Adv. Appl. Prob. **38,** 1155–1172 (2006)
Printed in Northern Ireland

ON SOME NONSTATIONARY, NONLINEAR RANDOM PROCESSES AND THEIR STATIONARY APPROXIMATIONS

SUHASINI SUBBA RAO,* *Universität Heidelberg*

Abstract

In this paper our object is to show that a certain class of nonstationary random processes can locally be approximated by stationary processes. The class of processes we are considering includes the time-varying autoregressive conditional heteroscedastic and generalised autoregressive conditional heteroscedastic processes, amongst others. The measure of deviation from stationarity can be expressed as a function of a derivative random process. This derivative process inherits many properties common to stationary processes. We also show that the derivative processes obtained here have alpha-mixing properties.

Keywords: Derivative process; local stationarity; nonlinearity; nonstationarity; state space model

2000 Mathematics Subject Classification: Primary 62M10
Secondary 62F10

1. Introduction

Linear time series models are often used in time series analysis and it is usually assumed that the underlying process is stationary. However, it may be that the assumption of stationarity is sometimes unrealistic, especially when we observe the process over long periods of time. Several nonstationary models have been introduced (see, for example, Priestley (1965) and Cramér (1961)), but many asymptotic results available for stationary time series are not immediately applicable to nonstationary time series. To circumvent this, Dahlhaus (1997) used a rescaling technique to define the notion of local stationarity. By using a time-varying spectral density function, Dahlhaus (1997) defined locally stationary processes. However, so far these methods have been used exclusively for the analysis of nonstationary linear processes. Here our object is to analyse nonstationary, nonlinear random processes.

In the last 20 years, nonlinear time series methods have received considerable attention, although they have mainly been restricted to stationary processes. Standard nonlinear models include autoregressive conditional heteroscedastic (ARCH) models (Engle (1982)), generalised ARCH (GARCH) models (Bollerslev (1986)), bilinear models (Subba Rao (1977), Terdik (1999)), and random-coefficient processes (Nicholls and Quinn (1982)). Often these stationary, nonlinear processes have a state space representation; see, for example, Brandt (1986), Bougerol and Picard (1992a), and Straumann and Mikosch (2006).

In this paper we consider nonstationary, nonlinear processes with state space representations and time-dependent parameters. In particular, we consider the nonstationary process $\{\mathcal{X}_{t,N}\}$

Received 23 January 2004; revision received 3 August 2006.
* Current address: Department of Statistics, Texas A&M University, 3143 TAMU, College Station, TX 77843-3143, USA. Email address: suhasini@stat.tamu.edu

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

which admits the time-varying state space representation

$$\mathcal{X}_{t,N} = \boldsymbol{A}_t\left(\frac{t}{N}\right)\mathcal{X}_{t-1,N} + \boldsymbol{b}_t\left(\frac{t}{N}\right), \qquad t = 1, \ldots, N, \tag{1.1}$$

where, for each t, $\mathcal{X}_{t,N}$ and $\boldsymbol{b}_t(t/N)$ are p-dimensional nonstationary random vectors and $\boldsymbol{A}_t(t/N)\}$ is a $(p \times p)$-dimensional nonstationary random matrix.

In Section 2 we will show that, under suitable conditions on the nonstationary random vectors $\{\boldsymbol{b}_t(t/N) \colon t \in \mathbb{Z}\}$ and matrices $\{\boldsymbol{A}_t(t/N) \colon t \in \mathbb{Z}\}$, $\{\mathcal{X}_{t,N}\}$ can locally be approximated by the stationary process $\{\mathcal{X}_t(u)\}$ given by

$$\mathcal{X}_t(u) = \boldsymbol{A}_t(u)\mathcal{X}_{t-1}(u) + \boldsymbol{b}_t(u), \tag{1.2}$$

where u is fixed and $\{\boldsymbol{b}_t(u) \colon t \in \mathbb{Z}\}$ and $\{\boldsymbol{A}_t(u) \colon t \in \mathbb{Z}\}$ are respectively p-dimensional stationary random vectors and $(p \times p)$-dimensional stationary random matrices. We will prove that $\mathcal{X}_t(u)$ can be regarded as a stationary approximation of $\mathcal{X}_{t,N}$ for values of t/N close to u. In Section 3 we define the derivative process which is a measure of the deviation of $\mathcal{X}_{t,N}$ from the stationary process $\mathcal{X}_t(u)$ and obtain an exact bound for this deviation, using a stochastic Taylor series expansion. We also show that the derivative process satisfies a stochastic differential equation. Using the derivative process, we consider some probabilistic results (such as mixing properties) associated with the observed process and the Taylor series expansions. In Section 5 we consider the particular example of the time-varying GARCH process, and show that all the results stated above apply to it. We mention that other processes, such as the time-varying random-coefficient autoregressive and GARCH processes also have the representation (1.1) and obey the results here. The idea of a stationary approximation with a derivative process was established for time-varying ARCH processes by Dahlhaus and Subba Rao (2006).

2. Nonlinear time-varying processes

2.1. Assumptions

In this section we state our assumptions and notation.

Let $\|\boldsymbol{x}\|_m$ and $\|\boldsymbol{A}\|_m$ respectively denote the ℓ_m-norms of the vector $\boldsymbol{x}$ and matrix $\boldsymbol{A}$. Let $\|\boldsymbol{A}\|_{\text{spec}}$ denote the spectral norm, where $\|\boldsymbol{A}\|_{\text{spec}} = \sup_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{A}\boldsymbol{x}\|_2$. Suppose that $B_{i,j}$ denotes the (i, j)th element of the matrix $\boldsymbol{B}$. Let $|\boldsymbol{B}|_{\text{abs}}$ denote the absolute matrix (or vector, as appropriate) of $\boldsymbol{B}$, where $(|\boldsymbol{B}|_{\text{abs}})_{i,j} = |B_{i,j}|$. We say that $\boldsymbol{A} \leq \boldsymbol{B}$ if $A_{i,j} \leq B_{i,j}$ for all i and j. Let $\lambda_{\text{spec}}(\boldsymbol{A})$ denote the largest absolute eigenvalue of the matrix $\boldsymbol{A}$, and let $\sup_u \boldsymbol{A}(u)$ be defined as $\sup_u \boldsymbol{A}(u) = \{\sup_u |\boldsymbol{A}(u)_{i,j}| \colon i = 1, \ldots p,\ j = 1, \ldots, q\}$. To simplify notation we will denote the ℓ_2-norm of a vector $\boldsymbol{x}$ (or of a matrix) as $\|\boldsymbol{x}\| \equiv \|\boldsymbol{x}\|_2$.

The Lyapunov exponent associated with a sequence of random matrices $\{\boldsymbol{A}_t \colon t \in \mathbb{N}\}$ is defined as

$$\inf\left\{\frac{1}{n}\,\mathrm{E}(\log \|\boldsymbol{A}_t\boldsymbol{A}_{t-1}\cdots\boldsymbol{A}_{t-n+1}\|_{\text{spec}}) \colon n \in \mathbb{N}\right\}.$$

We make the following assumption.

Assumption 2.1. *For every N, the stochastic process $\{\mathcal{X}_{t,N}\}$ has a time-varying state space representation defined as in (1.1), where, for all $u \in [0, 1]$, the random matrices $\{\boldsymbol{A}_t(u)\}$ and random vectors $\{\boldsymbol{b}_t(u)\}$ satisfy the following assumptions.*

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

(i) *There exists an $M \in \mathbb{N}$ such that, for each $r \in \{1, \dots, M\}$, there exists a sequence of independent, identically distributed, positive random matrices $\{\mathcal{A}_t(r) \colon t \in \mathbb{Z}\}$ such that*

$$|\boldsymbol{A}_t(u)|_{\text{abs}} \le \mathcal{A}_t(r) \quad \text{if } u \in \left[\frac{r-1}{M}, \frac{r}{M}\right)$$

and, for some $\delta < 0$, the Lyapunov exponent of $\{\mathcal{A}_t(r) \colon t \in \mathbb{Z}\}$ is less than δ. There exists a stationary sequence $\{\tilde{\boldsymbol{b}}_t\}$ such that $\sup_u |\boldsymbol{b}_t(u)|_{\text{abs}} \le \tilde{\boldsymbol{b}}_t$. Furthermore, for each $k \in \{1, \dots, M\}$ and for some $\varepsilon > 0$, $\mathrm{E}(\|\tilde{\boldsymbol{b}}_t\|_1^{\varepsilon}) < \infty$ and $\mathrm{E}(\|\mathcal{A}_t(r)\|_1^{\varepsilon}) < \infty$.

(ii) *There exist a $\beta \in (0, 1]$ and matrices $\{\mathcal{A}_t\}$ such that, for all $u, v \in [0, 1]$, the matrices $\{\boldsymbol{A}_t(\cdot)\}$ and vectors $\{\boldsymbol{b}_t(\cdot)\}$ satisfy*

$$|\boldsymbol{A}_t(u) - \boldsymbol{A}_t(v)|_{\text{abs}} \le C|u - v|^{\beta} \mathcal{A}_t, \qquad |\boldsymbol{b}_t(u) - \boldsymbol{b}_t(v)|_{\text{abs}} \le C|u - v|^{\beta} \tilde{\boldsymbol{b}}_t,$$

with $\tilde{\boldsymbol{b}}_t$ as defined in (i) and C a finite constant. Furthermore, $\mathrm{E}(\|\mathcal{A}_t\|^{\varepsilon}) < \infty$ for some $\varepsilon > 0$.

For convenience, from now on we let $\boldsymbol{A}_t(u) = \boldsymbol{0}$ for $u \le 0$ and $\prod_{i=0}^{-k} \boldsymbol{A}_i = \boldsymbol{I}$ (if $k \ge 1$), the identity matrix.

Assumption 2.1(i) means that the random matrices $\{\boldsymbol{A}_t(u)\}$ are dominated by random matrices which have a negative Lyapunov exponent. As will become clear below, this implies that $\{\mathcal{X}_{t,N}\}$ has a unique causal solution. Assumption 2.1(ii) is used to approximate $\{\mathcal{X}_{t,N}\}$ locally by a stationary process.

