{z}, \underline{x})} f_Z\left(\frac{y_s}{\mathcal{P}_{s,k,t}(\underline{z}) + \mathcal{Q}_{s,k,t}(\underline{z}, \underline{x})} \right). \tag{14}$$

3.3. Strong mixing of the tvARCH(p) process

The aim of 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 that $\{X_t\}$ is a tvARCH process which satisfies (11). For any $\underline{\eta} = (\eta_0, \ldots, \eta_{-p+1}) \in (\mathbb{R}^+)^p$, we then have*

$$
\begin{aligned}
&\sup_{G \in \mathcal{F}_\infty^{t+k}, H \in \mathcal{F}_t^{-\infty}} |P(G \cap H) - P(G)P(H)| \\
&\quad \le 2 \sum_{s=0}^{p-1} \sup_{\underline{x} \in \mathcal{E}} \int \mathbb{E}_{\underline{Z}} \biggl(\sup_{\underline{y}_{s-1} \in \mathbb{R}^s} \int_{\mathbb{R}} \mathcal{D}_{s,k,t}(y_s | \underline{y}_{s-1}, \underline{Z}, \underline{x}) \, \mathrm{d}y_s \biggr) \\
&\qquad + 4 \sum_{j=0}^{p-1} P(|X_{t-j}| \ge \eta_{-j+1})
\end{aligned} \tag{15}
$$

and

$$
\begin{aligned}
&\sup_{\{H_j\} \in \mathcal{F}_t^{-\infty}, \{G_j\} \in \underline{F}_\infty^{t+k}} \sum_{i,j} |P(G_i \cap H_j) - P(G_i)P(H_j)| \\
&\quad \le 2 \sum_{s=0}^{p-1} \sup_{\underline{x} \in \mathcal{E}} \mathbb{E}_{\underline{Z}} \biggl(\sup_{\underline{y}_{s-1} \in \mathbb{R}^s} \int_{\mathbb{R}} \sup_{\underline{x} \in \mathcal{E}} \mathcal{D}_{s,k,t}(y_s | \underline{y}_{s-1}, \underline{Z}, \underline{x}) \, \mathrm{d}y_s \biggr) \\
&\qquad + 4 \sum_{j=0}^{p-1} P(|X_{t-j}| \ge \eta_{-j+1}),
\end{aligned} \tag{16}
$$

where $\underline{z} = (z_1, \ldots, z_{k-1})$ and $\{G_i\}$ and $\{H_j\}$ are partitions of Ω and $\mathbb{E}_{\underline{Z}}(g(\underline{Z})) = \int g(\underline{z}) \times \prod_{i=1}^{k-1} f_Z(z_i) \, \mathrm{d}z_i$.

Proof. This can be found in Appendix A.2. □

To obtain a mixing rate for the tvARCH(p) process, we need to bound the integral in (15), then obtain the set E which minimizes (15). 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 (14)).

Lemma 3.3. *Let $\mathcal{D}_{s,k,t}$ and $\mathcal{Q}_{s,k,t}$ be defined as in (9) and (13), respectively.*

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

$$
\begin{aligned}
\sum_{s=0}^{p-1} \int \mathbb{E}_{\underline{Z}} \biggl(\sup_{\underline{y}_{s-1} \in \mathbb{R}^s} \int \mathcal{D}_{s,k,t}(y_s | \underline{y}_{s-1}, \underline{Z}, \underline{x}) \, \mathrm{d}y_s \biggr) &\le K \frac{\mathbb{E}[\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})]}{\inf_{t \in \mathbb{Z}} a_0(t)} \\
&\le K (1 - \tilde{\delta})^k \|\underline{x}\|,
\end{aligned} \tag{17}
$$

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

(ii) *Supposing that Assumption* 3.1(i)–(ii) *and* (vi) *hold*, *then for any set* $\mathcal{E}$ (*defined as in* (4)), *we have*

$$\sum_{s=0}^{p-1} \mathbb{E}_{\underline{Z}}\left(\sup_{\underline{y}_{s-1}\in\mathbb{R}^s} \int \sup_{\underline{x}\in\mathcal{E}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{Z}, \underline{x})\,\mathrm{d}y_s \right) \le \sup_{\underline{x}\in\mathcal{E}} K(1-\tilde{\delta})^k \|\underline{x}\|. \tag{18}$$

Proof. This can be found in Appendix A.2. □

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

Theorem 3.1. (i) *Supposing that Assumption* 3.1(i)–(iii) *holds*, *then*

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

(ii) *Supposing that Assumption* 3.1(i)–(ii) *and* (iv) *hold*, *then*

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

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

Proof. We will use (15) to prove (i). Equation (17) gives a bound for the integral difference in (15); therefore, all that remains is to bound the probabilities in (15). To do this, we first use Markov's inequality, to give $\sum_{j=0}^{p-1} P(|X_{t-j}| \ge \eta_{-j}) \le \sum_{j=0}^{p-1} \mathbb{E}|X_{t-j}|\eta_{-j}^{-1}$. By using the Volterra expansion of X_t (see [11], Section 5), it can be shown that $\sup_{t\in\mathbb{Z}} \mathbb{E}|X_t| \le (\sup_{t\in\mathbb{Z}} a_0(t))/(1 - \sup_{t\in\mathbb{Z}} \sum_{j=1}^{p} a_j(t))$. Using these bounds and substituting (17) into (15) gives, for every $\underline{\eta} \in (\mathbb{R}^+)^p$, the bound

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

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

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

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 using the bound in (18) instead of (17). We omit the details. □

