On some nonstationary, nonlinear random processes and their stationary approximations

Suhasini Subba Rao*
Universität Heidelberg,
Institut für Angewandte Mathematik,
69120 Heidelberg, Germany

August 31, 2006

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 ARCH and GARCH 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 satisfy alpha mixing properties.

Some key words: Derivative process; Local stationarity; Nonlinear; Nonstationary; State-space models.

*The author is currently at the Department of Statistics, Texas A&M, College Station, Texas 77843-3143

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 maybe 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 for example see Priestley (1965) and Crámer (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 all 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, though these have mainly been restricted to stationary processes. Standard nonlinear models include ARCH (Engle, 1982), GARCH (Bollerslev, 1986), Bilinear (Subba Rao (1977) and Terdik (1999)) and Random coefficient processes (Nicholls & Quinn, 1982) etc. Often these stationary, nonlinear processes have a state-space representation, for example, see Brandt (1986), Bougerol and Picard (1992a) and Straumann and Mikosch (2006).

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

$$\mathcal{X}_{t,N} = A_t(\frac{t}{N})\mathcal{X}_{t-1,N} + \underline{b}_t(\frac{t}{N}), \quad t = 1, \ldots, N, \tag{1}$$

where $\mathcal{X}_{t,N}$ and $\{\underline{b}_t(\frac{t}{N})\}_t$ are p-dimensional nonstationary random vectors and $\{A_t(\frac{t}{N})\}$ are $p \times p$ dimensional nonstationary random matrices.

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

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

where u is fixed and $\{\underline{b}_t(u) : t \in \mathbb{Z}\}$ and $\{A_t(u) : t \in \mathbb{Z}\}$ are p-dimensional stationary random vectors and $p \times p$-dimensional stationary random matrices respectively. We will prove that $\mathcal{X}_t(u)$ can be regarded as a stationary approximation of $\mathcal{X}_{t,N}$ for t/N close to u. In Section 3 we define the derivative process which is a measure of 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 probablistic results (such as mixing properties) associated with the observed process and Taylor series expansions. In Section 5 we consider the particular example of the time-varying GARCH process, and show that it satisfies all the results stated above. We mention that other processes, such as the time-varying random coefficient AR and GARCH process also have the representation (1) and satisfy the results in this paper. The idea of a stationary approximation and a derivative process was established for time-varying ARCH processes in Dahlhaus and Subba Rao (2006).

2 Nonlinear time-varying processes

2.1 Assumptions

In this section we state the assumptions and notation.

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

The Lyapunov exponent associated to a sequence of random matrices $\{A_t : t \in \mathbb{N}\}$ is defined as

$$\inf\{\frac{1}{n}\mathbb{E}(\log \|A_t A_{t-1}\dots A_{t-n+1}\|_{spec}) : n \in \mathbb{N}\}. \qquad (3)$$

We make the following assumptions.

Assumption 2.1 *The sequence of stochastic processes $\{\mathcal{X}_{t,N}\}$ has a time-varying state-space representation defined in (1), where the random matrices $\{A_t(u)\}$ and vectors $\{\underline{b}_t(u)\}$ satisfy the following assumptions:*

(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 $\{\mathcal{A}_t(r) : t \in \mathbb{Z}\}$ such that

$$|A_t(u)|_{abs} \leq \mathcal{A}_t(r) \quad \textit{if } u \in [\frac{r-1}{M}, \frac{r}{M}),$$

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

(ii) There exists a $\beta \in (0, 1]$ and $\{\mathcal{A}_t\}$, such that for all $u, v \in [0, 1]$, the matrices $\{A_t(\cdot)\}$ and $\{\underline{b}_t(\cdot)\}$ satisfy with $\tilde{\underline{b}}_t$ from (i):

$$|A_t(u) - A_t(v)|_{abs} \leq C|u - v|^{\beta}\mathcal{A}_t, \qquad |\underline{b}_t(u) - \underline{b}_t(v)|_{abs} \leq C|u - v|^{\beta}\tilde{\underline{b}}_t.$$

Furthermore, for some $\varepsilon > 0$, $\mathbb{E}\|\mathcal{A}_t\|^{\varepsilon} < \infty$.

For convenience from now on we let $A_t(u) = 0$ for $u \leq 0$ and $\prod_{i=0}^{-k} A_i = I$ (if $k \geq 1$).

Assumption 2.1(i) means that the random matrices $\{A_t(u)\}_t$ 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 locally approximate $\{\mathcal{X}_{t,N}\}$ by a stationary process.

2.2 The stationary approximation

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

$$\mathcal{X}_{t,N} = \sum_{k=0}^{\infty} A_t(\frac{t}{N}) \cdots A_{t-k+1}(\frac{t-k+1}{N})\underline{b}_{t-k}(\frac{t-k}{N}). \tag{4}$$

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

$$\underline{Y}_t = \sum_{k=1}^{\infty} \{ \sum_{r=1}^{M} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(r)\}\tilde{\underline{b}}_{t-k}. \tag{5}$$

Theorem 2.1 *Suppose Assumption 2.1 holds, let $\mathcal{X}_{t,N}$, $\mathcal{X}_t(u)$ and Y_t be defined as in (1), (2) and(5) respectively and suppose there exists an $\varepsilon > 0$ such that*

$$\sup_{t,N} \mathbb{E}\|\mathcal{X}_{t,N}\|_1^{\varepsilon} < \infty \text{ and } \sup_{u} \mathbb{E}\|Y_t\|_1^{\varepsilon} < \infty. \tag{6}$$

Then we have

$$|\mathcal{X}_{t,N} - \mathcal{X}_t(\frac{t}{N})|_{abs} \leq \frac{1}{N^{\beta}}\underline{V}_{t,N}, \tag{7}$$

$$|\mathcal{X}_t(u) - \mathcal{X}_t(w)|_{abs} \leq |u-w|^{\beta}\underline{W}_t \tag{8}$$

$$\textit{and } |\mathcal{X}_{t,N} - \mathcal{X}_t(u)|_{abs} \leq |\frac{t}{N} - u|^{\beta}\underline{W}_t + \frac{1}{N^{\beta}}\underline{V}_{t,N} \tag{9}$$

where

$$\underline{V}_{t,N} = C\sum_{k=1}^{\infty} k\{\prod_{j=0}^{k-1}\mathcal{A}_{t-j}(i_1)\}\{\mathcal{A}_{t-k}|\mathcal{X}_{t-k-1,N}|_{abs} + \underline{\tilde{b}}_{t-k}\}, \tag{10}$$

$$\underline{W}_t = C\sum_{k=1}^{\infty}\{\sum_{r=1}^{M}\prod_{j=0}^{k-1}\mathcal{A}_{t-j}(r)\}\{\mathcal{A}_{t-k}\underline{Y}_{t-k-1} + \underline{\tilde{b}}_{t-k}\} \tag{11}$$

with i_1 *such that* $\frac{i_1-1}{M} \leq \frac{t}{N} < \frac{i_1}{M}$. *Moreover* $\underline{V}_{t,N}$ *converges almost surely and the series* $\{\underline{W}_t\}_t$ *is a well defined stationary process.*