2.2. The stationary approximation

Using the arguments of Bougerol and Picard (1992b, Theorem 2.5), we can show that the unique causal solution of $\{\mathcal{X}_{t,N}\}$ is almost surely

$$\mathcal{X}_{t,N} = \sum_{k=0}^{\infty} \boldsymbol{A}_t\left(\frac{t}{N}\right) \cdots \boldsymbol{A}_{t-k+1}\left(\frac{t-k+1}{N}\right) \boldsymbol{b}_{t-k}\left(\frac{t-k}{N}\right). \tag{2.1}$$

One of our main results is the theorem below, where we show that $\mathcal{X}_{t,N}$ can locally be approximated by the stochastic process $\mathcal{X}_t(u)$. We let

$$\boldsymbol{Y}_t = \sum_{k=1}^{\infty} \sum_{r=1}^{M} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(r) \tilde{\boldsymbol{b}}_{t-k}. \tag{2.2}$$

Theorem 2.1. *Suppose that Assumption 2.1 holds, let $\mathcal{X}_{t,N}$, $\mathcal{X}_t(u)$, and $\boldsymbol{Y}_t$ be as defined in (1.1), (1.2), and (2.2), respectively, and suppose there exists an $\varepsilon > 0$ such that*

$$\sup_{t,N} \mathrm{E}(\|\mathcal{X}_{t,N}\|_1^{\varepsilon}) < \infty, \qquad \sup_{u} \mathrm{E}(\|\boldsymbol{Y}_t\|_1^{\varepsilon}) < \infty. \tag{2.3}$$

Then we have

$$\left|\mathcal{X}_{t,N} - \mathcal{X}_t\left(\frac{t}{N}\right)\right|_{\text{abs}} \le \frac{1}{N^{\beta}} \boldsymbol{V}_{t,N}, \tag{2.4}$$

$$|\mathcal{X}_t(u) - \mathcal{X}_t(w)|_{\text{abs}} \le |u - w|^{\beta} \boldsymbol{W}_t, \tag{2.5}$$

$$|\mathcal{X}_{t,N} - \mathcal{X}_t(u)|_{\text{abs}} \le \left|\frac{t}{N} - u\right|^{\beta} \boldsymbol{W}_t + \frac{1}{N^{\beta}} \boldsymbol{V}_{t,N}, \tag{2.6}$$

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

where

$$\boldsymbol{V}_{t,N} = C \sum_{k=1}^{\infty} k \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(i_1)\{\mathcal{A}_{t-k}|\mathcal{X}_{t-k-1,N}|_{\mathrm{abs}} + \tilde{\boldsymbol{b}}_{t-k}\}, \tag{2.7}$$

$$\boldsymbol{W}_t = C \sum_{k=1}^{\infty} \sum_{r=1}^{M} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(r)\{\mathcal{A}_{t-k}\boldsymbol{Y}_{t-k-1} + \tilde{\boldsymbol{b}}_{t-k}\}, \tag{2.8}$$

and i_1 is such that $(i_1 - 1)/M \le t/N < i_1/M$. Moreover, $\boldsymbol{V}_{t,N}$ converges almost surely and the series $\{\boldsymbol{W}_t\}_t$ is a well-defined stationary process.

Proof. See Appendix A.

The most notable result in the theorem above is (2.6), which states that the deviation between the nonstationary process $\mathcal{X}_{t,N}$ and the stationary process $\mathcal{X}_t(u)$ depends on the difference $|t/N - u|$, that is,

$$|\mathcal{X}_{t,N} - \mathcal{X}_t(u)|_{\mathrm{abs}} \le \left|\frac{t}{N} - u\right|^{\beta} O_p(\mathbf{1}) + \frac{1}{N^{\beta}} O_p(\mathbf{1}). \tag{2.9}$$

A simple application of the theorem above is the evaluation of the sampling properties of local averages of time-varying processes. For example, suppose that $|t_0/N - u_0| < 1/N$ and that we average $\mathcal{X}_{t,N}$ over a neighbourhood whose length, $(2M+1)$, increases as N increases although $M/N \to 0$. Then, using the theorem above, we have

$$\frac{1}{2M+1} \sum_{k=-M}^{M} \mathcal{X}_{t_0+k,N} = \frac{1}{2M+1} \sum_{k=-M}^{M} \mathcal{X}_{t_0+k}(u_0) + \mathcal{B}_{t_0,N}, \tag{2.10}$$

where

$$\|\mathcal{B}_{t_0,N}\|_1 \le \frac{1}{2M+1} \sum_{k=-M}^{M} \left(\left(\frac{k}{N}\right)^{\beta} \|\boldsymbol{W}_{t_0+k}\|_1 + \frac{1}{N^{\beta}} \|\boldsymbol{V}_{t_0+k,N}\|_1\right).$$

To evaluate the limit of this sum, which involves showing that $\mathcal{B}_{t_0,N}$ converges asymptotically to $\mathbf{0}$, we require the existence of moments of $\mathcal{X}_{t,N}$ and its related processes. We consider this in the section below.

2.3. Existence of moments

It is worth noting that the local approximation of $\{\mathcal{X}_{t,N}\}$ by a stationary process requires relatively weak assumptions on the moments of $\{\boldsymbol{A}_t(u)\}$ and $\{\boldsymbol{b}_t(u)\}$. However, under stronger assumptions on the moments of $\{\boldsymbol{A}_t(u)\}$ and $\{\boldsymbol{b}_t(u)\}$, we will show that $\mathrm{E}(\|\mathcal{X}_{t,N}\|_n^n)$ is uniformly bounded in t and N.

Define the matrix $[\boldsymbol{A}]_n$ as $[\boldsymbol{A}]_n = \{\mathrm{E}(|A_{i,j}|^n)^{1/n} : i = 1, \dots, p,\ j = 1, \dots, q\}$. We now give conditions for $\mathrm{E}(\|\mathcal{X}_{t,N}\|_n^n) < \infty$ and $\mathrm{E}(\|\mathcal{X}_t(u)\|_n^n) < \infty$ to hold.

Proposition 2.1. *Suppose that Assumption 2.1 holds and let $\mathcal{X}_{t,N}$, $\mathcal{X}_t(u)$, $\boldsymbol{V}_{t,N}$, and $\boldsymbol{W}_t$ be as defined in (1.1), (1.2), (2.7), and (2.8), respectively. Suppose, for all $r \in \{1, \dots, M\}$ and some $n \in [1, \infty)$, that $\mathrm{E}(\|\tilde{\boldsymbol{b}}_t\|_n^n) < \infty$, and, for some $\delta > 0$, that $\lambda_{\mathrm{spec}}([\boldsymbol{A}_t(r)]_n) < 1 - \delta$. Then*

$$\begin{aligned} &\sup_{t,N} \mathrm{E}(\|\mathcal{X}_{t,N}\|_n^n) < \infty, \qquad \sup_{t,N} \mathrm{E}(\|\boldsymbol{V}_{t,N}\|_n^n) < \infty, \\ &\sup_{u} \mathrm{E}(\|\mathcal{X}_t(u)\|_n^n) < \infty, \qquad \mathrm{E}(\|\boldsymbol{W}_t\|_n^n) < \infty. \end{aligned} \tag{2.11}$$

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

If the conditions of Proposition 2.1 are satisfied, then condition (2.3) is satisfied with $\varepsilon = n$; hence, the local stationarity conclusions of Theorem 2.1 immediately follow.

We now apply the above results to the local averages example in (2.10). Under the assumption that all the conditions in Proposition 2.1 are satisfied with $n = 1$, we have

$$\mathcal{B}_{t_0,N} \leq \left(\frac{M}{N}\right)^{\beta} O_p(\mathbf{1}) + \frac{1}{N^{\beta}} O_p(\mathbf{1})$$

and

$$\frac{1}{2M+1}\sum_{k=-M}^{M} \mathcal{X}_{t_0+k,N} = \frac{1}{2M+1}\sum_{k=-M}^{M} \mathcal{X}_{t_0+k}(u_0) + O_p\left(\left(\frac{M}{N}\right)^{\beta}\mathbf{1} + \frac{\mathbf{1}}{N^{\beta}}\right).$$

Therefore, if the process $\mathcal{X}_{t_0+k}(u_0)$ were ergodic, we would have $\mathcal{B}_{t_0,N} \xrightarrow{\mathrm{P}} \mathbf{0}$ and

$$\frac{1}{2M+1}\sum_{k=-M}^{M} \mathcal{X}_{t_0+k,N} \xrightarrow{\mathrm{P}} \mathrm{E}(\mathcal{X}_{t_0+k}(u_0)),$$

where $M \to 0$ and $M/N \to 0$ as $N \to \infty$. The results in the following section allow us to obtain a tighter bound for $\mathcal{B}_{t_0,N}$.

3. The derivative process and its state space representation

In the previous section we showed that time-varying processes can locally be approximated by stationary processes. In this section, under additional conditions on $\{\boldsymbol{A}_t(u)\}$ and $\{\boldsymbol{b}_t(u)\}$ we improve the approximation in (2.9) and show that a Taylor series expansion of the time-varying process in terms of stationary processes can be derived (see Theorem 3.2). In order to do this, we define the derivative process and show that it also has a state space representation.

The Taylor expansion of a given time-varying process in terms of stationary processes is of particular importance in theoretical investigations, since classical results for stationary sequences such as ergodic theorems and central limit theorems can fruitfully be used. In applications, it is unlikely that the stationary derivative process will be observed. It is more likely that the derivatives of the parameters $\{\boldsymbol{A}_t(t/N)\}$ will either be known or can be estimated. However, the state space representation motivates our definition of the time-varying derivative process. If the derivatives $\{\dot{\boldsymbol{A}}_t(t/N)\}$ (defined below) are known, the time-varying derivative process can be obtained from the original time-varying process.

