On some nonstationary, nonlinear random processes and their stationary approximations

Suhasini Subba Rao
Institut für Angewandte Mathematik,
Universität Heidelberg, Germany
suhasini@statlab.uni-heidelberg.de

February 7, 2005

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 time-varying ARCH, GARCH, random coefficients ARCH and random coefficient AR(1) processes. 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 derivative processes obtained here satisfy alpha mixing properties.

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

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), Crámer (1961) and Bibi (2003). 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; and these are mainly restricted to stationary processes. Standard nonlinear models include ARCH (Engle, 1982), GARCH (Bollerslev, 1986) and Random coefficient processes Nicholls and 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 (2003).

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

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

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 give examples of various time-varying processes which admit representation (1).

In Section 3 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

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

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 $\tilde{\mathcal{X}}_t(u)$ can be regarded as a stationary approximation of $\mathcal{X}_{t,N}$ for t/N close to u. In Section 3 we define a 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. The idea of a stationary approximation and a derivative process was established for time-varying ARCH processes in Dahlhaus and Subba Rao (2003).

2 Some useful nonstationary models

In this section we show that various time-dependent nonstationary processes admit represention (1).

Let $A_{i,j}$ denote the (i,j)th element of the matrix A. 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.

Time-varying GARCH

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

$$X_{t,N} = Z_t \sigma_{t,N}, \quad \sigma_{t,N}^2 = a_0(\frac{t}{N}) + \sum_{i=1}^{p} a_i(\frac{t}{N}) X_{t-i,N}^2 + \sum_{j=1}^{q} b_j(\frac{t}{N}) \sigma_{t-j,N}^2, \qquad t = 1,\ldots,N \quad (3)$$

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

$$A_t(u) = \begin{pmatrix} \underline{\tau}_t(u) & b_q(u) & \underline{a}(u) & a_p(u) \\ I_{q-1} & 0 & 0 & 0 \\ \underline{Z}_{t-1}^2 & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{pmatrix}, \quad (4)$$

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

Time-varying Random coefficient GARCH

The process $\{X_{t,N}\}$ is called a time-varying random coefficient GARCH(p,q) process if it has the representation

$$X_{t,N} = Z_t \sigma_{t,N}, \quad \sigma_{t,N}^2 = a_0(\frac{t}{N})\{1 + \xi_{0,t}\} + \sum_{i=1}^{p} a_i(\frac{t}{N})\{1 + \xi_{i,t}\} X_{t-i,N}^2 + \sum_{j=1}^{q} b_j(\frac{t}{N})\{1 + \xi_{p+j,t}\} \sigma_{t-j,N}^2,$$

where $\underline{\xi}_t^T = (\xi_{0,t}, \ldots, \xi_{p+q,t})$ are independent identically distributed random vectors and independent of $\{Z_t^2\}$ (where Z_t^2 is defined as above), with the support of the density function of $\xi_{i,t}$ lying in the interval $[-1, \infty)$, for $i = 0, \ldots, p+q$.

$\{X_{t,N}\}$ has the state-space representation in (1), where $\mathcal{X}_{t,N}^T = (\sigma_{t,N}^2, \ldots, \sigma_{t-q+1,N}^2, X_{t-1,N}^2, \ldots, X_{t-p+1,N}^2)$, $\underline{b}_t(u)^T = (a_0(u)\{\xi_{t,0}+1\}, 0, \ldots, 0) \in \mathbb{R}^{p+q-2}$, $A_t(u)$ is a $(p+q-1) \times (p+q-1)$ matrix defined by

$$A_t(u) = \begin{pmatrix} \underline{\tau}_t(u) & b_q(u)[1 + \xi_{t,p+q}] & \underline{a}_t(u) & a_p(u) \\ I_{q-1} & 0 & 0 & 0 \\ \underline{Z}_{t-1}^2 & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{pmatrix}, \quad (5)$$

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

Time-varying random coefficient AR

The sequence of processes $\{X_{t,N}\}$ is called a time-varying random coefficient Autoregressive (1) (tv-random coefficient AR(1)) process if it satisfies the representation

$$X_{t,N} = \{a(\frac{t}{N}) + \xi_t\}X_{t-1,N} + \sigma(\frac{t}{N})\varepsilon_t,$$

where $\{\xi_t\}$ and $\{\varepsilon_t\}$ are independent, identically distributed variables with $\mathbb{E}(\xi_t) = 0$ and $\mathbb{E}(\varepsilon_t) = 0$. $X_{t,N}$ has the time-varying state-space representation (1) with $\mathcal{X}_{t,N} = X_{t,N}$, $A_t(u) = (a_1(u) + \xi_t)$ and $\underline{b}_t(u) = \sigma(u)\varepsilon_t$.