PROOF. In the Appendix. □

The most notable result in the theorem above is (9). In other words 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

$$|\mathcal{X}_{t,N} - \mathcal{X}_t(u)|_{abs} \leq |\frac{t}{N} - u|^{\beta}O_p(1) + \frac{1}{N^{\beta}}O_p(1). \tag{12}$$

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

$$\frac{1}{2M+1}\sum_{k=-M}^{M}\mathcal{X}_{t_0+k,N} = \frac{1}{2M+1}\sum_{k=-M}^{M}\tilde{\mathcal{X}}_{t_0+k}(u_0) + \mathcal{B}_{t_0,N}, \tag{13}$$

where

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

To evaluate the limit of this sum, which involves showing that asymptotically $\mathcal{B}_{t_0,N}$ converges to zero, 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 $\{A_t(u)\}$ and $\{\underline{b}_t(u)\}$. However under stronger assumptions on the moments of $\{A_t(u)\}$ and $\{\underline{b}_t(u)\}$ we will show that $\mathbb{E}(\|\mathcal{X}_{t,N}\|_n^n)$ is uniformly bounded in t and N.

Define the matrix $[A]_n$ as $[A]_n = \left\{\left[\mathbb{E}(|a_{ij}|^n)\right]^{1/n} : i = 1, \ldots, p, j = 1, \ldots, q\right\}$. We now give conditions for $\mathbb{E}\|\mathcal{X}_{t,N}\|_n^n < \infty$ and $\mathbb{E}\|\mathcal{X}_t(u)\|_n^n < \infty$.

Proposition 2.1 *Suppose Assumption 2.1 holds, let $\mathcal{X}_{t,N}$, $\mathcal{X}_t(u)$, $\underline{V}_{t,N}$ and $\underline{W}_t$ be defined as in (1), (2), (10) and (11) respectively. Suppose for all $r \in \{1, \ldots, M\}$ and for some $n \in [1, \infty)$, that $\mathbb{E}\|\tilde{\underline{b}}_t\|_n^n < \infty$ and, for some $\delta > 0$, $\lambda_{spec}\{[A_t(r)]_n\} < 1 - \delta$. Then*

$$\sup_{t,N} \mathbb{E}\|\mathcal{X}_{t,N}\|_n^n < \infty \quad \sup_{t,N} \mathbb{E}\|\underline{V}_{t,N}\|_n^n < \infty, \quad \sup_{u} \mathbb{E}\|\mathcal{X}_t(u)\|_n^n < \infty \quad \textit{and} \quad \mathbb{E}\|\underline{W}_t\|_n^n < \infty \tag{14}$$

From Proposition 2.1 we see that if these conditions are satisfied then condition (6) is satisfied with $\varepsilon = n$, hence the locally stationary conclusions of Theorem 2.1 immediately follow.

We now apply the above results to the local average example in (13). Under the assumption that all the conditions in Proposition 2.1 are satisfied with $n = 1$, we have

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

and

$$\frac{1}{2M+1} \sum_{k=-M}^{M} \mathcal{X}_{t_0+k,N} = \frac{1}{2M+1} \sum_{k=-M}^{M} \tilde{\mathcal{X}}_{t_0+k}(u_0) + O_p((\frac{M}{N})^{\beta} + \frac{1}{N^{\beta}}). \tag{15}$$

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

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

where $M \to 0$, $M/N \to 0$ 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 have shown that time-varying processes can locally be approximated by a stationary process. In this section, under additional conditions on $\{A_t(u)\}_t$ and $\{\underline{b}_t(u)\}_t$, we improve the approximation in (12) and show that a Taylor series expansion of the time-varying process in terms of stationary processes, can be derived (Theorem 3.2). In order to do this we define the derivative process and show that the derivative process also has a state-space representation.

The Taylor expansion of a given time-varying process in terms of stationary processes is, in particular, of 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 $\{A_t(\frac{t}{N})\}$ 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 $\{\dot{A}_t(\frac{t}{N})\}_t$ (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 $A(u)$ is a $p \times q$ random matrix, we let $\dot{A}_t(u) = \{\frac{\partial A(u)_{i,j}}{\partial u} : i = 1, \ldots, p, j = 1, \ldots, q\}$. We make the following assumptions.

Assumption 3.1 *The sequence of stochastic processes $\{\mathcal{X}_{t,N}\}$ have a time-varying state-space representation defined in (1), $\{\mathcal{A}_t(i)\}$, $\{\tilde{\underline{b}}_t\}$ and $\{\mathcal{A}_t\}$ are defined as in Assumption 2.1, and $\{A_t(u)\}$ and $\{\underline{b}_t(u)\}$ satisfy the following properties:*

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

(ii) Let $\beta' > 0$. The matrices $\{A_t(\cdot)\}$ and $\{\underline{b}_t(\cdot)\}$ are assumed to satisfy:

$$|A_t(u) - A_t(v)|_{abs} \leq C|u - v|\mathcal{A}_t, \qquad |\underline{b}_t(u) - \underline{b}_t(v)|_{abs} \leq C|u - v|\tilde{\underline{b}}_t$$

$$|\dot{A}_t(u) - \dot{A}_t(v)|_{abs} \leq C|u - v|^{\beta'}\mathcal{A}_t, \qquad |\dot{\underline{b}}_t(u) - \dot{\underline{b}}_t(v)|_{abs} \leq C|u - v|^{\beta'}\tilde{\underline{b}}_t,$$

$$\sup_u |\dot{A}_t(u)| \leq C\mathcal{A}_t(r) \text{ and } \sup_u |A_t(u)| \leq C\mathcal{A}_t(r) \quad \text{if } u \in [\frac{r-1}{M}, \frac{r}{M}),$$