Here we focus our discussion on the first derivatives of the process. However, under suitable conditions all the results stated here apply to higher-order derivative processes as well.

Suppose that $\boldsymbol{A}(u)$ is a $p \times q$ random matrix; we let

$$\dot{\boldsymbol{A}}_t(u) = \left\{\frac{\partial \boldsymbol{A}(u)_{i,j}}{\partial u} : i = 1, \dots, p,\ j = 1, \dots, q\right\}.$$

We make the following assumption.

Assumption 3.1. *For every N, the stochastic process $\{\mathcal{X}_{t,N}\}$ has a time-varying state space representation defined as in (1.1), $\{\mathcal{A}_t(i)\}$, $\{\tilde{\boldsymbol{b}}_t\}$, and $\{\mathcal{A}_t\}$ are defined as in Assumption 2.1, and $\{\boldsymbol{A}_t(u)\}$ and $\{\boldsymbol{b}_t(u)\}$ satisfy the following assumptions.*

(i) *The process $\{\mathcal{X}_{t,N}\}$ satisfies Assumption 2.1 with $\beta = 1$.*

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

(ii) *Let $\beta' > 0$. For some $C \leq \infty$, the matrices $\{\boldsymbol{A}_t(\cdot)\}$ and vectors $\{\boldsymbol{b}_t(\cdot)\}$ satisfy*

$$|\boldsymbol{A}_t(u) - \boldsymbol{A}_t(v)|_{\text{abs}} \leq C|u - v|\mathcal{A}_t, \qquad |\boldsymbol{b}_t(u) - \boldsymbol{b}_t(v)|_{\text{abs}} \leq C|u - v|\tilde{\boldsymbol{b}}_t,$$
$$|\dot{\boldsymbol{A}}_t(u) - \dot{\boldsymbol{A}}_t(v)|_{\text{abs}} \leq C|u - v|^{\beta'}\mathcal{A}_t, \qquad |\dot{\boldsymbol{b}}_t(u) - \dot{\boldsymbol{b}}_t(v)|_{\text{abs}} \leq C|u - v|^{\beta'}\tilde{\boldsymbol{b}}_t,$$
$$\sup_u |\dot{\boldsymbol{A}}_t(u)|_{\text{abs}} \leq C\mathcal{A}_t(r), \quad \sup_u |\boldsymbol{A}_t(u)|_{\text{abs}} \leq C\mathcal{A}_t(r) \qquad \text{if } u \in \left[\frac{r-1}{M}, \frac{r}{M}\right),$$

and $\sup_u |\dot{\boldsymbol{b}}_t(u)|_{\text{abs}} \leq C\tilde{\boldsymbol{b}}_t$ *and* $\sup_u |\boldsymbol{b}_t(u)|_{\text{abs}} \leq C\tilde{\boldsymbol{b}}_t$, *where $C < \infty$. Therefore, $\{\boldsymbol{A}_t(u)\}$ and $\{\boldsymbol{b}_t(u)\}$ belong to the Lipschitz class* $\text{Lip}(1 + \beta')$. *This is a kind of Hölder continuity of order $1 + \beta'$ for random matrices.*

We now define the process $\{\dot{\mathcal{X}}_t(u)\}$, which we call the derivative process. By formally differentiating (1.2) with respect to u, we have

$$\dot{\mathcal{X}}_t(u) = \dot{\boldsymbol{A}}_t(u)\mathcal{X}_{t-1}(u) + \boldsymbol{A}_t(u)\dot{\mathcal{X}}_{t-1}(u) + \dot{\boldsymbol{b}}_t(u). \tag{3.1}$$

As we shall show below, an interesting aspect of the above difference differential equation is that its existence requires only weak assumptions on the derivative matrix $\dot{\boldsymbol{A}}_t(u)$. In other words, given that $\mathcal{X}_{t,N}$ is well defined, the existence of $\dot{\boldsymbol{A}}_t(u)$ is sufficient for the derivative process also to be well defined. This will become clear when we rewrite $\dot{\mathcal{X}}_t(u)$ as a state space model. Let $\mathcal{X}_t(2, u)^\top = (\dot{\mathcal{X}}_t(u)^\top, \mathcal{X}_t(u)^\top)$. It is then clear that $\{\mathcal{X}_t(2, u)\}$ has the representation

$$\mathcal{X}_t(2, u) = \boldsymbol{A}_t(2, u)\mathcal{X}_{t-1}(2, u) + \boldsymbol{b}_t(2, u), \tag{3.2}$$

where

$$\boldsymbol{A}_t(2, u) = \begin{pmatrix} \boldsymbol{A}_t(u) & \dfrac{\mathrm{d}\boldsymbol{A}_t(u)}{\mathrm{d}u} \\ \boldsymbol{0} & \boldsymbol{A}_t(u) \end{pmatrix}, \qquad \boldsymbol{b}_t(2, u) = \begin{pmatrix} \dot{\boldsymbol{b}}_t(u) \\ \boldsymbol{b}_t(u) \end{pmatrix}. \tag{3.3}$$

Motivated by the definition of the time-varying stationary process, we now define the time-varying derivative process. We call $\{\mathcal{X}_{t,N}(2)\}$ a time-varying derivative process if it satisfies

$$\mathcal{X}_{t,N}(2) = \boldsymbol{A}_t\left(2, \frac{t}{N}\right)\mathcal{X}_{t-1,N}(2) + \boldsymbol{b}_t\left(2, \frac{t}{N}\right), \tag{3.4}$$

with $\{\mathcal{X}_{t,N}(2)^\top\} = \{(\dot{\mathcal{X}}_{t,N}{}^\top, \mathcal{X}_{t,N}{}^\top)\}$. The main reason for defining this process is that it can be used to estimate the derivative process $\{\dot{\mathcal{X}}_t(u)\}$, which may not be observed and in practice may be difficult to estimate.

Let

$$\tilde{\boldsymbol{b}}_t(2) = \begin{pmatrix} C\tilde{\boldsymbol{b}}_t{}^\top \\ \tilde{\boldsymbol{b}}_t{}^\top \end{pmatrix}, \qquad \mathcal{A}_t(2, r) = \begin{pmatrix} \mathcal{A}_t(r) & C\mathcal{A}_t(r) \\ \boldsymbol{0} & \mathcal{A}_t(r) \end{pmatrix}, \qquad \mathcal{A}_t(2) = \begin{pmatrix} \mathcal{A}_t & C\mathcal{A}_t \\ \boldsymbol{0} & \mathcal{A}_t \end{pmatrix}, \tag{3.5}$$

where C is as defined in Assumption 3.1.

We now show that it is the triangular form of the transition matrix that allows the results of Section 2.2 to be directly applied to the derivative process. In order to do this, in the following lemma we show that if the sequence $\{\boldsymbol{A}_t(u)\}$ has a negative Lyapunov exponent, then the sequence $\{\boldsymbol{A}_t(2, u)\}$ also has a negative Lyapunov exponent.

Lemma 3.1. *Suppose that Assumption 3.1 is satisfied. Then, for $r = 1, \dots, M$, we have*

$$\inf\left\{\frac{1}{n}\,\mathrm{E}(\log \|\boldsymbol{A}_t(2,u)\cdots \boldsymbol{A}_{t-n+1}(2,u)\|_{\mathrm{spec}})\right\} < 0, \tag{3.6}$$

$$\inf\left\{\frac{1}{n}\,\mathrm{E}(\log \|\mathcal{A}_t(2,r)\cdots \mathcal{A}_{t-n+1}(2,r)\|_{\mathrm{spec}})\right\} < 0. \tag{3.7}$$

Proof. See Appendix A.

From the above result we can show that (3.2) and (3.4) have unique causal solutions similar to (2.1). The following theorem also follows from Lemma 3.1.

Theorem 3.1. *Suppose that Assumption 3.1 holds and let $\{\mathcal{X}_{t,N}(2)\}$ be as defined in (3.4). Then the process $\{\mathcal{X}_{t,N}(2)\}$ satisfies Assumption 2.1 with transition matrices $\{\boldsymbol{A}_t(2,u)\colon t \in \mathbb{Z},\ u \in (0,1]\}$ and innovations vectors $\{\boldsymbol{b}_t(2,u)\colon t \in \mathbb{Z},\ u \in (0,1]\}$.*

Proof. Under Assumption 3.1, we have $|\boldsymbol{A}_t(2,u)|_{\mathrm{abs}} \le \mathcal{A}_t(2,r)$ if $u \in ((r-1)/M, r/M]$, and $\sup_u |\boldsymbol{b}_t(2,u)|_{\mathrm{abs}} \le \tilde{\boldsymbol{b}}_t(2)$. Furthermore, according to Lemma 3.1, the random matrix sequences $\{\boldsymbol{A}_t(2,u)\}$ and $\{\mathcal{A}_t(2,r)\}$ have negative Lyapunov exponents. Hence, all the conditions of Assumption 2.1 are satisfied, and we have the result.