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

Moreover, since $\sigma(X_{t+k}) \subset \sigma(X_{t+k}, \dots, X_{t+p-1})$ and $\sigma(X_t) \subset \sigma(X_t, \dots, X_{t-p+1})$, 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 [17]). The squares of the ARCH(∞) process satisfy the representation

$$X_t = Z_t\left(a_0 + \sum_{j=1}^{\infty} a_j X_{t-j}\right), \tag{19}$$

where Z_t are i.i.d. positive random variables with $\mathbb{E}(Z_t) = 1$ and a_j are positive parameters. With a slight abuse of terminology, we will call the squared ARCH(∞) process an ARCH(∞) process. It is worth mentioning that the GARCH(p, q) process has an ARCH(∞) representation, where the a_j decay geometrically with j. Giraitis and Robinson [19], Robinson and Zaffaroni [34] and Subba Rao [37] 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.* [17] have shown that under Assumption 4.1(i), the ARCH(∞) process has a stationary solution and a finite mean (i.e., $\sup_{t\in\mathbb{Z}} \mathbb{E}(X_t) < \infty$). It is worth mentioning that since the ARCH(∞) process has a stationary solution, the shift t plays no role when obtaining mixing bounds, that is, $\sup_{G\in\sigma(X_{k+t}), H\in\sigma(X_t)} |P(G \cap H) - P(G)P(H)| = \sup_{G\in\sigma(X_k), H\in\sigma(X_0)} |P(G \cap H) - P(G)P(H)|$. 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 (20)) that, in general, the ARCH(∞) process is not Markovian.

Lemma 4.1. *Suppose that* $\{X_t\}$ *satisfies* (19). *Then*

$$X_k = \mathcal{P}_{0,k}(\underline{Z})Z_k + \mathcal{Q}_{0,k}(\underline{Z}, \underline{X})Z_k, \tag{20}$$

where

$$\begin{aligned} \mathcal{P}_{0,k}(\underline{Z}) &= \left[a_0 + \sum_{m=1}^{k} \sum_{k=j_m > \cdots > j_1 > 0} \left(\prod_{i=1}^{m-1} a_{j_{i+1}-j_i}\right)\left(\prod_{i=1}^{m-1} Z_{j_i}\right)\right], \\ \mathcal{Q}_{0,k}(\underline{Z}, \underline{X}) &= \sum_{r=1}^{k} \left\{\sum_{m=1}^{k} \sum_{k=j_m > \cdots > j_1 = r} \left(\prod_{i=1}^{m-1} a_{j_{i+1}-j_i}\right)\left(\prod_{i=1}^{m-1} Z_{j_i}\right)\right\} d_r(\underline{X}). \end{aligned} \tag{21}$$

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. This can be found in Appendix A.3 of the technical report. □

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 that* $\{X_t\}$ *satisfies* (19). *For* $s \geq 1$, *we then have*

$$\begin{aligned} X_{k+s} &= Z_{k+s}\{\mathcal{P}_{s,k}(\underline{Z}) + \mathcal{Q}_{s,k}(\underline{Z}, \underline{X})\}, \\ &\text{where } \mathcal{P}_{s,k}(\underline{Z}) = a_0 + \sum_{j=1}^{s} a_j X_{k+s-j} + \sum_{j=s+1}^{\infty} a_j Z_{k+s-j}\mathcal{P}_{0,k+s-j}(\underline{Z}), \\ \mathcal{Q}_{s,k}(\underline{Z}, \underline{X}) &= \sum_{j=s+1}^{k+s} a_j Z_{k+s-j}\mathcal{Q}_{0,k+s-j}(\underline{Z}, \underline{X}) + d_{k+s}(\underline{X}). \end{aligned} \tag{22}$$

Proof. This can be found in Appendix A.3 of the technical report. □

Using (20) and (22), 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

$$f_{s,k}(y_s | \underline{y}_{s-1}, \underline{z}, \underline{x}) = \frac{1}{\mathcal{P}_{s,k}(\underline{z}) + \mathcal{Q}_{s,k}(\underline{z}, \underline{x})} f_Z\left(\frac{y_s}{\mathcal{P}_{s,k}(\underline{z}) + \mathcal{Q}_{s,k}(\underline{z}, \underline{x})}\right). \tag{23}$$

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 (23), 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 that $\{X_t\}$ satisfies (19) and Assumption 4.1 is fulfilled. There then exists $\{\psi_j\}$ such that for all $|z| \le 1$, 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 [22] have shown that $\sum_j |j^\alpha \psi_j| < \infty$. For $k \le 0$, set $d_k(\underline{x}) = 0$ and $\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x}) = 0$. For $k \ge 1$, $\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x})$ then has the solution*

$$\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x}) = \sum_{j=0}^{\infty} \psi_j d_{k-j}(\underline{x}) = \sum_{j=0}^{k-1} \psi_j d_{k-j}(\underline{x}) = \sum_{j=0}^{k-1} \psi_j \left\{ \sum_{i=0}^{\infty} a_{k-j+i} x_{-i} \right\}, \tag{24}$$

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

Proof. This appears in Appendix A.3 of the technical report. □

4.3. Mixing for ARCH(∞) 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 (19). Giraitis *et al.* [17] 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 $|\mathrm{cov}(X_0, X_k)| \sim k^{-(1+\delta)}$. 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 [21]), we have

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

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

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 that $\{X_t\}$ satisfies (19) and let $\mathcal{D}_{s,k}$ and $\mathcal{Q}_{0,k}(\cdot)$ be defined as in (9) and (21), respectively. If Assumptions 3.1*(iii) *and 4.1 are fulfilled, then*

$$\begin{aligned} &\mathbb{E}_{\underline{Z}}\left(\int |f_{0,k}(y|\underline{Z}, \underline{x}) - f_{0,k}(y|\underline{Z}, 0)| \, \mathrm{d}y \right) \\ &\quad \le \frac{\mathcal{Q}_{0,k}(\underline{1}_{k-1}, \underline{x})}{a_0} = \sum_{j=0}^{k-1} |\psi_j| \left\{ \sum_{i=0}^{\infty} a_{k-j+i} x_{-i} \right\} \end{aligned} \tag{25}$$

and, for $s \geq 1$,

$$\mathbb{E}_{\underline{Z}}\left(\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\int \mathcal{D}_{s,k}(y_s|\underline{y}_{s-1},\underline{Z},\underline{x})\,\mathrm{d}y_s\right) \leq \frac{1}{a_0}\left\{\sum_{j=s+1}^{k+s} a_j \sum_{l=0}^{k+s-j}|\psi_l|\sum_{i=0}^{\infty}a_{k+s-j-l+i}x_{-i} + \sum_{i=0}^{\infty}a_{k+s+i}x_{-i}\right\}. \tag{26}$$