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

We now define the process $\{\dot{\mathcal{X}}_t(u)\}_t$, which we call the derivative process. By formally differentiating (2) with respect to u we have

$$\dot{\mathcal{X}}_t(u) = \dot{A}_t(u)\mathcal{X}_{t-1}(u) + A_t(u)\dot{\mathcal{X}}_{t-1}(u) + \dot{\underline{b}}_t(u). \tag{16}$$

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 $\dot{A}_t(u)$. In other words, given that $\mathcal{X}_{t,N}$ is well defined, the existence of $\dot{A}_t(u)$ is sufficient for the derivative process to also be well defined. This will become clear when we rewrite $\dot{\mathcal{X}}_t(u)$ as a state space model. Let $\mathcal{X}_t(2,u)^T = (\dot{\mathcal{X}}_t(u)^T, \mathcal{X}_t(u)^T)$, then it is clear that $\{\mathcal{X}_t(2,u)\}_t$ satisfies the representation

$$\mathcal{X}_t(2,u) = A_t(2,u)\mathcal{X}_{t-1}(2,u) + \underline{b}_t(2,u), \tag{17}$$

where

$$A_t(2,u) = \begin{pmatrix} A_t(u) & \frac{\partial A_t(u)}{\partial u} \\ 0 & A_t(u) \end{pmatrix} \text{ and } \underline{b}_t(2,u) = \begin{pmatrix} \dot{\underline{b}}_t(u) \\ \underline{b}_t(u) \end{pmatrix}. \tag{18}$$

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

$$\mathcal{X}_{t,N}(2) = A_t(2, \frac{t}{N})\mathcal{X}_{t-1,N}(2) + \underline{b}_t(2, \frac{t}{N}), \tag{19}$$

with $\{\mathcal{X}_{t,N}(2)^T\}_t = \{(\dot{\mathcal{X}}_{t,N}^T, \mathcal{X}_{t,N}^T)\}$. 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 maybe difficult to estimate.

Let $\tilde{\underline{b}}_t(2)^T = (C\tilde{\underline{b}}_t^T, \tilde{\underline{b}}_t^T)$,

$$\mathcal{A}_t(2,r) = \begin{pmatrix} \mathcal{A}_t(r) & C\mathcal{A}_t(r) \\ 0 & \mathcal{A}_t(r) \end{pmatrix} \quad \text{and} \quad \mathcal{A}_t(2) = \begin{pmatrix} \mathcal{A}_t & C\mathcal{A}_t \\ 0 & \mathcal{A}_t \end{pmatrix} \tag{20}$$

where C is defined as in Assumption 3.1.

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

Lemma 3.1 *Suppose Assumption 3.1 is satisfied, then for $r = 1, \ldots, M$, we have*

$$\inf\{\frac{1}{n}\mathbb{E}(\log \|A_t(2,u)\ldots A_{t-n+1}(2,u)\|_{spec})\} < 0, \tag{21}$$

$$\inf\{\frac{1}{n}\mathbb{E}(\log \|\mathcal{A}_t(2,r)\ldots \mathcal{A}_{t-n+1}(2,r)\|_{spec})\} < 0. \tag{22}$$

PROOF. In the Appendix □

From the above result we can show that (17) and (19) has a unique causal solution similar to (4). Furthermore, the following theorem follows from the lemma above.

Theorem 3.1 *Suppose Assumption 3.1 holds and let $\{\mathcal{X}_{t,N}(2)\}$ be defined as in (19). Then the process $\{\mathcal{X}_{t,N}(2)\}$ satisfies Assumption 2.1 with transition matrices $\{A_t(2,u); t \in \mathbb{Z}, u \in (0,1]\}$ and innovations vectors $\{\underline{b}_t(2,u) : t \in \mathbb{Z}, u \in (0,1]\}$.*

PROOF. Under Assumption 3.1 we have $|A_t(2,u)|_{abs} \leq \mathcal{A}_t(2,r)$ if $u \in (\frac{r-1}{M}, \frac{r}{M}]$ and $\sup_u |\underline{b}_t(2,u)|_{abs} \leq \underline{\tilde{b}}_t(2)$. Furthermore, by using Lemma 3.1, the random matrix sequences $\{A_t(2,u)\}_t$ and $\{\mathcal{A}_t(2,r)\}_t$ have negative Lyapunov exponents. Hence all the conditions in 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

$$\underline{Y}_t(2) = \sum_{k=1}^{\infty}\{\sum_{r=1}^{M}\prod_{j=0}^{k-1}\mathcal{A}_{t-j}(2,r)\}\underline{\tilde{b}}_{2,t-k}. \tag{23}$$

Corollary 3.1 *Suppose Assumption 3.1 holds. Let $\mathcal{X}_{t,N}(2)$, $\mathcal{X}_t(2,u)$, $\mathcal{A}_t(2,r)$, $\mathcal{A}_t(2)$, $\underline{\tilde{b}}_t(2)$, $\underline{W}_t$ and $\underline{Y}_t(2)$ be defined as in (19), (17), (20), (11), (23) and there exists a $\varepsilon > 0$ such that*

$$\sup_{t,N}\mathbb{E}\|\mathcal{X}_{t,N}(2)\|_1^{\varepsilon} < \infty \textit{ and } \mathbb{E}\|\underline{Y}_t(2)\|_1^{\varepsilon} < \infty.$$

Then $|\mathcal{X}_{t,N}(2) - \mathcal{X}_t(2,\frac{t}{N})|_{abs} \leq \frac{1}{N^{\beta'}}\underline{V}_{t,N}(2)$, $|\mathcal{X}_t(2,u) - \mathcal{X}_t(2,w)|_{abs} \leq |u-w|^{\beta'}\underline{W}_t(2)$,

$$|\mathcal{X}_{t,N}(2) - \mathcal{X}_t(2,u)|_{abs} \leq |\frac{t}{N} - u|^{\beta'}\underline{W}_t(2) + \frac{1}{N^{\beta'}}\underline{V}_{t,N}(2), \tag{24}$$

and $\|\underline{W}_t\|_1 \leq \|\underline{W}_t(2)\|_1$, *where* $\{\underline{V}_{t,N}(2)\}$ *is similar to* $\{\underline{V}_{t,N}\}$ *defined in (10), but with* $\mathcal{A}_{t-j}(i_1)$, $\underline{X}_{t-k-1,N}$ *and* $\mathcal{A}_{t-j}$ *replaced by* $\mathcal{A}_{t-j}(2,i_1)$, $\underline{X}_{t-k-1,N}(2)$ *and* $\mathcal{A}_t(2)$ *respectively and*

$$\underline{W}_t(2) = C\sum_{r=1}^{M}\sum_{k=1}^{\infty}\{\prod_{j=0}^{k-1}\mathcal{A}_{t-j}(2,r_1)\}\{\mathcal{A}_{t-k}(2)|\underline{Y}_{t-k-1}(2)|_{abs} + \underline{\tilde{b}}_{t-k}(2)\}. \quad (25)$$

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

PROOF. Under Assumption 3.1 and by using 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 (9). To do this we first show that the derivative process is almost surely Hölder continuous.