Now, with an additional weak assumption on the moments of $\{\mathcal{X}_{t,N}(2)\}$, Theorem 2.1 can also be applied to the processes $\{\mathcal{X}_{t,N}(2)\}$ and $\{\mathcal{X}_t(2,u)\}$. Let

$$\boldsymbol{Y}_t(2) = \sum_{k=1}^{\infty}\sum_{r=1}^{M}\prod_{j=0}^{k-1}\mathcal{A}_{t-j}(2,r)\tilde{\boldsymbol{b}}_{t-k}(2). \tag{3.8}$$

Corollary 3.1. *Suppose that Assumption 3.1 holds. Let $\mathcal{X}_{t,N}(2)$, $\mathcal{X}_t(2,u)$, $\mathcal{A}_t(2,r)$, $\mathcal{A}_t(2)$, $\tilde{\boldsymbol{b}}_t(2)$, $\boldsymbol{W}_t$, and $\boldsymbol{Y}_t(2)$ be as defined in (3.4), (3.2), (3.5), (2.8), and (3.8), respectively, and suppose that there exists an $\varepsilon > 0$ such that*

$$\sup_{t,N}\mathrm{E}(\|\mathcal{X}_{t,N}(2)\|_1^{\varepsilon}) < \infty, \qquad \mathrm{E}(\|\boldsymbol{Y}_t(2)\|_1^{\varepsilon}) < \infty.$$

Then

$$\begin{aligned}
\left|\mathcal{X}_{t,N}(2) - \mathcal{X}_t\left(2,\frac{t}{N}\right)\right|_{\mathrm{abs}} &\le \frac{1}{N^{\beta'}}\boldsymbol{V}_{t,N}(2),\\
|\mathcal{X}_t(2,u) - \mathcal{X}_t(2,w)|_{\mathrm{abs}} &\le |u-w|^{\beta'}\boldsymbol{W}_t(2),\\
|\mathcal{X}_{t,N}(2) - \mathcal{X}_t(2,u)|_{\mathrm{abs}} &\le \left|\frac{t}{N} - u\right|^{\beta'}\boldsymbol{W}_t(2) + \frac{1}{N^{\beta'}}\boldsymbol{V}_{t,N}(2),
\end{aligned} \tag{3.9}$$

and $\|\boldsymbol{W}_t\|_1 \le \|\boldsymbol{W}_t(2)\|_1$, where $\boldsymbol{V}_{t,N}(2)$ is similar to $\boldsymbol{V}_{t,N}$ defined in (2.7), but with $\mathcal{A}_{t-j}(i_1)$, $\mathcal{X}_{t-k-1,N}$, and $\mathcal{A}_{t-j}$ replaced by $\mathcal{A}_{t-j}(2,i_1)$, $\mathcal{X}_{t-k-1,N}(2)$, and $\mathcal{A}_t(2)$, respectively, and

$$\boldsymbol{W}_t(2) = C\sum_{r=1}^{M}\sum_{k=1}^{\infty}\prod_{j=0}^{k-1}\mathcal{A}_{t-j}(2,r_1)\{\mathcal{A}_{t-k}(2)|\boldsymbol{Y}_{t-k-1}(2)|_{\mathrm{abs}} + \tilde{\boldsymbol{b}}_{t-k}(2)\}. \tag{3.10}$$

Moreover, $\boldsymbol{V}_{t,N}(2)$ converges almost surely and the series $\{\boldsymbol{W}_t(2)\}$ is a well-defined stationary process.

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

Proof. Under Assumption 3.1 and according to Theorem 3.1, all the conditions in Theorem 2.1 are satisfied, giving the result.

Our object is to use the derivative process to obtain an exact expression for the difference in (2.6). To do this we first show that the derivative process is almost surely Hölder continuous.

Corollary 3.2. *Suppose that Assumption 2.1 holds. Let* $\{\mathcal{X}_t(2,u)\}$ *and* $\{\boldsymbol{W}_t(2)\}$ *be as defined in (3.2) and (3.10), respectively. Then*

$$\dot{\mathcal{X}}_t(u) = \sum_{k=0}^{\infty}\sum_{r=0}^{k-1}\left[\prod_{i=0}^{r-1}\boldsymbol{A}_{t-i}(u)\right]\dot{\boldsymbol{A}}_{t-r}(u)\left[\prod_{i=r+1}^{k-1}\boldsymbol{A}_{t-i}(u)\right]\boldsymbol{b}_{t-k}(u)+\sum_{k=1}^{\infty}\left[\prod_{i=0}^{k-1}\boldsymbol{A}_{t-i}(u)\right]\dot{\boldsymbol{b}}_{t-k}(u), \tag{3.11}$$

is almost surely the unique well-defined solution to (3.1). Furthermore,

$$\sup_{u,v}|\dot{\mathcal{X}}_t(u)-\dot{\mathcal{X}}_t(v)|_{\text{abs}} \le |u-v|^{\beta'}\boldsymbol{W}_t(2), \tag{3.12}$$

and almost surely all paths of $\mathcal{X}_t(u)$ *belong to the Lipschitz class* $\mathrm{Lip}(1+\beta')$.

Proof. By expanding (3.2) and using standard results of Brandt (1986), we can show that (3.11) holds. To show that the right-hand side of (3.11) is the derivative of $\mathcal{X}_t(u)$, we note that both $\|\mathcal{X}_t(u)\|_1$ and the sums of the absolute values of the three terms on the right-hand side of (3.11) are almost surely bounded. Thus, we can exchange the summation and derivative, and it immediately follows that (3.11) is the derivative of $\mathcal{X}_t(u)$.

Equation (3.12) follows immediately from Corollary 3.1, and by using (3.11) and (3.12) we have $\mathcal{X}_t(u,\omega) \in \mathrm{Lip}(1+\beta')$ for all $\omega \in \mathcal{N}^{\mathrm{c}}$, where $\mathcal{N}$ is a set of measure 0.

We now give a stochastic Taylor series expansion of $\{\mathcal{X}_{t,N}\}$ in terms of stationary processes.

Theorem 3.2. *Let* $\mathcal{X}_{t,N}(2)$ *and* $\mathcal{X}_t(2,u)$ *be as defined in (3.4) and (3.2), respectively. Suppose that the assumptions of Corollary 3.1 hold. Then*

$$\mathcal{X}_{t,N} = \mathcal{X}_t(u) + \left(\frac{t}{N}-u\right)\dot{\mathcal{X}}_t(u) + O_p\left(\left|\frac{t}{N}-u\right|^{\beta'+1}\mathbf{1}+\frac{\mathbf{1}}{N}\right) \tag{3.13}$$

$$= \mathcal{X}_t(u) + \left(\frac{t}{N}-u\right)\dot{\mathcal{X}}_{t,N} + O_p\left(\left|\frac{t}{N}-u\right|^{\beta'+1}\mathbf{1}+\frac{\mathbf{1}}{N}\right). \tag{3.14}$$

Proof. Let $\mathcal{N}_1$ be a set of zero measure such that $\mathcal{X}_{t,N}(\omega)$, $\boldsymbol{W}_t(2,\omega)$ (defined in Corollary 3.1), and $\boldsymbol{Y}_t(2,\omega)$ converge for all $\omega \in \mathcal{N}_1^{\mathrm{c}}$. Then, using Corollary 3.2, we have $\mathcal{X}_t(u,\omega) \in \mathrm{Lip}(1+\beta')$ if $\omega \in \mathcal{N}_1^{\mathrm{c}}$. By using (2.4), making a Taylor series expansion of $\mathcal{X}_t(t/N,\omega)$ about u, and using the mean value theorem, we obtain

$$\mathcal{X}_{t,N}(\omega) = \mathcal{X}_t(u,\omega) + \left(\frac{t}{N}-u\right)\dot{\mathcal{X}}_t(u,\omega) + \left(\left|\frac{t}{N}-u\right|^{1+\beta'}+\frac{1}{N}\right)\boldsymbol{R}_N(\omega),$$

where $\|\boldsymbol{R}_N(\omega)\|_1 \le \|\boldsymbol{V}_{t,N}(\omega)\|_1 + \|\boldsymbol{W}_t(2,\omega)\|_1$. Therefore, since $\mathrm{P}(\mathcal{N}_1^{\mathrm{c}}) = 1$ we obtain (3.13). We use (3.9) and repeat the method given above to prove (3.14).

Observe that (3.13) means the nonstationary process $\{\mathcal{X}_{t,N}\}$ can be written as a linear combination of stationary processes.

We now show that, under Assumption 3.1 and the conditions of Proposition 2.1, the moments of the derivative process are uniformly bounded.

Proposition 3.1. *Suppose that Assumption 3.1 holds. Let $\mathcal{X}_t(2, u)$, $\mathcal{X}_{t,N}(2)$, $\boldsymbol{V}_{t,N}(2)$, and $\boldsymbol{W}_t(2)$ be as defined in Corollary 3.1. Suppose, for each $r \in \{1, \dots, M\}$ and some $n \in [1, \infty)$, that $\mathrm{E}(\|\tilde{\boldsymbol{b}}_t\|_n^n) < \infty$, and, for some $\delta > 0$, that $\sup_{1 \le r \le M} \lambda_{\mathrm{spec}}([\mathcal{A}_t(r)]_n) < 1 - \delta$.*

Then the expectations $\mathrm{E}(\|\mathcal{X}_{t,N}(2)\|_n^n)$, $\mathrm{E}(\|\boldsymbol{V}_{t,N}(2)\|_n^n)$, $\mathrm{E}(\|\mathcal{X}_t(2, u)\|_n^n)$, $\mathrm{E}(\|\boldsymbol{Y}_t(2)\|_n^n)$, and $\mathrm{E}(\|\boldsymbol{W}_t(2)\|_n^n)$ are all uniformly bounded with respect to t, N, and u (as relevant).

Proof. Since, for each $r \in \{1, \dots, M\}$, $\mathcal{A}_t(2, n)$ is a block upper-triangular matrix, we observe that $\lambda_{\mathrm{spec}}([\mathcal{A}_t(2, r)]_n) = \lambda_{\mathrm{spec}}([\mathcal{A}_t(r)]_n)$. Therefore, the proof of Proposition 2.1 can be used to prove this result also.