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

$$\mathbb{E}_{\underline{Z}}\left(\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\int \sup_{\underline{x}\in\mathcal{E}}\mathcal{D}_{s,k}(y_s|\underline{y}_{s-1},\underline{Z},\underline{x})\,\mathrm{d}y_s\right) \leq \frac{1}{a_0}\left\{\sum_{j=s+1}^{k+s} a_j \sum_{l=0}^{k+s-j}|\psi_l|\sum_{i=0}^{\infty}a_{k+s-j-l+i}\eta_{-i} + \sum_{i=0}^{\infty}a_{k+s+i}\eta_{-i}\right\}, \tag{27}$$

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

Proof. This can be found in Appendix A.3 of the technical report. □

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

Lemma 4.5. *If* $\{c_i\}$, $\{d_i\}$ *and* $\{\eta_{-i}\}$ *are positive sequences, then*

$$\inf_{\underline{\eta}}\left\{\sum_{i=0}^{\infty}(c_i\eta_{-i} + d_i\eta_{-i}^{-\nu})\right\} = \left(\nu^{1/(1+\nu)} + \nu^{-\nu/(\nu+1)}\right)\sum_{i=0}^{\infty}c_i^{\nu/(\nu+1)}d_i^{1/(\nu+1)}. \tag{28}$$

Proof. This appears in Appendix A.3 of the technical report. □

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

Theorem 4.1. *Suppose that* $\{X_t\}$ *satisfies* (19).

(a) *Suppose Assumptions* 3.1(iii) *and* 4.1 *hold. We then have*

$$\sup_{G\in\mathcal{F}_{\infty}^{k},H\in\mathcal{F}_0^{-\infty}}|P(G\cap H) - P(G)P(H)| \leq K(\nu)\sum_{i=0}^{\infty}\left[\frac{1}{a_0}\sum_{s=0}^{\infty}\sum_{j=s+1}^{k+s}a_j\sum_{l=0}^{k+s-j}|\psi_l|a_{k+s-j-l+i} + \frac{1}{a_0}\sum_{s=0}^{\infty}a_{k+s+i}\right]^{\nu/(\nu+1)}[\mathbb{E}|X_0|^{\nu}]^{1/(\nu+1)}, \tag{29}$$

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

(i) *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*

$$\sup_{G\in\mathcal{F}_\infty^k, H\in\mathcal{F}_0^{-\infty}} |P(G\cap H) - P(G)P(H)| \le K\cdot[k(k+1)^{-\tilde{\delta}+3} + (k+1)^{-\tilde{\delta}+2}],$$

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*

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

where C *is a finite constant.*

(b) *If Assumptions* 3.1(iv) *and* 4.1 *hold, then we have*

$$\begin{aligned}\sup_{\{G_i\}\in\mathcal{F}_\infty^k, \{H_j\}\in\mathcal{F}_0^{-\infty}} & \sum_i\sum_j |P(G_i\cap H_j) - P(G_i)P(H_j)| \\ &\le K(\nu)\sum_{i=0}^{\infty}\Biggl[\frac{1}{a_0}\sum_{s=0}^{\infty}\sum_{j=s+1}^{k+s} a_j \sum_{l=0}^{k+s-j}|\psi_l| a_{k+s-j-l+i} \\ &\qquad + \frac{1}{a_0}\sum_{s=0}^{\infty} a_{k+s+i}\Biggr]^{\nu/(\nu+1)} [\mathbb{E}|X_0|^\nu]^{1/(\nu+1)},\end{aligned} \tag{30}$$

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 the fact 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

$$\begin{aligned}\sup_{G\in\mathcal{F}_{k+n}^k, H\in\underline{F}_0^{-\infty}} & |P(G\cap H) - P(G)P(H)| \\ &\le 2\sup_{\underline{x}\in\mathcal{E}}\sum_{s=0}^{n}\mathbb{E}_{\underline{Z}}\biggl(\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\biggl\{\int \mathcal{D}_{s,k}(y_s|\underline{y}_{s-1}, \underline{Z}, \underline{x})\,\mathrm{d}y_s\biggr\}\biggr) \\ &\quad + 4P(X_0 > \eta_0 \text{ or}, \ldots, X_{-n} > \eta_{-n}).\end{aligned} \tag{31}$$

To bound the integral in (31), we use (26) to obtain

$$\sup_{\underline{x}\in\mathcal{E}}\sum_{s=0}^{n}\mathbb{E}_{\underline{Z}}\biggl(\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\int_{\mathbb{R}}\mathcal{D}_{s,k}(y_s|\underline{y}_{s-1},\underline{Z},\underline{x})\,\mathrm{d}y_s\biggr)$$
$$=\frac{1}{a_0}\sum_{s=0}^{n}\Biggl\{\sum_{j=s+1}^{k+s}a_j\sum_{l=0}^{k+s-j}|\psi_l|\sum_{i=0}^{\infty}a_{k+s-j-l+i}\eta_{-i}+\sum_{i=0}^{\infty}a_{k+s+i}\eta_{-i}\Biggr\}.$$

Now, by using Markov's inequality, we have that $P(X_0>\eta_0\text{ or},\ldots,X_{-n}\ge\eta_{-n})\le\sum_{i=0}^{n}\frac{\mathbb{E}(|X_i|^\nu)}{\eta_{-i}^\nu}$. Substituting this and the above into (31) and letting $n\to\infty$ gives

$$\sup_{G\in\mathcal{F}_\infty^k,H\in\mathcal{F}_0^{-\infty}}|P(G\cap H)-P(G)P(H)|$$
$$\le\inf_{\eta}\Biggl[\frac{2}{a_0}\sum_{s=0}^{\infty}\Biggl\{\sum_{j=s+1}^{k+s}a_j\sum_{l=0}^{k+s-j}|\psi_l|\sum_{i=0}^{\infty}a_{k+s-j-l+i}\eta_{-i}\tag{32}$$
$$+\sum_{i=0}^{\infty}a_{k+s+i}\eta_{-i}\Biggr\}+4\mathbb{E}|X_0|^\nu\sum_{i=0}^{\infty}\eta_{-i}^{-\nu}\Biggr],$$

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

We now use (28) to minimize (32), which gives us (29). The proof of (i) can be found in the technical report. It is straightforward to prove (ii) using (28).