3 Nonlinear time-varying processes

3.1 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 $\|A\|_{spec}$ be the operator norm:

$$\|A\|_{spec} = \sup_{\|\underline{x}\|_2=1} \|A\underline{x}\|_2.$$

$|B|_{abs}$ denote the absolute vector or matrix of $B = (B)_{ij}$, 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 spectral radius of the matrix A, $\sup_u A(u)$ be defined as $\sup_u A(u) = \{\sup_u |A(u)|_{i,j} : i = 1,\ldots p, j = 1,\ldots,q\}$. 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

$$\inf\{\frac{1}{n}\mathbb{E}(\log \|A_tA_{t-1}\ldots A_{t-n+1}\|_{spec}) : n \in \mathbb{N}\}. \tag{6}$$

We make the following assumptions.

Assumption 3.1 *The sequence of stochastic processes $\{X_{t,N}\}$ 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 $r \in \{1,\ldots,M\}$, 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 $\varepsilon > 0$, $\mathbb{E}\|\mathcal{A}_t\|_1^{\varepsilon} < \infty$.

We will show that the processes defined in Section 2 satisfy the assumptions above. We will use the same notation as defined in Section 2.

Time-varying GARCH

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

$$\sup_u \Big\{ \sum_{i=1}^{p} a_i(u) + \sum_{j=1}^{q} b_j(u) \Big\} < (1 - \eta), \tag{7}$$

then Assumption 3.1 is satisfied for the tvGARCH(p, q) process. Using the β-Lipschitz continuity of the parameters we will show there exist matrices $\mathcal{A}_t(r)$ which bound $A_t(u)$ and satisfy Assumption 3.1(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 $M_\varepsilon = \varepsilon^{-1}$, then for each $r \in \{1, \ldots, M_\varepsilon\}$:

$$\alpha_i(r) = \big\{a_i[(k - 1)\varepsilon] + K_{\max}\varepsilon^{\beta}\big\} \text{ and } \beta_j(r) = \big\{b_j[(k - 1)\varepsilon] + K_{\max}\varepsilon^{\beta}\big\} \text{ for } r = 1, \ldots, M_\varepsilon, \tag{8}$$

$i = 1, \ldots, p$ and $j = 1 \ldots, q$. Therefore by using (7) and the above construction we have $\sup_{(r-1)\varepsilon < u \leq r\varepsilon} a_i(u) \leq \alpha_i(r)$, $\sup_{(r-1)\varepsilon < u \leq r\varepsilon} b_j(u) \leq \beta_j(r)$ and

$$\sum_{i=1}^{p} \sup_{(r-1)\varepsilon < u \leq r\varepsilon} a_i(u) + \sum_{j=1}^{q} \sup_{(r-1)\varepsilon < u \leq r\varepsilon} b_j(u) \leq \sum_{i=1}^{p} \alpha_i(r) + \sum_{j=1}^{q} \beta_j(r) \leq 1 - \eta/2. \tag{9}$$

Let

$$\mathcal{A}_t(r) = \begin{pmatrix} \tilde{\underline{\tau}}_t(r) & \beta_q(r) & \underline{\alpha}(r) & \alpha_p(r) \\ I_{q-1} & 0 & 0 & 0 \\ \underline{Z}_{t-1}^2 & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{pmatrix}, \tag{10}$$

$\underline{\tilde{\tau}}_t(r) = (\beta_1(r) + \alpha_1(r)Z_{t-1}^2, \beta_2(r), \ldots, \beta_{q-1}(r))$, $\underline{\alpha}(r) = (\alpha_2(r), \ldots, \alpha_{p-1}(r))$ and $\underline{Z}_{t-1}^2 = (Z_{t-1}^2, 0, \ldots, 0) \in \mathbb{R}^{p-1}$. Then it is clear that

$$\sup_{(r-1)\varepsilon < u \leq r\varepsilon} |A_t(u)|_{abs} \leq \mathcal{A}_t(r).$$

To summarise, we have partitioned the unit interval into M_ε 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 3.1, that $\lambda_{spec}\{\mathbb{E}(\mathcal{A}_t(r))\} \leq (1-\eta/2)^{1/(p+q-1)}$. By using (Kesten & Spitzer, 1984), equation (1.4), we can show that $\{A_t(r)\}$ has a negative Lyapunov exponent.