Corollary 3.2 *Suppose Assumption 2.1 holds. Let* $\{\mathcal{X}_t(2,u)\}$ *and* $\{\underline{W}_t(2)\}_t$ *be defined as in (17) and (25) respectively. Then*

$$\dot{\mathcal{X}}_t(u) = \sum_{k=0}^{\infty}\sum_{r=0}^{k-1}\{[\prod_{i=0}^{r-1}A_{t-i}(u)]\dot{A}_{t-r}(u)[\prod_{i=r+1}^{k-1}A_{t-i}(u)]\}\underline{b}_{t-k}(u) + \sum_{k=1}^{\infty}\{\prod_{i=0}^{k-1}A_{t-i}(u)\}\dot{\underline{b}}_{t-k}(u), \quad (26)$$

is almost surely the unique well defined solution of (16). Furthermore,

$$\sup_{u,v}\left|\dot{\mathcal{X}}_t(u) - \dot{\mathcal{X}}_t(v)\right|_{abs} \leq |u-v|^{\beta'}\{W_t(2)\}_{1,\ldots,p}, \quad (27)$$

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

PROOF. By expanding (17) and using standard results in Brandt (1986) we can show that (26) holds. To show that the right hand side of (26) is the derivative of $\mathcal{X}_t(u)$, we note that both $|\mathcal{X}_t(u)|$ and the absolute sum on right hand side of (26) are almost surely bounded. Thus we can exchange the summation and derivative, it immediately follows that (26) is the derivative of $\mathcal{X}_t(u)$.

(27) follows immediately from Corollary 3.1 and by using (26) and (27), we have $\mathcal{X}_t(u,\omega) \in \text{Lip}(1+\beta')$ for all $\omega \in \mathcal{N}^c$, where $P(\mathcal{N}) = 0$. □

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 defined as in (19) and (17) respectively. Suppose the assumptions of Corollary 3.1 hold. Then*

(i)

$$\mathcal{X}_{t,N} = \mathcal{X}_t(u) + (\frac{t}{N} - u)\dot{\mathcal{X}}_t(u) + O_p(|\frac{t}{N} - u|^{\beta'+1} + \frac{1}{N}); \tag{28}$$

(ii)

$$\mathcal{X}_{t,N} = \mathcal{X}_t(u) + (\frac{t}{N} - u)\dot{\mathcal{X}}_{t,N} + O_p(|\frac{t}{N} - u|^{\beta'+1} + \frac{1}{N}). \tag{29}$$

PROOF. Let $\mathcal{N}_1$ be a set of zero measure such that $\mathcal{X}_{t,N}(\omega)$, $\underline{W}_t(2,\omega)$ (defined in Corollary 3.2) and $\underline{Y}_t(2,\omega)$ converge for all $\omega \in \mathcal{N}_1^c$. Then by using Corollary 3.2 we have that $\mathcal{X}_t(u,\omega) \in Lip(1+\beta')$ if $\omega \in \mathcal{N}_1^c$. By using (7), making a Taylor series expansion of $\mathcal{X}_t(\frac{t}{N},\omega)$ about u and using the mean value theorem we obtain

$$\mathcal{X}_{t,N}(\omega) = \mathcal{X}_t(u,\omega) + (\frac{t}{N} - u)\dot{\mathcal{X}}_t(u,\omega) + \big(|\frac{t}{N} - u|^{1+\beta'} + \frac{1}{N}\big)R_N(\omega),$$

where $|R_N(\omega)| \leq \|V_{t,N}(\omega)\|_1 + \|\underline{W}_t(2,\omega)\|_1$. Therefore since $\mathbb{P}(\mathcal{N}_1^c) = 1$ we obtain (28). We use (24) and the method given above to prove (29). □

We observe that (28) means that 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 in Proposition 2.1 that the moments of the derivative process are uniformly bounded.

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

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

PROOF. Since for each $r \in \{1,\ldots,M\}$, $\mathcal{A}_t(2,n)$ is a block upper triangular matrix we observe that $\lambda_{spec}([\mathcal{A}_t(2,r)]_n) = \lambda_{spec}([\mathcal{A}_t(r)]_n)$. Therefore the same proof as in Proposition 2.1 can be used to prove the result. □

We now return to the local average example in (13), and obtain a tighter bound for the remainder $\mathcal{B}_{t_0,N}$. By using (28) we have

$$\frac{1}{2M+1}\sum_{k=-M}^{M}(\mathcal{X}_{t_0+k,N}-\mathcal{X}_{t_0+k}(u_0)) = \frac{1}{2M+1}\sum_{k=-M}^{M}\left(\frac{k}{N}\right)\dot{\mathcal{X}}_{t_0+k}(u_0)+R_{t_0,N} \qquad (30)$$

where

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

It follows from (30) that the size of the remainder or bias due to nonstationarity depends on the magnitude of the derivative processes $\{\dot{\mathcal{X}}_t(u)\}_t$. Furthermore, if the conditions in Proposition 3.1 are satisfied with $n=2$, then $(\mathbb{E}\|\mathcal{B}_{t_0,N}\|_2^2)^{1/2} = O(\frac{M}{N}+\frac{1}{N})$. 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 the derivative process is strongly mixing). Let us suppose $\{\dot{\mathcal{X}}_t(u)\}_t$ is a short memory process, then we have $\frac{1}{2M+1}\sum_{k=-M}^{M}\left(\frac{k}{N}\right)\dot{\mathcal{X}}_{t_0+k}(u_0) = \frac{\sqrt{M}}{N}O_p(1)$. Therefore if $\sqrt{M}/N << (M/N)^{1+\beta'}$, we have

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

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

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

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 by using $\mathbb{E}\|\mathcal{B}_{t_0,N}\|_2^2$ and $\mathrm{var}[\frac{1}{2M+1}\sum_{k=-M}^{M}\mathcal{X}_{t+k}(u)]$ we are able to evaluate the mean squared error of the local average and thus obtain the optimal segment length M.