We now return to the local averages example in (2.10) and obtain a tighter bound for the remainder $\mathcal{B}_{t_0,N}$. Using (3.13), we have

$$\frac{1}{2M+1} \sum_{k=-M}^{M} (\mathcal{X}_{t_0+k,N} - \mathcal{X}_{t_0+k}(u_0)) = \frac{1}{2M+1} \sum_{k=-M}^{M} \frac{k}{N} \dot{\mathcal{X}}_{t_0+k}(u_0) + \boldsymbol{R}_{t_0,N} \qquad (3.15)$$

where

$$\|\boldsymbol{R}_{t_0,N}\|_1 \le \frac{1}{2M+1} \sum_{k=-M}^{M} \left(\frac{|k|^{1+\beta'}}{N^{1+\beta'}} + \frac{1}{N} \right) (\|\boldsymbol{W}_{t_0+k}(2)\|_1 + \|\boldsymbol{V}_{t_0+k,N}\|_1).$$

It follows from (3.15) that the size of the remainder or bias due to nonstationarity depends on the magnitude of the derivative processes $\{\dot{\mathcal{X}}_t(u)\}$. Furthermore, if the conditions of Proposition 3.1 are satisfied with $n = 2$, then $\mathrm{E}(\|\mathcal{B}_{t_0,N}\|_2^2)^{1/2} = O(M/N + 1/N)$. However, we can reduce this bound by assuming that the derivative process satisfies some mixing conditions (note that conditions are given in Section 4 which guarantee that the derivative process be strongly mixing). Let us suppose that $\{\dot{\mathcal{X}}_t(u)\}_t$ is a short-memory process; then

$$\frac{1}{2M+1} \sum_{k=-M}^{M} \frac{k}{N} \dot{\mathcal{X}}_{t_0+k}(u_0) = \frac{\sqrt{M}}{N} O_p(1).$$

Therefore, if $\sqrt{M}/N \ll (M/N)^{1+\beta'}$, we have

$$\mathrm{E}(\|\mathcal{B}_{t_0,N}\|_2^2) = \mathrm{E}\left(\left\| \frac{1}{2M+1} \sum_{k=-M}^{M} [\mathcal{X}_{t_0+k,N} - \mathcal{X}_{t_0+k}(u_0)] \right\|_2^2 \right) \le O\left(\left[\frac{M}{N} \right]^{1+\beta'} + \frac{1}{N} \right)^2.$$

In addition, if the second derivatives $\{\ddot{A}_t(u)\}$ and $\{\ddot{\boldsymbol{b}}_t(u)\}$ were to exist, then the process $\{\ddot{\mathcal{X}}_t(u)\}$ could be defined in the same way as $\{\dot{\mathcal{X}}_t(u)\}$ and we would have

$$\frac{1}{2M+1} \sum_{k=-M}^{M} [\mathcal{X}_{t_0+k,N} - \mathcal{X}_{t_0+k}(u_0)] \approx \frac{1}{2M+1} \sum_{k=-M}^{M} \frac{k^2}{N} \ddot{\mathcal{X}}_{t_0+k}(u_0).$$

From the above we can see that the sum of second derivatives is the dominating term in the remainder $\mathcal{B}_{t_0,N}$. Therefore, using $\mathrm{E}(\|\mathcal{B}_{t_0,N}\|_2^2)$ and

$$\mathrm{var}\left[\frac{1}{2M+1} \sum_{k=-M}^{M} \mathcal{X}_{t+k}(u) \right],$$

we are able to evaluate the mean squared error of the local average and thus obtain the optimal segment length M.

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

4. Mixing properties of the derivative process

We now consider the mixing properties for the stationary derivative process $\{\mathcal{X}_t(2,u)\}$. To establish geometric mixing of $\{\mathcal{X}_t(2,u)\}$, we use Tweedie (1983, Theorem 4(ii)), which requires the ϕ-irreducibility of the derivative process. We here state a state space version of this theorem given in Basrak *et al.* (2002).

Lemma 4.1. (Basrak *et al.* (2002, Theorem 2.8 and Remark 2.9).) *Suppose that the matrices $\{\boldsymbol{A}_t\}$ and the vectors $\{\boldsymbol{b}_t\}$ are independent, identically distributed processes such that* $\mathrm{E}(\log \|\boldsymbol{A}_t\|_{\text{spec}}) < 0$, *and that there exists an $\varepsilon > 0$ with $\mathrm{E}(\|\boldsymbol{A}_t\|_{\text{spec}}^{\varepsilon}) < \infty$. If the process $\{\mathcal{X}_t\}$ satisfies $\mathcal{X}_t = \boldsymbol{A}_t\mathcal{X}_{t-1} + \boldsymbol{b}_t$ and is ϕ-irreducible, then it is geometrically ergodic and, hence, strongly mixing with a geometric rate.*

To show that $\{\mathcal{X}_t(2,u)\}$ is a geometrically ergodic process we require the following lemma.

Lemma 4.2. *Suppose that Assumption 2.1 holds and that $\boldsymbol{A}_t(2,u)$ is as defined in (3.3). Then*

$$\mathrm{E}(\|\boldsymbol{A}_t(2,u)\|_{\text{spec}}^{\varepsilon}) < \infty \tag{4.1}$$

and the sequence $\{\boldsymbol{A}_t(2,u)\}$ has a negative Lyapunov exponent.

Proof. Under Assumption 2.1 and according to (A.4) with $n = 1$, we have $\|\boldsymbol{A}_t(2,u)\|_{\text{spec}} \leq C\|\boldsymbol{A}_t(2,u)\|_{\text{spec}}$ and, thus, $\mathrm{E}(\|\boldsymbol{A}_t(2,u)\|_{\text{spec}}) \leq 2C'\,\mathrm{E}(\|\boldsymbol{A}_t(u)\|_{\text{spec}}) < \infty$, which gives (4.1). From (3.6) we see that $\{\boldsymbol{A}_t(2,u)\}$ has a negative Lyapunov exponent.

We now use this lemma to prove the strong mixing with geometric rate of the stationary derivative process.

Theorem 4.1. *Suppose that Assumption 3.1 holds (with $d = 1$) and let the process $\{\mathcal{X}_t(2,u)\}$ defined in (3.2) be ϕ-irreducible. Then $\{\mathcal{X}_t(2,u)\}$ is geometrically ergodic and, thus, strongly mixing with a geometric rate.*

Proof. We first show that the conditions of Lemma 4.1 are satisfied; the result then follows. According to Lemma 4.2, there exist an $m > 0$ and a $\delta < 0$ such that

$$\frac{1}{m}\,\mathrm{E}(\log \|\boldsymbol{A}_t(2,u)\cdots \boldsymbol{A}_{t-m+1}(2,u)\|_{\text{spec}}) \leq \delta.$$

We iterate $\mathcal{X}_t(2,u)$ m times and define the mth iterate process $\{\mathcal{X}_{m,t}(2,u)\}$, where $\mathcal{X}_{m,t}(2,u) = \mathcal{X}_{mt}(2,u)$ and

$$\mathcal{X}_{m,t}(2,u) = \boldsymbol{C}_{m,t}(2,u)\mathcal{X}_{m,t-1} + \boldsymbol{d}_{m,t}(2,u),$$

where $\boldsymbol{C}_{m,t}(2,u) = \boldsymbol{A}_{mt}(2,u)\cdots \boldsymbol{A}_{m(t-1)+1}(2,u)$ and

$$\boldsymbol{d}_{m,t}(2,u) = \sum_{k=1}^{m-1} \boldsymbol{A}_{mt}(2,u)\cdots \boldsymbol{A}_{mt-k}(2,u)\boldsymbol{b}_{t-k}(2,u) + \boldsymbol{b}_t(2,u).$$

From (4.1), we see that $\mathrm{E}(\|\boldsymbol{C}_{m,t}(2,u)\|_{\text{spec}}) < \infty$.

From the above it is clear that $\{\mathcal{X}_{m,t}(2,u)\}$ satisfies the conditions of Lemma 4.1 and is therefore geometrically ergodic. It follows that $\{\mathcal{X}_t(2,u)\}$ is also geometrically ergodic.

A process which is strongly mixing with a geometric rate has many interesting properties. We now state one such property.

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

Corollary 4.1. *Let $\{\mathcal{X}_t(2,u)\}$ be as defined in (3.2). Suppose that Assumption 3.1 holds (with $d=1$), that $\{\mathcal{X}_t(2,u)\}$ is ϕ-irreducible, and that* $\mathrm{E}(\|\mathcal{X}_t(2,u)\|^2)<\infty$. *Then we have*

$$\sum_{k=0}^{\infty} |\operatorname{cov}(\mathcal{X}_t(2,u)_i, \mathcal{X}_{t+k}(2,u)_i)| < \infty \quad \textit{for } i=1,\ldots,2p,$$

where $\mathcal{X}_t(2,u)_i$ denotes the ith element of the vector $\mathcal{X}_t(2,u)$.

Proof. From Theorem 4.1, we see that $\{\mathcal{X}_t(2,u)\}$ is geometrically ergodic; therefore, by using Davidson (1994, Corollary 14.3), we have the result.

It follows from Corollary 4.1 that $\{\mathcal{X}_t(u)\}$ and $\{\dot{\mathcal{X}}_t(u)\}$ are short-memory processes.

5. An example: the time-varying GARCH process

In this section we show that the time-varying GARCH (tvGARCH) process admits the representation (1.1), and that the results in the previous section apply to it. We mention both that the results below also apply to the time-varying ARCH process, as it is a special case of the tvGARCH process, and that the conditions stated here are slightly more general than the conditions given in Dahlhaus and Subba Rao (2006). Let $\mathcal{I}_{p\times q}$ denote a $(p\times q)$-dimensional matrix with $(\mathcal{I}_{p\times q})_{i,j}=1$ for all i and j, and let $\boldsymbol{I}_p$ denote the $(p\times p)$-dimensional identity matrix.