Let $\tilde{\underline{b}}_t^T = (\sup_u a_0(u), 0, \ldots, 0) \in \mathbb{R}^{p+q-2}$. Since, $\mathbb{E}(Z_t^2) = 1$ we have for some $K > 0$ $\mathbb{E}(\log \|\mathcal{A}_t(r)\|_{spec}) \leq \mathbb{E}(\|\mathcal{A}_t(r)\|_{spec}) \leq K\mathbb{E}\|\mathcal{A}_t(r)\|_1 < \infty$. Therefore all the conditions of Assumption 3.1(i) are satisfied.

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

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

where $\mathcal{A}_t = (1 + Z_{t-1}^2)\mathbb{I}_{(p+q-1)\times(p+q-1)}$. Thus Assumption 3.1(ii) is satisfied.

Time-varying Random coefficient GARCH

Let us consider the tvRandom coefficient GARCH(p,q) process. We now show if $\mathbb{E}(Z_t^2) = 1$, the parameters $\{a_i(\cdot)\}$ and $\{b_j(\cdot)\}$ are β-Lipschitz continuous and satisfy

$$\sup_u \Big\{ \sum_{i=1}^p a_i(u)[1 + \mathbb{E}(\xi_{t,i})] + \sum_{j=1}^q b_j(u)[1 + \mathbb{E}(\xi_{t,p+j})] \Big\} < (1-\eta), \tag{11}$$

and $\inf_u a_0(u)[1 + \mathbb{E}(\xi_{t,0})] > \rho_0$ for some $\rho_0 > 0$, then Assumption 3.1 is satisfied.

Using a similar method given above (for the tvGARCH process) we will show that there exists random matrices $\{\mathcal{A}_t(r)\}$ which dominate $A_t(u)$ and satisfy Assumption 3.1, we now outline the method.

Let K_{max} be such that $\sup_{u,v} |a_i(u) - a_i(v)| \leq K_{\max}|u-v|^\beta$ and $\sup_{u,v} |b_j(u) - b_j(v)| \leq K_{\max}|u-v|^\beta$. Let $\tilde{\mu} = \max_{1 \leq i \leq p+q} \mathbb{E}(1+\xi_{t,i})$ and 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 $\{\alpha_i(r)\}$ and $\{\beta_j(r)\}$ be defined as in (9) but with the new M_ε and ε. By using a similar construction to that given in the tvGARCH case we can define the matrices $\mathcal{A}_t(r)$ (which includes the additional $\xi_{t,i}$ terms), which satisfy for $u \in (\frac{r-1}{M}, \frac{r}{M}]$, that $A_t(u) \leq \mathcal{A}_t(r)$. Since

$$\Big(\sum_{i=1}^p \alpha_i(r)\{1 + \mathbb{E}(\xi_{t,i})\} + \sum_{j=1}^q \beta_j(r)\{1 + \mathbb{E}(\xi_{t,p+i})\}\Big) < 1 - \eta/2,$$

by using the same arguments given for the tvGARCH process we have that $\{\mathcal{A}_t(r)\}$ has a negative Lyapunov exponent. Hence Assumption 3.1(i) is satisfied. Moreover it is straightforward to show that Assumption 3.1(ii) satisfied.

Time-varying Random coefficient AR

Under the condition that $a(\cdot)$ and $\sigma(\cdot)$ are β-Lipschitz continuous, $\{\xi_t\}$ and $\{\varepsilon_t\}_t$ are iid random variables with $\mathbb{E}(\xi_t) = 0, \mathbb{E}(\varepsilon_t) = 0$ and $\{\sup_u |a(u)| + \mathbb{E}(|\xi_t|)\} < 1 - \eta$ (for some $\eta > 0$) it is straightforward to show that Assumption 3.1 is satisfied.

3.2 The stationary approximation

By showing that the random matrices $\prod_{i=0}^{k} A_{t-i}(\frac{t-i}{N})$ decay exponentially, we will prove that (1) has a unique solution. Typically, this is achieved by showing that the random transition matrix associated with the model has a negative Lyapunov exponent. Under Assumption 3.1, $\{A_t(\frac{t}{N})\}_t$ has a negative Lyapunov exponent, which we use to show that $\{\mathcal{X}_{t,N}\}_t$ has a unique, well defined solution.