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 ϕ-irreducibility of the derivative process. We state below a state-space version of this theorem given in Basrak et al. (2002), Theorem 2.8 and Remark 2.9.

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

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

Lemma 4.2 *Suppose Assumption 2.1 is satisfied and* $A_t(2,u)$ *is defined as in (18). Then*

$$\mathbb{E}\|A_t(2,u)\|_{spec}^{\varepsilon} < \infty \tag{31}$$

and $\{A_t(2,u)\}_t$ *has a negative Lyapunov exponent.*

PROOF. Under Assumption 2.1 and by using (49) with $n = 1$ we have $\|A_t(2,u)\|_{spec} \leq C\|A_t(2,u)\|_{spec}$, thus $\mathbb{E}\|A_t(2,u)\|_{spec} \leq 2C'\mathbb{E}\|A_t(u)\|_{spec} < \infty$, which gives us (31). From (21) we have that $\{A_t(2,u)\}$ has a negative Lyapunov exponent. $\square$

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

Theorem 4.1 *Suppose Assumption 3.1 (with* $d = 1$*) is satisfied and let the process* $\{\mathcal{X}_t(2,u)\}_t$ *defined in (17) be* ϕ*-irreducible. Then the process* $\{\mathcal{X}_t(2,u)\}$ *is geometrically ergodic and thus strongly mixing with a geometric rate.*

PROOF. We now show that the conditions of Lemma 4.1 are satisfied, the result then follows.

By using Lemma 4.2 there exists an $m > 0$ and $\delta < 0$ such that

$$\frac{1}{m}\mathbb{E}\log\|A_t(2,u)\ldots A_{t-m+1}(2,u)\|_{spec} \leq \delta. \tag{32}$$

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

$$\mathcal{X}_{m,t}(2) = C_{m,tm}(2,u)\mathcal{X}_{m,t-1}(2) + \underline{d}_{m,tm}(u),$$

with $C_{m,t}(2,u) = A_t(2,u) \dots A_{t-m+1}(2,u)$ and

$$\underline{d}_{m,t}(2,u) = \sum_{k=1}^{m-1} A_t(2,u) \dots A_{t-k+1}(2,u)\underline{b}_{t-k}(2,u) + \underline{b}_t(2,u). \tag{33}$$

By using (31) we have that $\mathbb{E}\|C_{m,t}(2,u)\|_{spec} < \infty$.

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

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

Corollary 4.1 *Let $\{\mathcal{X}_t(2,u)\}$ be defined as in (17). Suppose Assumption 3.1 holds (with $d = 1$), $\{\mathcal{X}_t(2,u)\}$ is ϕ-irreducible and $\mathbb{E}\|\mathcal{X}_t(2,u)\|^2 < \infty$. Then we have*

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

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

PROOF. By using Theorem 4.1 we have 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 $\{\mathcal{X}_t(u)\}_t$ and $\{\dot{\mathcal{X}}_t(u)\}_t$ are short memory processes.

5 An Example: time-varying GARCH

In this section we show that the time-varying GARCH admits the represention (1) and the results in the previous section apply to the tvGARCH process. We mention that the results below also apply to the tvARCH process, as it is a special case of the tvGARCH process. And the conditions stated here are slightly more general than the conditions given in Dahlhaus and Subba Rao (2006). Let 0_p denote a p-dimensional zero vector, $0_{p\times q}$ a $p \times q$-dimensional zero matrix and $\mathbb{I}_{p\times q}$ a $p \times q$-dimensional with $(\mathbb{I}_{p\times q})_{i,j} = 1$ for all i and j.

We first note that the sequence of processes $\{X_{t,N}\}$ is called a time-varying Generalised Autoregressive Conditional Heteroscedasticity (p,q) (tvGARCH(p,q)) process if it satisfies

$$X_{t,N} = Z_t\sigma_{t,N}, \quad \sigma_{t,N}^2 = a_0(\frac{t}{N}) + \sum_{i=1}^{p} a_i(\frac{t}{N})X_{t-i,N}^2 + \sum_{j=1}^{q} b_j(\frac{t}{N})\sigma_{t-j,N}^2, \qquad t = 1, \dots, N \tag{34}$$

where $\{Z_t\}$ are independent, identically distributed random variables, with $\mathbb{E}(Z_t) = 0$ and $\mathbb{E}(Z_t^2) = 1$. It is straightforward to show that the tvGARCH process $X_{t,N}^2$ admits the state-space representation (1), where $\mathcal{X}_{t,N}^T = (\sigma_{t,N}^2, \ldots, \sigma_{t-q+1,N}^2, X_{t-1,N}^2, \ldots, X_{t-p+1,N}^2)$, $\underline{b}_t(u)^T = (a_0(u), 0, \ldots, 0) \in \mathbb{R}^{p+q-2}$, $A_t(u)$ is a $(p+q-1) \times (p+q-1)$ matrix defined as

$$A_t(u) = \begin{pmatrix} \underline{\tau}_t(u) & b_q(u) & \underline{a}(u) & a_p(u) \\ I_{q-1} & 0 & 0 & 0 \\ \underline{Z}_{t-1}^2 & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{pmatrix}, \tag{35}$$

$\tau_t(u) = (b_1(u) + a_1(u)Z_{t-1}^2, b_2(u), \ldots, b_{q-1}(u))$, $\underline{a}(u) = (a_2(u), \ldots, a_{p-1}(u))$ and $\underline{Z}_{t-1}^2 = (Z_{t-1}^2, 0, \ldots, 0) \in \mathbb{R}^{q-1}$ (we assume without loss of generality $p, q \geq 2$).

The tvGARCH process and Assumption 2.1

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

$$\sup_u \Big\{ \sum_{i=1}^p a_i(u) + \sum_{j=1}^q b_j(u) \Big\} < (1-\eta), \tag{36}$$

then Assumption 2.1 is satisfied for the tvGARCH(p,q) process. Using the β-Lipschitz continuity of the parameters we now show that there exist matrices $\mathcal{A}_t(r)$ which bound $A_t(u)$ and satisfy Assumption 2.1(i).