The proof of (b) is very similar to the proof of (a), but uses (27) rather than (26). We omit the details. □

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

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

Theorem 4.2. *Suppose that* $\{X_t\}$ *satisfies* (19) *and that Assumption* 3.1(iii) *and* 4.1 *hold. We then have*

$$\sup_{G\in\sigma(X_k),H\in\mathcal{F}_0^{-\infty}}|P(G\cap H)-P(G)P(H)|$$
$$\le K(\nu)\sum_{i=0}^{\infty}\Biggl[\frac{1}{a_0}\sum_{j=0}^{k-1}a_j|\psi_j|a_{k-j+i}\Biggr]^{\nu/(\nu+1)}[\mathbb{E}|X_0|^\nu]^{1/(\nu+1)},\tag{33}$$

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

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*

$$\sup_{G\in\sigma(X_k),H\in\mathcal{F}_0^{-\infty}} |P(G\cap H) - P(G)P(H)| \le K\cdot k(k+1)^{-\tilde{\delta}+1}, \tag{34}$$

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

Proof. We use a similar proof to that of Theorem 4.1. The integral difference is replaced with the bound in (25) and we again use Markov's inequality: together they give the bound

$$\begin{aligned}&\sup_{G\in\sigma(X_k),H\in\mathcal{F}_0^{-\infty}} |P(G\cap H) - P(G)P(H)| \\ &\quad\le \inf_{\underline{\eta}}\left[2\frac{1}{a_0}\sum_{j=0}^{k-1}|\psi_j|\left\{\sum_{i=0}^{\infty}a_{k-j+i}\eta_{-i}\right\} + 4\mathbb{E}|X_0|^{\nu}\sum_{i=0}^{\infty}\frac{1}{\eta_{-i}^{\nu}}\right].\end{aligned} \tag{35}$$

Finally, to obtain (33) and (34), we use (35) and a proof similar to that of Theorem 4.1(i). We omit the details. □

Remark 4.2. Comparing (34) 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 σ-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.

Appendix: 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 of Proposition* 2.1. *We then have*

$$\begin{aligned}&|P(G\cap H\cap E) - P(G\cap E)P(H)| \\ &\quad\le 2P(H)\sup_{\underline{x}\in\mathcal{E}}\left|P(G|\underline{X}_t^{t-r_1}=\underline{x}) - P(G|\underline{X}_t^{t-r_1}=0)\right| \\ &\qquad + \inf_{\underline{x}\in\mathcal{E}} P(G|\underline{X}_t^{t-r_1}=\underline{x})\{P(H)P(E^c) + P(H\cap E^c)\}.\end{aligned} \tag{36}$$

Proof. To prove the result, we first observe that

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

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

$$\begin{aligned}
&\inf_{\underline{x}\in\mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_{t}^{t-r_1} = \underline{x}) P(H \cap E) \\
&\quad \le P(G \cap H \cap E) \le \sup_{\underline{x}\in\mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_{t}^{t-r_1} = \underline{x}) P(H)
\end{aligned} \tag{37}$$

and

$$\begin{aligned}
&\inf_{\underline{x}\in\mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_{t}^{t-r_1} = \underline{x}) P(E) \\
&\quad \le P(G \cap E) \le \sup_{\underline{x}\in\mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_{t}^{t-r_1} = \underline{x}) P(E).
\end{aligned} \tag{38}$$

Subtracting (37) from (38) and using $P(H \cap E) = P(H) - P(H \cap E^c)$ gives the inequalities

$$\begin{aligned}
&P(G \cap H \cap E) - P(G \cap E)P(H) \\
&\quad \le \sup_{\underline{x}\in\mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_{t}^{t-r_1} = \underline{x}) P(H) \\
&\qquad - \inf_{\underline{x}\in\mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_{t}^{t-r_1} = \underline{x}) P(H) + P(E^c)P(H),
\end{aligned} \tag{39}$$

$$\begin{aligned}
&P(G \cap H \cap E) - P(G \cap E)P(H) \\
&\quad \ge \inf_{\underline{x}\in\mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_{t}^{t-r_1} = \underline{x}) P(H) \\
&\qquad - \sup_{\underline{x}\in\mathcal{E}} P(\underline{X}_{t+k+r_2}^{t+k} \in \mathcal{G} | \underline{X}_{t}^{t-r_1} = \underline{x}) P(H) - P(E^c \cap H).
\end{aligned} \tag{40}$$

Combining (39) and (40), we obtain

$$\begin{aligned}
&|P(G \cap H \cap E) - P(G \cap E)P(H)| \\
&\quad \le P(H) \Big| \sup_{\underline{x}\in\mathcal{E}} P(G | \underline{X}_{t}^{t-r_1} = \underline{x}) - \inf_{\underline{x}\in\mathcal{E}} P(G | \underline{X}_{t}^{t-r_1} = \underline{x}) \Big| \\
&\qquad + \inf_{\underline{x}\in\mathcal{E}} P(G | \underline{X}_{t}^{t-r_1} = \underline{x}) \{P(H)P(E^c) + P(H \cap E^c)\}.
\end{aligned} \tag{41}$$

Using the triangle inequality, we have

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

Substituting the above into (41) gives (36), as required. □

We now obtain a bound for the first term on the right-hand side of (36).

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,*

$$\left|P(G|\underline{X}_t^{t-r_1}=\underline{x}) - P(G|\underline{X}_t^{t-r_1}=0)\right| \le \int_{\mathcal{G}} \mathcal{D}_{0,k,t}(\underline{y}|\underline{x})\,\mathrm{d}\underline{y}. \tag{42}$$

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

$$\int_{\mathcal{G}} \left|f_{X_{t+k}|\underline{X}_t^{t-r_1}}(y|\underline{x}) - f_{X_{t+k}|\underline{X}_t^{t-r_1}}(y|0)\right|\mathrm{d}y \le \mathbb{E}_{\underline{W}}\left(\int_{\mathbb{R}} \mathcal{D}_{0,k,t}(y|\underline{W},\underline{x})\,\mathrm{d}y\right) \tag{43}$$

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

$$\begin{aligned}&\int_{\mathcal{G}} \left|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) - f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\right|\mathrm{d}\underline{y}\\ &\quad\le \sum_{s=0}^{r_2} \mathbb{E}_{\underline{W}}\left(\sup_{\underline{y}_{s-1}} \int_{\mathbf{G}_s} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1},\underline{w},\underline{x})\,\mathrm{d}y_s\right).\end{aligned} \tag{44}$$

Proof. The proof of (42) is clear from the definition of $\mathcal{D}_{s,k,t}$, hence we omit the details.