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

Proposition 3.1 *Suppose Assumption 3.1 holds, and let $\mathcal{X}_{t,N}$ be defined as in (1). Then*

$$\mathcal{X}_{t,N} = \sum_{k=0}^{\infty} A_t(\frac{t}{N}) \cdots A_{t-k+1}(\frac{t-k+1}{N})\underline{b}_{t-k}(\frac{t-k}{N}) \tag{12}$$

converges almost surely and is the unique solution of (1). Moreover the series

$$\tilde{\mathcal{X}}_t(u) = \sum_{k=0}^{\infty} A_t(u) \cdots A_{t-k+1}(u)\underline{b}_{t-k}(u) \tag{13}$$

converges almost surely and is the unique, stationary solution of (2).

PROOF. A formal expansion of (1) gives (12). We now show that (12) almost surely converges. Let $I_r = (\frac{r-1}{M}, \frac{r}{M}]$, $J_r = \{k \geq 0 : \frac{t-k}{N} \in I_r\}$ and $J_{r,l} = J_r \cap \{0, 1, \ldots, l-1\}$. Then we have

$$|\mathcal{X}_{t,N}|_{abs} \leq \sum_{k=0}^{\infty} \{\prod_{i=0}^{k-1} |A_{t-i}(\frac{t-i}{N})|_{abs}\}\underline{\tilde{b}}_{t-k} \leq \sum_{k=0}^{\infty} \{\prod_{r=1}^{M} \prod_{i \in J_{r,k}} \mathcal{A}_{t-i}(r)\}\underline{\tilde{b}}_{t-k}. \tag{14}$$

Therefore by using Lemma A.2 we can show that the series (14) converges almost surely. The almost sure convergence of $|\mathcal{X}_{t,N}|_{abs}$ implies the convergence of $\mathcal{X}_{t,N}$. Finally, using the same arguments given in Bougerol and Picard (1992b), Theorem 2.5, we can show that (12) is the unique solution of (1).

If $u \in (\frac{r-1}{M}, \frac{r}{M}]$ then $|A_t(u)|_{abs} \leq \mathcal{A}_t(r)$ and we have $\|A_t(u) \dots A_{t-k}(u)\|_{spec} \leq \|\mathcal{A}_t(r) \dots \mathcal{A}_{t-k}(r)\|_{spec}$. Thus (13) converges almost surely and is the unique, stationary solution of (2). $\square$

In the theorem below we will show that $\mathcal{X}_{t,N}$ can locally be approximated by the stochastic process $\tilde{\mathcal{X}}_t(u)$.

Let

$$\underline{Y}_t = \sum_{k=1}^{\infty} \{ \sum_{r=1}^{M} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(r) \} \underline{\tilde{b}}_{t-k}. \tag{15}$$

Theorem 3.1 *Suppose Assumption 3.1 holds, let $\mathcal{X}_{t,N}$, $\tilde{\mathcal{X}}_t(u)$ and $\underline{Y}_t$ be defined as in (1), (2) and(15) respectively and suppose there exists an $\varepsilon > 0$ such that*

$$\sup_{t,N} \mathbb{E}\|\mathcal{X}_{t,N}\|_1^{\varepsilon} < \infty \text{ and } \sup_{u} \mathbb{E}\|\underline{Y}_t\|_1^{\varepsilon} < \infty. \tag{16}$$

Then for any $t \in \{1, \dots, N\}$ we have

$$|\mathcal{X}_{t,N} - \tilde{\mathcal{X}}_t(\frac{t}{N})|_{abs} \leq \frac{1}{N^{\beta}} \underline{V}_{t,N}, \tag{17}$$

$$|\tilde{\mathcal{X}}_t(u) - \tilde{\mathcal{X}}_t(w)|_{abs} \leq |u - w|^{\beta} \underline{W}_t \tag{18}$$

$$\text{and } |\mathcal{X}_{t,N} - \tilde{\mathcal{X}}_t(u)|_{abs} \leq |\frac{t}{N} - u|^{\beta} \underline{W}_t + \frac{1}{N^{\beta}} \underline{V}_{t,N} \tag{19}$$

where

$$\underline{V}_{t,N} = C \sum_{k=1}^{\infty} k \{ \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(i_1) \} \{ \mathcal{A}_{t-k} |\mathcal{X}_{t-k-1,N}|_{abs} + \underline{\tilde{b}}_{t-k} \}, \tag{20}$$

$$\underline{W}_t = C \sum_{k=1}^{\infty} \{ \sum_{r=1}^{M} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(r) \} \{ \mathcal{A}_{t-k} \underline{Y}_{t-k-1} + \underline{\tilde{b}}_{t-k} \} \tag{21}$$

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. $\square$

The keypoint is (19). In other words the deviation between the nonstationary process $\mathcal{X}_{t,N}$ and the stationary process $\tilde{\mathcal{X}}_t(u)$ depends on the difference $|t/N - u|$. We summarise this result in the following corollary.