We first note that the stochastic process $\{X_{t,N}\}$ is called a tvGARCH(p,q) process if it satisfies

$$X_{t,N} = Z_t\sigma_{t,N},$$
$$\sigma_{t,N}^2 = a_0\left(\frac{t}{N}\right) + \sum_{i=1}^{p} a_i\left(\frac{t}{N}\right)X_{t-i,N}^2 + \sum_{j=1}^{q} b_j\left(\frac{t}{N}\right)\sigma_{t-j,N}^2, \qquad t=1,\ldots,N,$$

where $\{Z_t\}$ are independent, identically distributed random variables with $\mathrm{E}(Z_t)=0$ and $\mathrm{E}(Z_t^2)=1$, and a_i, $0\le i\le p$, and b_j, $1\le j\le q$, are functions from $[0,1]$ to $\mathbb{R}$. It is straightforward to show that the tvGARCH process $\{X_{t,N}^2\}$ admits the state space representation (1.1) with

$$\mathcal{X}_{t,N}{}^{\top} = (\sigma_{t,N}^2,\ldots,\sigma_{t-q+1,N}^2, X_{t-1,N}^2,\ldots,X_{t-p+1,N}^2),$$
$$\boldsymbol{b}_t(u)^{\top} = (a_0(u),0,\ldots,0)\in\mathbb{R}^{p+q-2},$$

and

$$\boldsymbol{A}_t(u) = \begin{pmatrix} \boldsymbol{\tau}_t(u) & b_q(u) & \boldsymbol{a}(u) & a_p(u) \\ \boldsymbol{I}_{q-1} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \boldsymbol{Z}_{t-1}^2 & 0 & \mathbf{0} & 0 \\ \mathbf{0} & \mathbf{0} & \boldsymbol{I}_{p-2} & \mathbf{0} \end{pmatrix},$$

a $(p+q-1)\times(p+q-1)$ matrix where $\boldsymbol{\tau}_t(u)=(b_1(u)+a_1(u)Z_{t-1}^2, b_2(u),\ldots,b_{q-1}(u))$, $\boldsymbol{a}(u)=(a_2(u),\ldots,a_{p-1}(u))$, and $\boldsymbol{Z}_{t-1}^2=(Z_{t-1}^2,0,\ldots,0)\in\mathbb{R}^{q-1}$ (we assume without loss of generality that $p,q\ge 2$).

5.1. The tvGARCH process and Assumption 2.1

Let us consider the tvGARCH(p, q) process. We will show that if $\mathrm{E}(Z_t^2) = 1$ and the parameters $\{a_i(\cdot)\}$ and $\{b_j(\cdot)\}$ are β-Lipschitz continuous (that is, $|a_i(u) - a_i(v)| \le K|u - v|^\beta$ and $|b_j(u) - b_j(v)| \le K|u - v|^\beta$, where K is a finite constant) and satisfy

$$\sup_u \left\{ \sum_{i=1}^{p} a_i(u) + \sum_{j=1}^{q} b_j(u) \right\} < 1 - \eta, \tag{5.1}$$

then Assumption 2.1 holds for the tvGARCH(p, q) process.

Using the β-Lipschitz continuity of the parameters, we first show that there exist matrices $\mathcal{A}_t(r)$ which bound $\boldsymbol{A}_t(u)$ and satisfy Assumption 2.1(i). Let $K_{\max}$ be such that $\sup_{u,v} |a_i(u) - a_i(v)| \le K_{\max}|u - v|^\beta$ and $\sup_{u,v} |b_j(u) - b_j(v)| \le K_{\max}|u - v|^\beta$. Define an ε such that $\varepsilon \le \{\eta/(2K_{\max}(p + q))\}^{1/\beta}$ and $\varepsilon^{-1} \in \mathbb{N}$. Let $M(\varepsilon) = \varepsilon^{-1}$ and, for each $r \in \{1, \dots, M(\varepsilon)\}$, $i = 1, \dots, p$, and $j = 1 \dots, q$, define

$$\alpha_i(r) = \{a_i((k - 1)\varepsilon) + K_{\max}\varepsilon^\beta\}, \qquad \beta_j(r) = \{b_j((k - 1)\varepsilon) + K_{\max}\varepsilon^\beta\}. \tag{5.2}$$

Therefore, from (5.1) and the above construction, we have $\sup_{(r-1)\varepsilon \le u < r\varepsilon} a_i(u) \le \alpha_i(r)$, $\sup_{(r-1)\varepsilon \le u < r\varepsilon} b_j(u) \le \beta_j(r)$, and

$$\sum_{i=1}^{p} \sup_{(r-1)\varepsilon \le u < r\varepsilon} a_i(u) + \sum_{j=1}^{q} \sup_{(r-1)\varepsilon \le u < r\varepsilon} b_j(u) \le \sum_{i=1}^{p} \alpha_i(r) + \sum_{j=1}^{q} \beta_j(r) \le 1 - \frac{\eta}{2}. \tag{5.3}$$

Let

$$\mathcal{A}_t(r) = \begin{pmatrix} \tilde{\boldsymbol{\tau}}_t(r) & \beta_q(r) & \boldsymbol{\alpha}(r) & \alpha_p(r) \\ \boldsymbol{I}_{q-1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{Z}_{t-1}^2 & 0 & \boldsymbol{0} & 0 \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{p-2} & \boldsymbol{0} \end{pmatrix}, \tag{5.4}$$

where $\tilde{\boldsymbol{\tau}}_t(r) = (\beta_1(r) + \alpha_1(r)Z_{t-1}^2, \beta_2(r), \dots, \beta_{q-1}(r))$, $\boldsymbol{\alpha}(r) = (\alpha_2(r), \dots, \alpha_{p-1}(r))$, and (as above) $\boldsymbol{Z}_{t-1}^2 = (Z_{t-1}^2, 0, \dots, 0) \in \mathbb{R}^{p-1}$. Then it is clear that $\sup_{(r-1)\varepsilon \le u < r\varepsilon} |\boldsymbol{A}_t(u)|_{\text{abs}} \le \mathcal{A}_t(r)$. To summarise, we have partitioned the unit interval into $M(\varepsilon)$ intervals, such that all the matrices $\boldsymbol{A}_t(u)$ in a given interval, say $[(r - 1)\varepsilon, r\varepsilon)$, are bounded above by the matrix $\mathcal{A}_t(r)$. It is clear that, for each r, $\{\mathcal{A}_t(r)\}$ is an independent, identically distributed sequence of random matrices. Since $\mathrm{E}(Z_0^2) = 1$ and $\sum_{i=1}^{p} \alpha_i(r) + \sum_{j=1}^{q} \beta_j(r) < 1 - \eta/2$, it follows from Lemma 5.1 that $\lambda_{\text{spec}}(\mathrm{E}(\mathcal{A}_t(r))) \le (1 - \eta/2)^{1/(p+q-1)}$. By using Kesten and Spitzer (1984, Equation (1.4)), we can show that the sequence $\{\boldsymbol{A}_t(r)\}$ has a negative Lyapunov exponent.

Let $\tilde{\boldsymbol{b}}_t{}^\top = (\sup_u a_0(u), 0, \dots, 0) \in \mathbb{R}^{p+q-2}$. Since $\mathrm{E}(Z_t^2) = 1$, for some $K > 0$ we have $\mathrm{E}(\log \|\mathcal{A}_t(r)\|_{\text{spec}}) \le \mathrm{E}(\|\mathcal{A}_t(r)\|_{\text{spec}}) \le K\,\mathrm{E}(\|\mathcal{A}_t(r)\|_1) < \infty$. Therefore, all the conditions of Assumption 2.1(i) are satisfied.

Finally, it is clear that there exists a constant K such that

$$|\boldsymbol{A}_t(u) - \boldsymbol{A}_t(v)|_{\text{abs}} \le K|u - v|^\beta \mathcal{A}_t, \qquad |\boldsymbol{b}_t(u) - \boldsymbol{b}_t(v)|_{\text{abs}} \le K|u - v|^\beta \tilde{\boldsymbol{b}}_t,$$

where $\mathcal{A}_t = (1 + Z_{t-1}^2)\mathcal{I}_{(p+q-1)\times(p+q-1)}$. Thus, Assumption 2.1(ii) is also satisfied.

5.2. The tvGARCH process and the stationary approximation

We now define the stationary GARCH process $\{X_t(u)\}$ which has the representation

$$X_t(u)^2 = \left\{a_0(u) + \sum_{i=1}^{p} a_i(u) X_{t-i}(u)^2 + \sum_{j=1}^{q} b_j(u)\sigma_{t-j}(u)^2\right\} Z_t^2. \tag{5.5}$$

In order to show that $X_t(u)^2$ locally approximates $X_{t,N}^2$, we need to verify the conditions of Theorem 2.1. We have shown above that Assumption 2.1 is satisfied; hence, we now only need to show the existence of the moments $\mathrm{E}(\|\mathcal{X}_{t,N}\|_1^{\varepsilon})$ and $\mathrm{E}(\|\boldsymbol{Y}_t\|_1^{\varepsilon})$. We do this by verifying the conditions of Proposition 2.1 for $\varepsilon = n$ (although it is enough to prove the result for $\varepsilon = 1$). Suppose that $n \in [1, \infty)$ and let $\mu_n = \{\mathrm{E}(Z_t^{2n})\}^{1/n}$. In addition, we assume that

$$\mu_n \sup_u \left\{\sum_{i=1}^{p} a_i(u) + \sum_{j=1}^{q} b_j(u)\right\} < 1 - \eta \tag{5.6}$$

for some $\eta > 0$. Using a similar construction to the above, we now construct matrices $\mathcal{A}_t(r)$ for which $\boldsymbol{A}_t(u) \leq \mathcal{A}_t(r)$ for $(r-1)/M(\varepsilon) \leq u < r/M(\varepsilon)$ and such that $\lambda_{\text{spec}}([\mathcal{A}_t(r)]_n) \leq (1-\eta/2)^{1/(p+q-1)}$, thus satisfying the conditions of Proposition 2.1. We then let $\varepsilon = (\eta/[2\mu_n K_{\max}(p+q)])^{1/\beta}$ and, observing the inequalities in (5.3), define $M(\varepsilon)$, $\alpha_i(r)$, and $\beta_j(r)$ (as in (5.2)), and $\mathcal{A}_t(r)$ (as in (5.4)) using the new ε. It is straightforward to show that

$$\sup_u \mu_n \left\{\sum_{i=1}^{p} \alpha_i(r) + \sum_{j=1}^{q} \beta_j(r)\right\} < 1 - \frac{\eta}{2}. \tag{5.7}$$

To show that $\lambda_{\text{spec}}([\mathcal{A}_t(r)]_n) \leq (1-\eta/2)^{1/(p+q-1)}$, we will use the following result, which is an adaptation of Bougerol and Picard (1992a, Corollary 2.2), where the result was proved for $\mu = 1$.