To prove (43), 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

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

Substituting the above into $\int_{\mathcal{G}} |f_{X_{t+k}|\underline{X}_t^{t-r_1}}(y|\underline{x}) - f_{X_{t+k}|\underline{X}_t^{t-r_1}}(y|0)|\,\mathrm{d}y$ and using the definition of $\mathbb{E}_{\underline{W}}$ now gives (43).

To prove (44), we note that, by using the same argument used to prove (43), we have

$$f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) = \int_{\mathbb{R}^{k-1}} f_{\underline{W}}(\underline{w}) \prod_{s=0}^{r_2} f_{s,k,t}(y_s|\underline{y}_{s-1},\underline{w},\underline{x})\,\mathrm{d}\underline{w}. \tag{45}$$

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

$$
\begin{aligned}
&f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) - f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0) \\
&\quad = \sum_{s=0}^{r_2} \int_{\mathbb{R}^{k-1}} f_{\underline{W}}(\underline{w}) \left\{ \prod_{a=0}^{s-1} f_{a,k,t}(y_a|\underline{y}_{a-1}, \underline{w}, \underline{x}) \right\} \\
&\qquad\qquad \times \left\{ \prod_{b=s+1}^{r_2} f_{b,k,t}(y_b|\underline{y}_{b-1}, \underline{w}, 0) \right\} \\
&\qquad\qquad \times \{ f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) - f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, 0) \} \, \mathrm{d}\underline{w}.
\end{aligned}
\tag{46}
$$

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

$$
\begin{aligned}
&\int_{\mathcal{G}} \left| f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x}) - f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0) \right| \mathrm{d}\underline{y} \\
&\quad \le \sum_{s=0}^{r_2} \int_{\mathbb{R}^{k-1}} f_{\underline{W}}(\underline{w}) \left\{ \left[\prod_{a=0}^{s-1} \int_{\mathcal{G}_a} f_{a,k,t}(y_a|\underline{y}_{a-1}, \underline{w}, \underline{x}) \, \mathrm{d}y_a \right.\right. \\
&\qquad\qquad \left. \times \prod_{b=s+1}^{r_2} \int_{\mathcal{G}_b} f_{b,k,t}(y_b|\underline{y}_{b-1}, \underline{w}, \underline{x}) \, \mathrm{d}y_b \right] \\
&\qquad\qquad \left. \times \sup_{\underline{y}_{s-1}} \int_{\mathcal{G}_s} \left| f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) - f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, 0) \right| \mathrm{d}y_s \right\} \mathrm{d}\underline{w}.
\end{aligned}
\tag{47}
$$

Next, we observe that since $\mathcal{G}_j \subset \mathbb{R}$ and $\int_{\mathbb{R}} f_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{w}, \underline{x}) \, \mathrm{d}y_s = 1$, we have $(\prod_{a=0}^{s-1} \int_{\mathcal{G}_a} f_{a,k,t}(y_a|\underline{y}_{a-1}, \underline{w}, \underline{x}) \, \mathrm{d}y_a)(\prod_{b=s+1}^{r_2} \int_{\mathcal{G}_b} f_{b,k,t}(y_b|\underline{y}_{b-1}, \underline{w}, \underline{x}) \, \mathrm{d}y_b) \le 1$. Finally, substituting this bound into (47) gives (44). $\square$

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. We then have*

$$
\begin{aligned}
&\sum_{i,j} |P(G_i \cap H_j \cap E) - P(G_i \cap E) P(H_j)| \\
&\quad \le 2 \sum_i \sup_{\underline{x} \in \mathcal{E}} \left| P(G_i|\underline{X}_t^{t-r_1} = \underline{x}) - P(G_i|\underline{X}_t^{t-r_1} = 0) \right| + 2P(E^c) \quad \textit{and}
\end{aligned}
\tag{48}
$$

$$
\sum_{i,j} |P(G_i \cap H_j \cap E^c) - P(G_i \cap E^c) P(H_j)| \le 2P(E^c). \tag{49}
$$

Proof. Substituting the inequality in (36) into $\sum_{i,j}|P(G_i\cap H_j\cap E)-P(G_i\cap E)P(H_j)|$ gives

$$\begin{aligned}
&\sum_{i,j}|P(G_i\cap H_j\cap E)-P(G_i\cap E)P(H_j)|\\
&\quad\le 2\sum_j P(H_j)\sum_i\sup_{\underline{x}\in\mathcal{E}}\bigl|P(G_i|\underline{X}_t^{t-r_1}=\underline{x})-P(G_i|\underline{X}_t^{t-r_1}=0)\bigr| \\
&\qquad+\sum_{i,j}\inf_{\underline{x}\in\mathcal{E}}P(G_i|\underline{X}_t^{t-r_1}=\underline{x})\{P(H_j)P(E^c)+P(H_j\cap E^c)\}.
\end{aligned}\tag{50}$$

The sets $\{H_j\}$ are partitions of Ω, hence $\sum_i P(H_j)=1$ and $\sum_i P(H_j\cap E^c)\le 1$. Using these observations together with (50) gives (48).