Let K_{max} be such that $\sup_{u,v} |a_i(u) - a_i(v)| \leq K_{\max}|u-v|^\beta$ and $\sup_{u,v} |b_j(u) - b_j(v)| \leq K_{\max}|u-v|^\beta$. Define ε such that $\varepsilon \leq \big\{\eta/(2K_{\max}(p+q))\big\}^{1/\beta}$ and $\varepsilon^{-1} \in \mathbb{N}$. Let $M(\varepsilon) = \varepsilon^{-1}$, and for each $r \in \{1, \ldots, M(\varepsilon)\}$, $i = 1, \ldots, p$ and $j = 1 \ldots, q$ define

$$\alpha_i(r) = \big\{a_i[(k-1)\varepsilon] + K_{\max}\varepsilon^\beta\big\} \text{ and } \beta_j(r) = \big\{b_j[(k-1)\varepsilon] + K_{\max}\varepsilon^\beta\big\}, \tag{37}$$

Therefore by using (36) and the above construction we have $\sup_{(r-1)\varepsilon \leq u < r\varepsilon} a_i(u) \leq \alpha_i(r)$, $\sup_{(r-1)\varepsilon \leq u < r\varepsilon} b_j(u) \leq \beta_j(r)$ and

$$\sum_{i=1}^p \sup_{(r-1)\varepsilon \leq u < r\varepsilon} a_i(u) + \sum_{j=1}^q \sup_{(r-1)\varepsilon \leq u < r\varepsilon} b_j(u) \leq \sum_{i=1}^p \alpha_i(r) + \sum_{j=1}^q \beta_j(r) \leq 1 - \eta/2. \tag{38}$$

Let

$$\mathcal{A}_t(r) = \begin{pmatrix} \underline{\tilde{\tau}}_t(r) & \beta_q(r) & \underline{\alpha}(r) & \alpha_p(r) \\ I_{q-1} & 0 & 0 & 0 \\ \underline{Z}_{t-1}^2 & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{pmatrix}, \tag{39}$$

$\underline{\tilde{\tau}}_t(r) = (\beta_1(r) + \alpha_1(r)Z_{t-1}^2, \beta_2(r), \ldots, \beta_{q-1}(r))$, $\underline{\alpha}(r) = (\alpha_2(r), \ldots, \alpha_{p-1}(r))$ and $\underline{Z}_{t-1}^2 = (Z_{t-1}^2, 0, \ldots, 0) \in \mathbb{R}^{p-1}$. Then it is clear that $\sup_{(r-1)\varepsilon \le u < r\varepsilon} |A_t(u)|_{abs} \le \mathcal{A}_t(r)$. To summarise, we have partitioned the unit interval into $M(\varepsilon)$ intervals, such that all the matrices $A_t(u)$ 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, $\{\mathcal{A}_t(r)\}_t$ are independent, identically distributed sequences of random matrices. Since $\mathbb{E}(Z_0^2) = 1$ 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 $\lambda_{spec}\{\mathbb{E}(\mathcal{A}_t(r))\} \le (1-\eta/2)^{1/(p+q-1)}$. By using Kesten and Spitzer (1984), equation (1.4), we can show that $\{\mathcal{A}_t(r)\}$ has a negative Lyapunov exponent.

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

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

$$|A_t(u) - A_t(v)|_{abs} \le K|u-v|^\beta \mathcal{A}_t \quad \text{and } |\underline{b}_t(u) - \underline{b}_t(v)|_{abs} \le K|u-v|^\beta \underline{\tilde{b}}_t,$$

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

The tvGARCH process and the stationary approximation

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

$$X_t(u)^2 = \big\{a_0(u) + \sum_{i=1}^p a_i(u)X_{t-i}(u)^2 + \sum_{j=1}^q b_j(u)\sigma_{t-j}(u)^2\big\}Z_t^2. \tag{40}$$

In order to show that $X_t(u)^2$ locally approximates $X_{t,N}^2$, we need to verify the conditions in 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, $\mathbb{E}\|\mathcal{X}_{t,N}\|_1^\varepsilon$ and $\mathbb{E}\|Y_t\|_1^\varepsilon$. We do this by verifying the conditions of Proposition 2.1, for $\varepsilon = n$ (though it is enough to prove the result for $\varepsilon = 1$). Suppose $n \in [1, \infty)$ and let $\mu_n = \{\mathbb{E}(Z_t^{2n})\}^{1/n}$. In addition we assume

$$\mu_n \sup_u \big\{ \sum_{i=1}^p a_i(u) + \sum_{j=1}^q b_j(u) \big\} < 1 - \eta, \tag{41}$$

for some $\eta > 0$. Using a similar construction to the above we now construct matrices $\mathcal{A}_t(r)$ where $A_t(u) \le \mathcal{A}_t(r)$ for $(r-1)/M(\varepsilon) \le u < r/M(\varepsilon)$ such that $\lambda_{spec}([A_t(r)]_n) \le (1 - \eta/2)^{1/(p+q-1)}$, thus verifying the conditions of Proposition 2.1. Let $\varepsilon = \big(\eta/(2\mu_n K_{\max}(p+q))\big)^{1/\beta}$, now by using the methods given in (38), define $M(\varepsilon)$, $\alpha_i(r)$ and $\beta_j(r)$ (defined in

(37)) and $\mathcal{A}_t(r)$ as in (39), using the new ε. It is straightforward to show

$$\sup_u \mu_n \big\{ \sum_{i=1}^{p} \alpha_i(r) + \sum_{j=1}^{q} \beta_j(r) \big\} < 1 - \eta/2. \tag{42}$$

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

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

$$A = \begin{pmatrix} \underline{\tau} & b_q & \underline{a} & a_p \\ I_{q-1} & 0 & 0 & 0 \\ \underline{\mu} & 0 & 0 & 0 \\ \underline{0} & 0 & I_{p-2} & 0 \end{pmatrix}, \tag{43}$$

where $\underline{\tau} = (b_1 + a_1\mu, b_2, \ldots, b_{q-1}) \in \mathbb{R}^{q-1}$, $\underline{a} = (a_2, \ldots, a_{p-1}) \in \mathbb{R}^{p-2}$ and $\underline{\mu} = (\mu, 0, \ldots, 0) \in \mathbb{R}^{q-1}$. Suppose $\mu(\sum_{i=1}^{p} a_i + \sum_{j=1}^{q} b_j) < 1 - \delta$, where $p, q \geq 2$ and $\delta > 0$. Then we have $\lambda_{spec}(A) \leq (1-\delta)^{1/(p+q-1)}$.