Corollary 3.1 *Let $\mathcal{X}_{t,N}$ and $\tilde{\mathcal{X}}_t(u)$ be defined as in (1) and (2). Suppose the assumptions in Theorem 3.1 hold. Then for any $t \in \{1, \ldots, N\}$ we have*

$$|\mathcal{X}_{t,N} - \tilde{\mathcal{X}}_t(u)|_{abs} \leq |\frac{t}{N} - u|^{\beta} O_p(1) + \frac{1}{N^{\beta}} O_p(1). \tag{22}$$

A simple application of the theorem above is in the evaluation of the sampling properties of 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 (19) we have

$$\frac{1}{2M+1} \sum_{k=-M}^{M} \mathcal{X}_{t_0+k,N} = \frac{1}{2M+1} \sum_{k=-M}^{M} \tilde{\mathcal{X}}_{t_0+k}(u_0) + \mathcal{B}_{t_0,N}, \tag{23}$$

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 some moments of $\mathcal{X}_{t,N}$ and its related processes.

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.

Let $\|\underline{X}\|_n^E = \{\sum_{i=1}^{p} \mathbb{E}(|X_i|^n)\}^{1/n}$, where $\underline{X}^T = (X_1, \ldots, X_p)$. Let A be a $p \times q$ matrix, where $A = (a_{ij} : i = 1, \ldots, p, j = 1, \ldots, q)$. We define the matrix function $[\cdot]_n : \mathbb{R}^{p \times q} \to \mathbb{R}^{p \times q}$ as

$$[A]_n = \left\{ \left[\mathbb{E}(|a_{ij}|^n)\right]^{1/n} : i = 1, \ldots, p, j = 1, \ldots, q \right\}. \tag{24}$$

We now give conditions for $\|\mathcal{X}_{t,N}\|_n^E < \infty$ and $\|\tilde{\mathcal{X}}_t(u)\|_n^E < \infty$.

Proposition 3.2 *Suppose Assumption 3.1 holds, let $\mathcal{X}_{t,N}$, $\tilde{\mathcal{X}}_t(u)$, $\underline{V}_{t,N}$ and $\underline{W}_t$ be defined as in (1), (2), (20) and (21) respectively. Suppose for all $r \in \{1, \ldots, M\}$ and for some $n \in [1, \infty)$, that $\|\tilde{\underline{b}}_t\|_n^E < \infty$ and, for some $\delta > 0$, $\lambda_{spec}\{[A_t(r)]_n\} < 1 - \delta$. Then*

$$\|\mathcal{X}_{t,N}\|_n^E < \infty \text{ and } \|\underline{V}_{t,N}\|_n^E < \infty \text{ uniformly in } t \text{ and } N, \tag{25}$$

$$\|\tilde{\mathcal{X}}_t(u)\|_n^E < \infty \textit{ uniformly in } u, \tag{26}$$

$$\textit{and } \|\underline{W}_t\|_n^E < \infty. \tag{27}$$

We use the proposition above to show that the nth moment of the tvGARCH process exists under appropriate conditions. To do this we need the following lemma.

Lemma 3.1 *Let $\mu > 1$ and $\{a_i : i = 1, \ldots, p\}$, $\{b_j : j = 1 \ldots, q\}$ be positive sequences and*

$$A = \begin{pmatrix} \underline{\tau} & b_q & \underline{a} & a_p \\ I_{q-1} & 0 & 0 & 0 \\ \underline{\mu} & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{pmatrix}, \tag{28}$$

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 $\lambda_{spec}(A) \leq (1-\delta)^{1/(p+q-1)}$.