Inequality (49) immediately follows from the fact that $\{H_j\}$ and $\{G_i\}$ are disjoint sets. □

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

$$\begin{aligned}
|P(G\cap H)-P(G)P(H)|&\le|P(G\cap H\cap E)-P(G\cap E)P(H)|\\
&\quad+|P(G\cap H\cap E^c)-P(G\cap E^c)P(H)|.
\end{aligned}$$

Now, by substituting (42) into (36) and using the above, we get

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

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

Proof of Proposition 2.1, equation (6). It is worth noting that the proof of (6) is similar to the proof of (5). Using (48) and the same arguments as those in the proof of (5), we have

$$\begin{aligned}
&\sum_{i,j}|P(G_i\cap H_j)-P(G_i)P(H_j)|\\
&\quad\le 2\sum_i\sup_{\underline{x}\in\mathcal{E}}\int_{\mathcal{G}_i}\bigl|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x})-f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\bigr|\,\mathrm{d}\underline{y}+4P(E^c)\\
&\quad\le 2\sum_i\int_{\mathcal{G}_i}\sup_{\underline{x}\in\mathcal{E}}\bigl|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x})-f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\bigr|\,\mathrm{d}\underline{y}+4P(E^c)\\
&\quad\le 2\int_{\mathbb{R}^{r_2+1}}\sup_{\underline{x}\in\mathcal{E}}\bigl|f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|\underline{x})-f_{\underline{X}_{t+k+r_2}^{t+k}|\underline{X}_t^{t-r_1}}(\underline{y}|0)\bigr|\,\mathrm{d}\underline{y}+4P(E^c),
\end{aligned}\tag{51}$$

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). □

Proof of Proposition 2.1, equation (7). To prove the result, we substitute the bound in (44) into (5) to obtain (7). □

Proof of Proposition 2.1, equation (8). To prove (8), we substitute (44) into (6) to obtain (8). □

A.2. Proofs in Section 3

Proof of Lemma 3.1. We first prove (12) with $s = 0$. Suppose that $k \geq 1$. Focusing on the first element of $\underline{X}_{t+k}^{t+k-p+1}$ in (12) and factoring out Z_{t+k} gives

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

which is (12) (with $s = 0$). To prove (12) for $1 \leq s \leq p$, we note that using the tvARCH(p) representation in (11) and (12) for $s = 0$ gives

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

where $\mathcal{P}_{s,k,t}$ and $\mathcal{Q}_{s,k,t}$ are defined in (13). Hence, this gives (12). 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. □

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 σ-algebras generated by $\underline{X}_{t+k+r_2}^{t+k}$ and $(\underline{Z}_{t+k+r_2}^{t+k+p}, \underline{X}_{t+k+p-1}^{t+k})$ are the same. Moreover, by using the fact that for all $\tau > t$, Z_τ is independent of X_t, we have

$$\begin{aligned} &\sup_{G\in\mathcal{F}_\infty^{t+k}, H\in\mathcal{F}_t^{-\infty}} |P(G\cap H) - P(G)P(H)| \\ &\quad = \sup_{G\in\mathcal{F}_{t+k+p-1}^{t+k}, H\in\mathcal{F}_t^{t-p+1}} |P(G\cap H) - P(G)P(H)|. \end{aligned} \tag{52}$$

Now, by using the above, Proposition 2.1, equation (7), and the fact 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

$$
\begin{aligned}
&\sup_{G\in\mathcal{F}_{t+k+p-1}^{t+k}, H\in\mathcal{F}_t^{t-p+1}} |P(G\cap H)-P(G)P(H)| \\
&\quad\le 2\sup_{\underline{x}\in\mathcal{E}}\sum_{s=0}^{p-1}\mathbb{E}_{\underline{Z}}\biggl(\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\int \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1},\underline{z},\underline{x})\,\mathrm{d}y_s\biggr) \qquad (53)\\
&\qquad + 4P(X_t>\eta_0 \text{ or},\ldots,X_{t-p+1}>\eta_{-p+1}).
\end{aligned}
$$

Finally, using the fact that $P(X_t>\eta_0 \text{ or } X_{t-1}>\eta_{-1},\ldots,X_{t-p+1}>\eta_{-p+1})\le\sum_{j=0}^{p-1}P(X_{t-j}>\eta_{-j})$ gives (15).

The proof of (16) is similar to the proof above, but uses (8) instead of (7), so we omit the details. □

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

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

$$
\int_{\mathbb{R}}\biggl|\frac{1}{A+B}f_Z\biggl(\frac{y}{A+B}\biggr)-\frac{1}{A}f_Z\biggl(\frac{y}{A}\biggr)\biggr|\,\mathrm{d}y\le K\biggl(\frac{B}{A}+\frac{B}{A+B}\biggr). \qquad (54)
$$

If 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*

$$
\int_{\mathbb{R}}\sup_{\underline{x}\in\mathcal{E}}\biggl|\frac{1}{A+B(\underline{x})}f_Z\biggl(\frac{y}{A+B(\underline{x})}\biggr)-\frac{1}{A}f_Z\biggl(\frac{y}{A}\biggr)\biggr|\,\mathrm{d}y\le K\sup_{\underline{x}\in\mathcal{E}}\biggl(\frac{B(\underline{x})}{A}+\frac{B(\underline{x})}{A+B(\underline{x})}\biggr). \qquad (55)
$$

Proof. To prove (54), we observe that

$$
\int_{\mathbb{R}}\biggl|\frac{1}{A+B}f_Z\biggl(\frac{y}{A+B}\biggr)-\frac{1}{A}f_Z\biggl(\frac{y}{A}\biggr)\biggr|\,\mathrm{d}y=I+II,
$$

where

$$
I=\int_{\mathbb{R}}\frac{1}{A+B}\biggl|f_Z\biggl(\frac{y}{A+B}\biggr)-f_Z\biggl(\frac{y}{A}\biggr)\biggr|\,\mathrm{d}y \quad\text{and}\quad II=\int_{\mathbb{R}}\biggl(\frac{1}{A+B}-\frac{1}{A}\biggr)f_Z\biggl(\frac{y}{A}\biggr).
$$

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

$$
I\le\int_{\mathbb{R}}\biggl|f_Z(u)-f_Z\biggl(u\biggl(1+\frac{B}{A}\biggr)\biggr)\biggr|\,\mathrm{d}u\le K\frac{B}{A}.
$$

It is straightforward to show that $II\le\frac{B}{A+B}$. Hence, the bounds for I and II give (54).