Lemma 5.1. *Let $\mu > 1$, let $\{a_i : i = 1, \dots, p\}$ and $\{b_j : j = 1 \dots, q\}$ be positive sequences, and let*

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{\tau} & b_q & \boldsymbol{a} & a_p \\ \boldsymbol{I}_{q-1} & \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{\mu} & 0 & \boldsymbol{0} & 0 \\ \boldsymbol{0} & \boldsymbol{0} & \boldsymbol{I}_{p-2} & \boldsymbol{0} \end{pmatrix}, \tag{5.8}$$

where $\boldsymbol{\tau} = (b_1 + a_1\mu, b_2, \dots, b_{q-1}) \in \mathbb{R}^{q-1}$, $\boldsymbol{a} = (a_2, \dots, a_{p-1}) \in \mathbb{R}^{p-2}$, and $\boldsymbol{\mu} = (\mu, 0, \dots, 0) \in \mathbb{R}^{q-1}$. Suppose that

$$\mu\left(\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j\right) < 1 - \delta,$$

where $p, q \geq 2$ and $\delta > 0$. Then $\lambda_{\text{spec}}(\boldsymbol{A}) \leq (1-\delta)^{1/(p+q-1)}$.

Now we construct the matrix $\mathcal{A}(r)^*$, which is the same as $\boldsymbol{A}$ in (5.8) with a_i, b_i, and μ replaced by $\alpha_i(r)$, $\beta_i(r)$, and μ_n, respectively. It is clear that $[\mathcal{A}(r)]_n \leq \mathcal{A}(r)^*$. Now, from Lemma 5.1 and (5.7), we then have $\lambda_{\text{spec}}(\mathcal{A}(r)^*) \leq (1-\eta/2)^{1/(p+q-1)}$. Thus, the conditions of Proposition 2.1 are satisfied and we have $\sup_{t,N} \mathrm{E}(\|\mathcal{X}_{t,N}\|_n^n) < \infty$ (which implies that $\sup_{t,N} \mathrm{E}(X_{t,N}^{2n}) < \infty$) and $\mathrm{E}(\|\boldsymbol{Y}_t\|_n^n) < \infty$.

Therefore, if (5.6) holds for some $n \geq 1$, the conditions of Theorem 2.1 are fulfilled and we have

$$X_{t,N}^2 = X_t(u)^2 + \left(\left|\frac{t}{N} - u\right|^{\beta} + \frac{1}{N^{\beta}}\right) R_{t,N}, \quad \text{where} \quad \sup_{t,N} \mathrm{E}(R_{t,N}^n) < \infty.$$

5.3. The GARCH and derivative processes

We now consider the stationary derivative process associated with the tvGARCH process. Formally differentiating (5.5) gives

$$\frac{\mathrm{d}X_t(u)^2}{\mathrm{d}u} = \frac{\mathrm{d}a_0(u)}{\mathrm{d}u} + \sum_{i=1}^{p} \left\{ \frac{\mathrm{d}a_i(u)}{\mathrm{d}u} X_{t-i}(u)^2 + a_i(u) \frac{\mathrm{d}X_{t-i}(u)^2}{\mathrm{d}u} \right\}$$
$$+ \sum_{j=1}^{q} \left\{ \frac{\mathrm{d}b_j(u)}{\mathrm{d}u} \sigma_{t-j}(u)^2 + b_j(u) \frac{\mathrm{d}\sigma_{t-j}(u)^2}{\mathrm{d}u} \right\},$$

which was shown in Section 3 to admit a state space representation. By applying Theorem 3.2, we obtain the Taylor series expansion

$$X_{t,N}^2 = X_t(u)^2 + \left(\frac{t}{N} - u\right) \frac{\mathrm{d}X_t(u)^2}{\mathrm{d}u} + \left(\left|\frac{t}{N} - u\right|^{1+\beta'} + \frac{1}{N}\right).$$

Finally, if (5.6) holds then the conditions of Proposition 3.1 are satisfied and

$$\mathrm{E}\left(\left(\frac{\mathrm{d}X_t(u)^2}{\mathrm{d}u}\right)^n\right) < \infty.$$

6. Applications

The notion of stationary approximations and the derivative process can fruitfully be used in many applications. The key is the representation, (3.13), of the nonstationary process in terms of stationary processes. As indicated by the local averages example, by using this representation classical results for stationary processes such as the ergodic theorem or central limit theorems can (more or less easily) be used in the theoretical investigations of nonstationary processes. An example was given in Dahlhaus and Subba Rao (2006, Theorem 3), where the properties of a local likelihood estimator were investigated. The results of that paper can be used to derive similar results for the models used as examples in the present paper (among others). The derivative process in (3.13) then typically leads to bias terms due to the nonstationarity of the process. Another application for the results in this paper is recursive online estimation for such models. Problems of this type will be considered in future work.

Appendix A.

In this appendix we sketch some of the proofs of the results stated earlier. Full details can be found in the technical report available from the author (and online at http://www.stat.tamu.edu/~suhasini/tvstate-space.ps).

Most the results in this paper are based on the following theorem, which is a nonstationary version of Brandt (1986, Theorem 1) and Bougerol and Picard (1992b, Theorem 2.5). The proof is similar to that of the latter, so we omit the details.

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

Lemma A.1. *Suppose that* $\{\mathcal{A}_t(i)\colon i = 1, \dots, M\}$ *satisfy Assumption 2.1(i). Let* $\{\boldsymbol{d}_t\}$ *be a random sequence which satisfies* $\sup_t \mathrm{E}(\|\boldsymbol{d}_t\|^{\varepsilon}) < \infty$ *for some* $\varepsilon > 0$*, let the sequence* $\{n_r\}$ *be such that* $n_0 \le n_1 \le \cdots \le n_M$*, let* $(s, t]$ *denote the integer sequence* $\{s+1, s+2, \dots, t\}$*, and let* $J^t_{r,k} = [t-k, t] \cap [n_{r-1}, n_r]$*. Then, for any* γ*,* $0 \le \gamma \le 1$*,*

$$\boldsymbol{Y}_t = \sum_{k\ge 1} k^{\gamma} \prod_{r=1}^{M} \prod_{i \in J^t_{r,k}} \mathcal{A}_{t-i}(r)\boldsymbol{d}_{t-k} \tag{A.1}$$

converges almost surely.

We use the lemma above to prove Theorem 2.1.

Proof of Theorem 2.1. By the triangle inequality, we have

$$|\mathcal{X}_{t,N} - \mathcal{X}_t(u)|_{\mathrm{abs}} \le \left|\mathcal{X}_{t,N} - \mathcal{X}_t\left(\frac{t}{N}\right)\right|_{\mathrm{abs}} + \left|\mathcal{X}_t\left(\frac{t}{N}\right) - \mathcal{X}_t(u)\right|_{\mathrm{abs}}.$$

We first derive a bound for $|\mathcal{X}_{t,N} - \mathcal{X}_t(t/N)|_{\mathrm{abs}}$. By expanding $\mathcal{X}_{t,N}$ and $\mathcal{X}_t(t/N)$, under Assumption 2.1 we have

$$\begin{aligned}
\left|\mathcal{X}_{t,N} - \mathcal{X}_t\left(\frac{t}{N}\right)\right|_{\mathrm{abs}} &= \left|\boldsymbol{A}_t\left(\frac{t}{N}\right)\left\{\mathcal{X}_{t-1,N} - \mathcal{X}_{t-1}\left(\frac{t}{N}\right)\right\}\right|_{\mathrm{abs}} \\
&= \left|\boldsymbol{A}_t\left(\frac{t}{N}\right)\left\{\boldsymbol{A}_{t-1}\left(\frac{t-1}{N}\right) - \boldsymbol{A}_{t-1}\left(\frac{t}{N}\right)\right\}\mathcal{X}_{t-2,N}\right. \\
&\qquad + \boldsymbol{A}_{t-1}\left(\frac{t}{N}\right)\left\{\mathcal{X}_{t-2,N} - \mathcal{X}_{t-2}\left(\frac{t}{N}\right)\right\} \\
&\qquad \left. + \boldsymbol{A}_t\left(\frac{t}{N}\right)\left\{\boldsymbol{b}_{t-1}\left(\frac{t-1}{N}\right) - \boldsymbol{b}_{t-1}\left(\frac{t}{N}\right)\right\}\right|_{\mathrm{abs}} \\
&\le \frac{1}{N^{\beta}}\mathcal{A}_t(i_1)\boldsymbol{b}_{t-1} + \frac{1}{N^{\beta}}\mathcal{A}_t(i_1)\mathcal{A}_{t-1}|\mathcal{X}_{t-2,N}|_{\mathrm{abs}} \\
&\qquad + \mathcal{A}_{t-1}(i_1)\left|\mathcal{X}_{t-2,N} - \mathcal{X}_{t-2}\left(\frac{t}{N}\right)\right|_{\mathrm{abs}}.
\end{aligned}$$

Now, by continuing the iteration above we obtain $|\mathcal{X}_{t,N} - \mathcal{X}_t(t/N)|_{\mathrm{abs}} \le (1/N^{\beta})\boldsymbol{V}_{t,N}$, where $\boldsymbol{V}_{t,N}$ is as defined in (2.7). Under Assumption 2.1 and from (2.3), we have

$$\mathrm{E}(\log\|\mathcal{A}_t\mathcal{X}_{t-1,N}\|) < \infty.$$

Therefore, by using (A.1) with $\boldsymbol{d}_t = \mathcal{A}_t\mathcal{X}_{t-1,N} + \tilde{\boldsymbol{b}}_t$, we see that $\boldsymbol{V}_{t,N}$ converges almost surely.

Using a similar method to the above, we can show that

$$\|\mathcal{X}_t(u) - \mathcal{X}_t(w)\|_1 \le |u-w|^{\beta}\boldsymbol{W}_t,$$

where $\boldsymbol{W}_t$ is as defined in (2.8). From Lemma A.1, we see that $\{\boldsymbol{W}_t\}$ converges almost surely. Finally, (2.6) follows from (2.4) and (2.5).