Now we construct the matrix $\mathcal{A}(r)^*$ which is the same as A in (43), where a_i, b_i and μ have been 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 by using Lemma 5.1 and (42) we then have that $\lambda_{spec}\{\mathcal{A}(r)^*\} \leq (1-\eta/2)^{1/(p+q-1)}$. Thus the conditions of Proposition 2.1 are satisfied and we have $\sup_{t,N} \mathbb{E}\|\mathcal{X}_{t,N}\|_n^n < \infty$, (which implies $\sup_{t,N} \mathbb{E}(X_{t,N}^{2n}) < \infty$) and $\mathbb{E}\|Y_t\|_n^n < \infty$.

Therefore if (41) holds with some $n \geq 1$, the conditions of Theorem 2.1 are fulfilled, and we have

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

The GARCH and derivative process

We now consider the stationary derivative process associated with the tvGARCH. Formally differentiating (40) gives

$$\begin{aligned} \frac{\partial \tilde{X}_t(u)^2}{\partial u} &= a_0'(u) + \sum_{i=1}^{p} \big\{ \frac{a_i(u)}{\partial u} \tilde{X}_{t-i}(u)^2 + a_i(u) \frac{\partial \tilde{X}_{t-i}(u)}{\partial u} \big\} + \\ &\quad + \sum_{j=1}^{q} \big\{ \frac{\partial b_j(u)}{\partial u} \tilde{\sigma}_{t-j}(u)^2 + b_j(u) \frac{\partial \tilde{\sigma}_{t-j}(u)^2}{\partial u} \big\}, \end{aligned}$$

which we have shown in Section 3 admits a state space representation. By applying Theorem 3.2 we have the Taylor series expansion

$$X_{t,N}^2 = \tilde{X}_t(u)^2 + (\frac{t}{N} - u)\frac{\partial \tilde{X}_t(u)^2}{\partial u} + (|\frac{t}{N} - u|^{1+\beta'} + \frac{1}{N}). \tag{44}$$

Finally if (41) holds, then the conditions of Proposition 3.1 are satisfied, and $\mathbb{E}\|\frac{\partial \tilde{X}_t(u)^2}{\partial u}\|_n^n < \infty$.

6 Applications

The notion of the stationary approximations and the derivative process can fruitfully be used in many applications. The key is the representation (28) of the nonstationary process in terms of stationary processes. As indicated by the local average 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 is given in Dahlhaus and Subba Rao (2006), Theorem 3, where the properties of a local likelihood estimator have been investigated. The results of this paper can be used to derive similar results for the models used as examples in the present paper (among others). The derivative process in (28) 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.

Acknowledgments

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).

A Appendix

In this section we sketch some of the proofs of the results stated earlier. Full details can be found in the technical report available from the author.

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

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

$$\underline{Y}_t = \sum_{k \geq 1} \{ \prod_{r=1}^{M} \prod_{i \in J^t_{r,k}} \mathcal{A}_{t-i}(r) \} \underline{d}_{t-k} + \underline{d}_t \tag{45}$$

and

$$\underline{Y}_t = \sum_{k \geq 1} k \{ \prod_{r=1}^{M} \prod_{i \in J^t_{r,k}} \mathcal{A}_{t-i}(r) \} \underline{d}_{t-k} \tag{46}$$

converges almost surely.

We use the lemma above to prove Theorem 2.1.

PROOF of Theorem 2.1 By the triangular inequality we have

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

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

$$\begin{aligned}
|\mathcal{X}_{t,N} - \mathcal{X}_t(\frac{t}{N})|_{abs} &= |A_t(\frac{t}{N})\{\mathcal{X}_{t-1,N} - \tilde{\mathcal{X}}_{t-1}(\frac{t}{N})\}|_{abs} \\
&= |A_t(\frac{t}{N})\{A_{t-1}(\frac{t-1}{N}) - A_{t-1}(\frac{t}{N})\}\mathcal{X}_{t-2,N} + A_{t-1}(\frac{t}{N})\{\mathcal{X}_{t-2,N} - \tilde{\mathcal{X}}_{t-2}(\frac{t}{N})\} + \\
&\quad + A_t(\frac{t}{N})\{\underline{b}_{t-1}(\frac{t-1}{N}) - \underline{b}_{t-1}(\frac{t}{N})\}|_{abs} \\
&\leq \frac{1}{N^{\beta}}\mathcal{A}_t(i_1)\underline{b}_{t-1} + \frac{1}{N^{\beta}}\mathcal{A}_t(i_1)\mathcal{A}_{t-1}|\mathcal{X}_{t-2,N}|_{abs} + \mathcal{A}_{t-1}(i_1)|\mathcal{X}_{t-2,N} - \tilde{\mathcal{X}}_{t-2}(\frac{t}{N})|_{abs}.
\end{aligned}$$

Now by continuing the iteration above we obtain $|\mathcal{X}_{t,N} - \mathcal{X}_t(\frac{t}{N})|_{abs} \leq \frac{1}{N^{\beta}}\underline{V}_{t,N}$, where $\underline{V}_{t,N}$ is defined in (10). Under Assumption 2.1 and by using (6) we have that $\mathbb{E}(\log\|\mathcal{A}_t\mathcal{X}_{t-1,N}\|) < \infty$. Therefore by using (46) with $d_t := (\mathcal{A}_t\mathcal{X}_{t-1,N} + \tilde{\underline{b}}_t)$, we have that $\underline{V}_{t,N}$ converges almost surely.

Using a similar method to the above we can show

$$|\mathcal{X}_t(u) - \mathcal{X}_t(w)| \leq |u - w|^{\beta}\underline{W}_t,$$

where $\underline{W}_t$ is defined as in (11). By using Lemma A.1 we have that $\{\underline{W}_t\}_t$ converges almost surely. Finally, (9) follows from (7) and (8). $\square$

Below we make frequent use of the following inequalities (which can proved by repeated use of the Minkowski inequality). Suppose $\{A_t\}_t$ are $p \times p$-dimensional independent random matrices and $\mathcal{X}$ a p-dimensional random vector and $\{A_t\}_t$ and $\mathcal{X}$ are independent, then

$$(\mathbb{E}\|A\mathcal{X}\|_n^n)^{1/n} \leq K\|[A]_n\|_{spec}(\mathbb{E}\|\mathcal{X}\|_n^n)^{1/n} \quad \text{and} \quad \|[A_1 \ldots A_n]_n\|_{spec} \leq K\|[A_1]_n \ldots [A_m]_n\|_{spec}. \tag{47}$$

for some finite constant K.