The proof of (55) is the same as above, but uses Assumption 3.1(iv) instead of Assumption 3.1(iii), so we omit the details. □

Proof of Lemma 3.3. We first show that

$$\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\int \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1},\underline{z},\underline{x})\,\mathrm{d}y_s \le \frac{K}{\inf_{t\in\mathbb{Z}} a_0(t)}\mathcal{Q}_{s,k,t}(\underline{z},\underline{x}) \tag{56}$$

and use this to prove (17). We note that when $\underline{x}=0$, $\mathcal{Q}_{s,k,t}(\underline{z},0)=0$ and $f_{s,k,t}(y_s|\underline{y}_{s-1},\underline{z},0)=\mathcal{P}_{s,k,t}(\underline{z})^{-1}f_Z(\frac{y_s}{\mathcal{P}_{s,k,t}(\underline{z})})$. Therefore, using (14) gives

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

Now, recalling that $\mathcal{P}_{s,k,t}$ and $\mathcal{Q}_{s,k,t}$ are both positive and setting $A=\mathcal{P}_{s,k,t}(\underline{z})$, $B=\mathcal{Q}_{s,k,t}(\underline{z},\underline{x})$ and using (54), we have

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

Finally, since $\mathcal{P}_{s,k,t}(\underline{z}) > \inf_{t\in\mathbb{Z}} a_0(t)$, we have $\int_{\mathbb{R}}\mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1},\underline{z},\underline{x})\,\mathrm{d}y_s \le K\frac{\mathcal{Q}_{s,k,t}(\underline{z},\underline{x})}{\inf_{t\in\mathbb{Z}} a_0(t)}$, thus giving (56). By using (56), we now prove (17). Substituting (56) into the integral on the left-hand side of (17), using the fact that $\mathbb{E}[\mathcal{Q}_{s,k,t}(\underline{Z},\underline{x})]=\mathcal{Q}_{s,k,t}(\underline{1}_{k-1},\underline{x})$ and substituting (56) into (15) gives

$$\mathbb{E}_{\underline{Z}}\left(\sup_{\underline{y}_{s-1}\in\mathbb{R}^s}\int_{\mathbb{R}}\mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1},\underline{Z},\underline{x})\,\mathrm{d}y_s\right)\le K\frac{\mathbb{E}[\mathcal{Q}_{s,k,t}(\underline{Z},\underline{x})]}{\inf_{t\in\mathbb{Z}} a_0(t)} = K\frac{\mathcal{Q}_{s,k,t}(\underline{1}_{k-1},\underline{x})}{\inf_{t\in\mathbb{Z}} a_0(t)}. \tag{57}$$

We now find a bound for $\mathcal{Q}_{s,k,t}$. By the definition of $\mathcal{Q}_{s,k,t}$ in (13) and using the matrix norm inequality $[A\underline{x}]_1 \le K\|A\|_{spec}\|\underline{x}\|$ ($\|\cdot\|_{spec}$ is the spectral norm), we have

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

To bound the above, we note that by Assumption 3.1(i), $\sup_{t\in\mathbb{Z}}\sum_{j=1}^{p} a_j(t)\le(1-\delta)$, therefore there exists a $\tilde{\delta}$, where $0<\tilde{\delta}<\delta<1$ and such that, for all t, we have $\|A_{t+k+s-i}\times$

$\{\prod_{d=0}^{k-1} A_{t+k+s-i-d}\}\|_{spec} \leq K(1-\tilde{\delta})^{k+1}$ for some finite K. Combining all of this gives

$$\begin{aligned} \mathcal{Q}_{s,k,t}(\underline{1}_{k-1}, \underline{x}) &\leq \frac{K}{\inf_{t\in\mathbb{Z}} a_0(t)} \sum_{i=s}^{p} a_i(t+k+s) \left\| A_{t+k+s-i} \left\{ \prod_{d=0}^{k+s-i} A_{t+k+s-i-d} \right\} \right\|_{spec} \|\underline{x}\| \\ &\leq \frac{K}{\inf_{t\in\mathbb{Z}} a_0(t)} \sum_{i=s}^{p} a_i(t+k+s)(1-\tilde{\delta})^{k+s-i} \|\underline{x}\|. \end{aligned} \tag{58}$$

Substituting the above into (57) gives (17).

We now prove (18). We use the same proof to show (56), but apply (54) instead of (55) to obtain

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

By substituting the above into (16) and using the same proof to prove (17), we obtain

$$\begin{aligned} &\sum_{s=0}^{p-1} \int \prod_{i=1}^{k-1} f_Z(z_i) \sup_{\underline{y}_{s-1}\in\mathbb{R}^s} \left\{ \int_{\mathbb{R}} \sup_{\underline{x}\in\mathcal{E}} \mathcal{D}_{s,k,t}(y_s|\underline{y}_{s-1}, \underline{z}, \underline{x})\,\mathrm{d}y_s \right\} \mathrm{d}\underline{z} \\ &\quad \leq K \frac{\mathbb{E}[\sup_{\underline{x}\in\mathcal{E}} \mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})]}{\inf_{t\in\mathbb{Z}} a_0(t)}. \end{aligned} \tag{59}$$

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 $\mathbb{E}[\sup_{\underline{x}\in\mathcal{E}} \mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})] \leq \sup_{\underline{x}\in\mathcal{E}} \mathbb{E}[\mathcal{Q}_{s,k,t}(\underline{Z}, \underline{x})] = \sup_{\underline{x}\in\mathcal{E}} \mathcal{Q}_{s,k,t}(\underline{1}_{k-1}, \underline{x})$. Hence, by using (58), 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 (59) gives (18). □

Acknowledgements

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

References

[1] Athreya, K.B. and Pantula, S.G. (1986). Mixing properties of Harris chains and autoregressive processes. *J. Appl. Probab.* **23** 880–892. MR0867185

[2] Basrak, B., Davis, R.A. and Mikosch, T. (2002). Regular variation of GARCH processes. *Stochastic Process. Appl.* **99** 95–115. MR1894253

[3] Berkes, I., Hörmann, S. and Schauer, J. (2009). Asymptotic results for the empirical process of stationary sequences. *Stochastic Process. Appl.* **119** 1298–1324. MR2508575

[4] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*. New York: Springer. MR1640691

[5] Bougerol, P. and Picard, N. (1992). Stationarity of GARCH processes and some nonnegative time series. *J. Econometrics* **52** 115–127. MR1165646

[6] Bousamma, F. (1998). Ergodicité, mélange et estimation dans les modèles GARCH. Ph.D. thesis, Paris 7.