Below we make frequent use of the following inequalities (which can proved by repeated use of the Minkowski inequality). Suppose that $\boldsymbol{A}$ and $\{\boldsymbol{A}_t\}$ are $(p \times p)$-dimensional independent

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

random matrices and $\mathcal{X}$ a p-dimensional random vector independent of $\{\boldsymbol{A}_t\}$. Then

$$\begin{aligned} \mathrm{E}(\|\boldsymbol{A}\mathcal{X}\|_n^n)^{1/n} &\leq K\|[\boldsymbol{A}]_n\|_{\mathrm{spec}}\,\mathrm{E}(\|\mathcal{X}\|_n^n)^{1/n}, \\ \|[\boldsymbol{A}_1\cdots\boldsymbol{A}_n]_n\|_{\mathrm{spec}} &\leq K\|[\boldsymbol{A}_1]_n\cdots[\boldsymbol{A}_m]_n\|_{\mathrm{spec}}, \end{aligned} \tag{A.2}$$

for some finite constant K.

Proof of Proposition 2.1. We now show that $\mathrm{E}(\|\mathcal{X}_{t,N}\|_n^n)$ is uniformly bounded over t and N. Since $\{\mathcal{A}_t(i)\}$ and $\{\tilde{\boldsymbol{b}}_t\}$ are independent, from (A.2) we have

$$\mathrm{E}\left(\left\|\prod_{i=0}^{k-1}\mathcal{A}_{t-i}\left(\frac{t-i}{N}\right)\tilde{\boldsymbol{b}}_{t-k}\right\|_n^n\right)^{1/n} \leq K\left\|\prod_{i=0}^{k-1}\left[\mathcal{A}_{t-i}\left(\frac{t-i}{N}\right)\right]_n\right\|_{\mathrm{spec}}\mathrm{E}(\|\tilde{\boldsymbol{b}}_{t-k}\|_n^n)^{1/n}.$$

From the above and

$$|\mathcal{X}_{t,N}|_{\mathrm{abs}} \leq \sum_{k=0}^{\infty}\prod_{i=0}^{k-1}\left|\boldsymbol{A}_{t-i}\left(\frac{t-i}{N}\right)\right|_{\mathrm{abs}}\tilde{\boldsymbol{b}}_{t-k} \leq \sum_{k=0}^{\infty}\prod_{r=1}^{M}\prod_{i\in J_{r,k}}\mathcal{A}_{t-i}(r)\tilde{\boldsymbol{b}}_{t-k},$$

where

$$J_{r,l} = \left\{k \geq 0\colon \frac{t-k}{N} \in \left[\frac{r-1}{M}, \frac{r}{M}\right)\right\} \cap \{0, 1, \ldots, l-1\},$$

we have

$$\begin{aligned} \mathrm{E}(\|\mathcal{X}_{t,N}\|_n^n)^{1/n} &\leq \sum_{k=0}^{\infty}\mathrm{E}\left(\left\|\prod_{r=1}^{M}\prod_{i\in J_{r,k}}\mathcal{A}_{t-i}(r)\tilde{\boldsymbol{b}}_{t-k}\right\|_n^n\right)^{1/n} \\ &\leq \sum_{k=0}^{\infty}\prod_{r=1}^{M}\left\|\left[\prod_{i\in J_{r,k}}\mathcal{A}_{t-i}(r)\right]_n\right\|_{\mathrm{spec}}\mathrm{E}(\|\tilde{\boldsymbol{b}}_t\|_n^n)^{1/n} \\ &\leq \sum_{k=0}^{\infty}\prod_{r=1}^{M}\|[\mathcal{A}_0(r)]_n^{\#(J_{r,k})}\|_{\mathrm{spec}}\,\mathrm{E}(\|\tilde{\boldsymbol{b}}_t\|_n)^{1/n}, \end{aligned} \tag{A.3}$$

where $\#(J_{r,k})$ denotes the cardinality of the set $J_{r,k}$. Since $\lambda_{\mathrm{spec}}([\boldsymbol{A}_0(i)]_n) \leq 1-\delta$ for $i = 1, \ldots, M$, and as a result of Moulines *et al.* (2005, Lemma 12), there exists a K independent of $\mathcal{A}_0(r)$ and m such that $\|[\mathcal{A}_0(r)]_n^m\|_{\mathrm{spec}} \leq K(1-\delta/2)^m$. Therefore,

$$\|[\mathcal{A}_0(r)]_n^{\#(J_{r,k})}\|_{\mathrm{spec}} \leq K(1-\delta/2)^{\#(J_{r,k})}.$$

By substituting the above into (A.3) and using $\sum_{r=1}^{M}\#(J_{r,k}) = k$, we can show that

$$\sup_{t,N}\mathrm{E}(\|\mathcal{X}_{t,N}\|_n^n) < \infty.$$

Using a similar method we can prove the other inequalities in (2.11).

Proof of Lemma 3.1. We now prove (3.6). Under Assumption 3.1, it is straightforward to show that

$$|\boldsymbol{A}_t(2,u)\cdots\boldsymbol{A}_{t-n+1}(2,u)|_{\mathrm{abs}} \leq \boldsymbol{B}_k(t,n),$$

https://doi.org/10.1017/S000186780000149X Published online by Cambridge University Press

where

$$\boldsymbol{B}_k(t,n) = \begin{pmatrix} \mathcal{A}_t(r)\cdots\mathcal{A}_{t-n+1}(r) & Cn\mathcal{A}_t(r)\cdots\mathcal{A}_{t-n+1}(r) \\ \mathbf{0} & \mathcal{A}_t(r)\cdots\mathcal{A}_{t-n+1}(r) \end{pmatrix}$$

and $u \in [(k-1)/M, k/M)$. From this, we have

$$\boldsymbol{B}_k(t,n)\boldsymbol{B}_k(t,n)^\top = \begin{pmatrix} (Cn^2+1)\boldsymbol{R}_k(t,n) & Cn\boldsymbol{R}_k(t,n) \\ Cn\boldsymbol{R}_k(t,n) & \boldsymbol{R}_k(t,n) \end{pmatrix},$$

where

$$\boldsymbol{R}_k(t,n) = (\mathcal{A}_t(r)\cdots\mathcal{A}_{t-n+1}(r))(\mathcal{A}_t(r)\cdots\mathcal{A}_{t-n+1}(r))^\top.$$

By choosing a C_1 such that $C_1(n^2+1) \ge (Cn^2+1)$ and $C_1(n^2+1) \ge Cn$ for all n, we obtain

$$\boldsymbol{B}_k(t,n)\boldsymbol{B}_k(t,n)^\top \le C_1(n^2+1)\begin{pmatrix} \boldsymbol{R}_k(t,n) & \boldsymbol{R}_k(t,n) \\ \boldsymbol{R}_k(t,n) & \boldsymbol{R}_k(t,n) \end{pmatrix}.$$

It is clear that the largest eigenvalue of this matrix is $C_1(n^2+1)\|\mathcal{A}_t(r)\cdots\mathcal{A}_{t-n+1}(r)\|_{\text{spec}}$. Therefore, we have

$$\begin{aligned} \|\boldsymbol{A}_t(2,u)\cdots\boldsymbol{A}_{t-n+1}(2,u)\|_{\text{spec}} &\le \|\boldsymbol{B}_k(t,n)\|_{\text{spec}} \\ &\le C_1(n^2+1)\|\mathcal{A}_t(r)\cdots\mathcal{A}_{t-n+1}(r)\|_{\text{spec}}. \end{aligned} \tag{A.4}$$

It immediately follows that the sequence $\{\boldsymbol{A}_t(2,u)\}$ has a negative Lyapunov exponent. The proof of (3.7) is similar, so we omit this proof.

Acknowledgements

The author would like to thank Professor Rainer Dahlhaus and an anonymous referee for making many extremely interesting suggestions and improvements. This research was supported by the Deutsche Forschungsgemeinschaft (DA 187/12-2).

References

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

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics* **31,** 307–327.

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

Bougerol, P. and Picard, N. (1992b). Strict stationarity of generalized autoregressive processes. *Ann. Prob.* **20,** 1714–1730.

Brandt, A. (1986). The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients. *Adv. Appl. Prob.* **18,** 211–220.

Cramér, H. (1961). On some classes of nonstationary stochastic processes. In *Proc. 4th Berkeley Symp. Math. Statist. Prob.*, Vol. II, ed. J. Neyman, University of California Press, Berkeley, pp. 57–78.

Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **16,** 1–37.

Dahlhaus, R. and Subba Rao, S. (2006). Statistical inference for time-varying ARCH processes. *Ann. Statist.* **34,** 1075–1114.

Davidson, J. (1994). *Stochastic Limit Theory*. Clarendon Press, New York.

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

Kesten, H. and Spitzer, F. (1984). Convergence in distribution for products of random matrices. *Z. Wahrscheinlichkeitsth.* **67,** 363–386.

Moulines, E., Priouret, P. and Roueff, F. (2005). On recursive estimation for time varying autoregressive processes. *Ann. Statist.* **33,** 2610–2654.

NICHOLLS, D. F. AND QUINN, B. G. (1982). *Random Coefficient Autoregressive Models: an Introduction* (Lecture Notes Statist. **11**). Springer, New York.

PRIESTLEY, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. R. Statist. Soc. B* **27,** 204–237.

STRAUMANN, D. AND MIKOSCH, T. (2006). Quasi-maximum-likelihood estimation in conditionally heteroscedastic time series: a stochastic recurrence equations approach. To appear in *Ann. Statist.*

SUBBA RAO, T. (1977). On the estimation of bilinear time series models. *Bull. Internat. Statist. Inst.* **41,** 413–417.

TERDIK, G. (1999). *Bilinear Stochastic Models and Related Problems of Nonlinear Time Series Analysis. A Frequency Domain Approach* (Lecture Notes Statist. **142**). Springer, New York.

TWEEDIE, R. L. (1983). Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In *Probability, Statistics and Analysis*, eds J. Kingman and G. Reuter, Cambridge University Press, pp. 260–277.