PROOF of Proposition 2.1 We now show that $\mathbb{E}\|\mathcal{X}_{t,N}\|_n^n$ is uniformly bounded over t and N. Since $\{\mathcal{A}_t(i)\}_t$ and $\{\underline{\tilde{b}}_t\}_t$ are independent by using (47) we have

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

By the above and

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

where $I_r = [\frac{r-1}{M}, \frac{r}{M})$, $J_r = \{k \geq 0 : \frac{t-k}{N} \in I_r\}$ and $J_{r,l} = J_r \cap \{0, 1, \ldots, l-1\}$, we have

$$\begin{aligned}
\mathbb{E}(\|\mathcal{X}_{t,N}\|_n^n)^{1/n} &\leq \sum_{k=0}^{\infty}(\mathbb{E}\|(\prod_{r=1}^{M}\prod_{i\in J_{r,k}}\mathcal{A}_{t-i}(r))\underline{\tilde{b}}_{t-k}\|_n^n)^{1/n} \\
&\leq \sum_{k=0}^{\infty}\prod_{r=1}^{M}\|[\prod_{i\in J_{r,k}}\mathcal{A}_{t-i}(r)]_n\|_{spec}(\mathbb{E}\|\underline{\tilde{b}}_t\|_n^n)^{1/n} \\
&\leq \sum_{k=0}^{\infty}\prod_{r=1}^{M}\|[\mathcal{A}_0(r)]_n^{\#(J_{r,k})}\|_{spec}(\mathbb{E}\|\underline{\tilde{b}}_t\|_n)^{1/n} \qquad (48)
\end{aligned}$$

where $\#(J_{r,k})$=number of elements in the set $J_{r,k}$. Since $\lambda([A_0(i)]_n) \leq 1-\delta$ for $i = 1, \ldots, M$ and by using Moulines et al. (2005), Lemma 12, there exists a K independent of $\mathcal{A}_0(r)$ and m such that $\|[\mathcal{A}_0(r)]_n^m\|_{spec} \leq K(1-\delta/2)^m$. Therefore

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

By substituting the above into (48) and using $\sum_{r=1}^{M} \#(J_{r,k}) = k$ we can show $\sup_{t,N} \mathbb{E}\|\mathcal{X}_{t,N}\|_n^n < \infty$. Using a similar method we can prove the other inequalities in (14). $\Box$

PROOF of Lemma 3.1 We now prove (21). Under Assumption 3.1 it is straightforward to show

$$|A_t(2,u)\ldots A_{t-n+1}(2,u)|_{abs} \leq B_k(t,n)$$

where

$$B_k(t,n) = \begin{pmatrix} \mathcal{A}_t(r)\ldots\mathcal{A}_{t-n+1}(r) & Cn\mathcal{A}_t(r)\ldots\mathcal{A}_{t-n+1}(r) \\ 0 & \mathcal{A}_t(r)\ldots\mathcal{A}_{t-n+1}(r) \end{pmatrix}$$

and $u \in [\frac{k-1}{M}, \frac{k}{M})$. By using the above we have

$$B_k(t,n)B_k(t,n)^T = \begin{pmatrix} (Cn^2+1)R_k(t,n) & CnR_k(t,n) \\ CnR_k(t,n) & R_k(t,n) \end{pmatrix},$$

where $R_k(t,n) = (\mathcal{A}_t(r)\ldots\mathcal{A}_{t-n+1}(r))(\mathcal{A}_t(r)\ldots\mathcal{A}_{t-n+1}(r))^T$. By choosing C_1 such that $C_1(n^2+1) \geq (Cn^2+1)$ and $C_1(n^2+1) \geq Cn$ for all n, we obtain

$$B_k(t,n)B_k(t,n)^T \leq C_1(n^2+1)\begin{pmatrix} R_k(t,n) & R_k(t,n) \\ R_k(t,n) & R_k(t,n) \end{pmatrix}.$$

It is clear the largest eigenvalue of the matrix above is $C_1(n^2+1)\|\mathcal{A}_t(r)\ldots\mathcal{A}_{t-n+1}(r)\|_{spec}$. Therefore by using the above we have

$$\|A_t(2,u)\ldots A_{t-n+1}(2,u)\|_{spec} \leq \|B_k(t,n)\|_{spec} \leq C_1(n^2+1)\|\mathcal{A}_t(r)\ldots\mathcal{A}_{t-n+1}(r)\|_{spec}. \quad (49)$$

It immediately follows that $\{A_t(2,u)\}_t$ has a negative Lyapounov exponent. The proof of (22) is similar, hence we omit this proof. $\Box$

References

Basrak, B., Davis, R., & Mikosch, T. (2002). Regular variation of GARCH processes. *Stochastic Processes and their Applications*, *99*, 95-115.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econometrics*, *31*, 301-327.

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

Bougerol, P., & Picard, N. (1992b). Strict stationarity of generalised autoregressive processes. *Ann. Probab.*, *20*, 1714-1730.

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

Crámer, H. (1961). On some classes of nonstationary stochastic processes. In *Proceedings of the Fourth Berkeley Symposium* (p. 57-78). Berkeley: University of California Press.

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

Dahlhaus, R., & Subba Rao, S. (2006). Statistical inference of time varying ARCH processes. *Ann. Statist.*, *34*, 1074-1114.

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

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, *50*, 987-1006.

Kesten, H., & Spitzer, F. (1984). Convergence in distribution for products of random matrices. *Z. Wahrsch. Verw. Gebiete*, *67*.

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

Nicholls, D., & Quinn, B. (1982). *Random Coefficient Autoregressive Models, An Introduction.* New York: Springer-Verlag.

Priestley, M. B. (1965). Evolutionary spectra and non-stationary processes. *J. Roy. Statist. Soc. Ser. B*, *27*, 204-37.

Straumann, D., & Mikosch, T. (2006). Quasi-maximum likelihood estimation in conditionally hetroscedastic time series: a stochastic recurrence equation approach. *Ann. Statist.*

Subba Rao, T. (1977). On the estimation of bilinear time series models. In *Bull. Inst. Internat. Statist. (paper presented at 41st session of ISI, New Delhi, India)* (Vol. 41).

Terdik, G. (1999). *Bilinear stochastic models and related problems of nonlinear time series analysis; a frequency domain approach* (Vol. 142). New York: Springer Verlag.

Tweedie, R. (1983). Criteria for rates of convergence of Markov chains, with application to queueing and storage theory. In J. Kingman & G. Reuter (Eds.), *Probability, statistics and analysis* (p. 260-277). Cambridge: Cambridge University Press.