[7] Bradley, R.C. (2007). *Introduction to Strong Mixing Conditions Volumes 1, 2 and 3*. Herber City, UT: Kendrick Press. MR2325294

[8] Carrasco, M. and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econom. Theory* **18** 17–39. MR1885348

[9] Chanda, K.C. (1974). Strong mixing properties of linear stochastic processes. *J. Appl. Probab.* **11** 401–408. MR0358951

[10] Dahlhaus, R. and Polonik, W. (2006). Nonparametric quasi-maximum likelihood estimation for Gaussian locally stationary processes. *Ann. Statist.* **34** 2790–2842. MR2329468

[11] Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference of time varying ARCH processes. *Ann. Statist.* **34** 1074–1114. MR2278352

[12] Davidson, J. (1994). *Stochastic Limit Theory*. Oxford: Oxford Univ. Press. MR1430804

[13] Feigin, P.D. and Tweedie, R.L. (1985). Random coefficient autoregressive processes. A Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* **6** 1–14. MR0792428

[14] Francq, C. and Zakoïan, J.-M. (2006). Mixing properties of a general class of GARCH(1, 1) models without moment assumptions on the observed process. *Econom. Theory* **22** 815–834. MR2291218

[15] Fryzlewicz, P., Sapatinas, T. and Subba Rao, S. (2008). Normalised least squares estimation in time-varying ARCH models. *Ann. Statist.* **36** 742–786. MR2396814

[16] Fryzlewicz, P. and Subba Rao, S. (2009). BaSTA: Consistent multiscale multiple change-point detection for piecewise-stationary ARCH processes. Preprint. Available at http://stats.lse.ac.uk/fryzlewicz/basta/basta.pdf.

[17] Giraitis, L., Kokoskza, P. and Leipus, R. (2000). Stationary ARCH models: Dependence structure and central limit theorem. *Econom. Theory* **16** 3–22. MR1749017

[18] Giraitis, L., Leipus, R. and Surgailis, D. (2005). Recent advances in ARCH modelling. In *Long Memory in Economics* (A. Kirman and G. Teyssiere, eds.) 3–39. Berlin: Springer. MR2265054

[19] Giraitis, L. and Robinson, P.M. (2001). Whittle estimation of ARCH models. *Econom. Theory* **17** 608–631. MR1841822

[20] Gorodetskii, V.V. (1977). On the strong mixing propery for linear sequences. *Theory Probab. Appl.* **22** 411–413.

[21] Hall, P. and Heyde, C.C. (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press. MR0624435

[22] Hannan, E.J. and Kavaliers, L. (1986). Regression, autoregression models. *J. Time Ser. Anal.* **7** 27–49. MR0832351

[23] Hörmann, S. (2008). Augmented GARCH sequences: Dependence structure and asymptotics. *Bernoulli* **14** 543–561. MR2544101

[24] Ibragimov, I.A. (1962). Some limit theorems for stationary processes. *Theory Probab. Appl.* **7** 349–382. MR0148125

[25] Liebscher, E. (2005). Towards a unified approach for proving geometric ergodicity and mixing properties of nonlinear autoregressive processes. *J. Time Ser. Anal.* **26** 669–689. MR2188304

[26] Lindner, A. (2009). Stationarity, mixing, distributional properties and moments of GARCH(p,q)-processes. In *Handbook of Financial Time Series* (T. Andersen, R. Davis, J.-P. Kreiss and T. Mikosch, eds.) 43–69. Berlin: Springer.

[27] Meitz, M. and Saikkonen, P. (2008). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econom. Theory* **24** 1291–1320. MR2440741

[28] Meyn, S.P. and Tweedie, R.L. (1993). *Markov Chains and Stochastic Stability*. Berlin: Springer. MR1287609

[29] Mikosch, T. and Stărică, C. (2003). Long-range dependence effects and arch modelling. In *Theory and Applications of Long Range Dependence* (P. Doukhan, G. Oppenheim and M.S. Taqqu, eds.) 439–459. Boston: Birkhäuser. MR1957503

[30] Mokkadem, A. (1990). Propertiés de mélange des processus autorégressifs polnomiaux. *Ann. Inst. H. Poincaré Probab. Statist.* **26** 219–260. MR1063750

[31] Pham, D.T. (1986). The mixing propery of bilinear and generalised random coefficient autorregressive models. *Stochastic Process. Appl.* **23** 291–300. MR0876051

[32] Pham, D.T. and Tran, T.T. (1985). Some mixing properties of time series models. *Stochastic Process. Appl.* **19** 297–303. MR0787587

[33] Robinson, P.M. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasity in multiple regression. *J. Econom.* **47** 67–78. MR1087207

[34] Robinson, P.M. and Zaffaroni, P. (2006). Pseudo-maximum likelihood estimation of ARCH(∞) models. *Ann. Statist.* **34** 1049–1074. MR2278351

[35] Sorokin, A.A. (2006). Uniform bound for strong mixing coefficient and maximum of residual empirical process of ARCH sequence (in Russian). Available at arXiv:math/0610747v1.

[36] Straumann, D. and Mikosch, T. (2006). Quasi-maximum likelihood estimation in conditionally hetroscedastic time series: A stochastic recurrence equation approach. *Ann. Statist.* **34** 2449–2495. MR2291507

[37] Subba Rao, S. (2006). A note on uniform convergence of an ARCH(∞) estimator. *Sankhyā* **68** 600–620. MR2370392

[38] Tjostheim, D. (1990). Nonlinear time series and Markov chains. *Adv. in Appl. Probab.* **22** 587–611. MR1066965

[39] Volkonskii, V.A. and Rozanov, Y.A. (1959). Some limit theorems for random functions I. *Theory Probab. Appl.* **4** 178–197. MR0121856

Received November 2008 and revised February 2010