PROOF. The proof is close to the proof of Bougerol and Picard (1992a), Corollary 2.2. By expanding the determinant of A with respect to the last column and arguing by recursion on p we have that

$$\mathrm{Det}(A - zI) = z^{p+q-1}(1 - \sum_{i=1}^p a_i A_i z^{-i} - \sum_{j=1}^q b_j B_j z^{-j})$$

where $A_i \in \{1, \mu\}$ and $B_j \in \{1, \mu\}$. By using the inequality $|a - b| > ||a| - |b||$ we have

$$|\mathrm{Det}(A - zI)| \geq |z^{p+q-1}|\big|1 - \sum_{i=1}^p a_i A_i |z^{-i}| - \sum_{j=1}^q b_j B_j |z^{-j}|\big|.$$

Since $A_i \geq 1$, $B_j \geq 1$, $(1-\delta)^{-i/(p+q-1)} \geq (1-\delta)^{-1}$ (if $1 \leq i \leq p+q-1$) and $\mu(\sum_{i=1}^p a_i + \sum_{j=1}^q b_j) < 1 - \delta$, for $|z| \geq (1-\delta)^{1/(p+q-1)}$ we have

$$\begin{aligned} |\mathrm{Det}(A - zI)| &\geq (1-\delta)\big\{1 - \sum_{i=1}^p a_i A_i (1-\delta)^{-i/(p+q-1)} - \sum_{j=1}^q b_j B_j (1-\delta)^{-j/(p+q-1)}\big\} \\ &\geq (1-\delta)\big\{1 - \mu \sum_{i=1}^p a_i (1-\delta)^{-1} - \mu \sum_{j=1}^q b_j (1-\delta)^{-1}\big\} > 0. \end{aligned}$$

Since $|\mathrm{Det}(A - zI)| > 0$ for all z which satisfy $|z| \geq (1-\delta)^{1/(p+q-1)}$, then we have that $\lambda_{spec}(A) < (1-\delta)^{1/(p+q-1)}$. □

Example 3.1 *We consider again the tvGARCH process, under the tvGARCH assumptions stated earlier. We have shown that Assumption 3.1 is satisfied. Therefore the conditions of Proposition 3.1 are satisfied, which implies that the tvGARCH(p, q) process has a unique well defined solution.*

Suppose $n \in [1, \infty)$ and let $\mu_n = \{\mathbb{E}(Z_t^{2n})\}^{1/n}$. In addition we assume

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

Then by verifying the conditions in Proposition 3.2 we will show that $\sup_{t,N} \mathbb{E}(X_{t,N}^{2n}) < \infty$. We will show that there exists matrices $\{\mathcal{A}_t(r)\}$ which satisfy Assumption 3.1 (that is bound $A_t(u)$) and where $\lambda_{spec}\{[\mathcal{A}_t(r)]_n\} < (1 - \eta/2)^{1/(p+q-1)}$.

The method is the same as that given in (9), here we let $\varepsilon \leq \big(\eta/(2\mu_n K_{\max}(p+q))\big)^{1/\beta}$, and define M_ε ε, $\alpha_i(r)$ and $\beta_j(r)$ (defined in (8)) and $\mathcal{A}_t(r)$ as in (10), using the new ε. It is straightforward to show that

$$\sup_u \mu_n \Big\{ \sum_{i=1}^{p} \alpha_i(r) + \sum_{j=1}^{q} \beta_j(r) \Big\} < 1 - \eta/2. \tag{29}$$

We define the deterministic matrix $\mathcal{A}(r)^*$ as follows

$$\mathcal{A}(r)^* = \begin{pmatrix} \underline{\tau}(r) & \beta_q(r) & \underline{\alpha}(r) & a_p(r) \\ I_{q-1} & 0 & 0 & 0 \\ \underline{\mu} & 0 & 0 & 0 \\ 0 & 0 & I_{p-2} & 0 \end{pmatrix}, \tag{30}$$

where $\underline{\tau}(r) = (\beta_1(r) + \alpha_1(r)\mu_n, \beta_2(r), \ldots, \beta_{q-1}(r)) \in \mathbb{R}^{q-1}$, $\underline{\alpha}(r) = (\alpha_2(r), \ldots, \alpha_{p-1}(r)) \in \mathbb{R}^{p-2}$ and $\underline{\mu} = (\mu_n, 0, \ldots, 0) \in \mathbb{R}^{q-1}$. It is clear that $[\mathcal{A}_t(r)]_r \leq \mathcal{A}(r)^*$, therefore since $[\mathcal{A}_t(r)]_r$ is a positive matrix, we have that $\lambda_{spec}\{[\mathcal{A}_t(r)]_n\} \leq \lambda_{spec}\{\mathcal{A}(r)^*\}$. Now by using Lemma 3.1 and (29) we then have that $\lambda_{spec}\{\mathcal{A}(r)^*\} \leq (1-\eta/2)^{1/(p+q-1)}$. Thus the conditions of Proposition 3.2 are satisfied and we have $\sup_{t,N} \|\mathcal{X}_{t,N}\|_n^E < \infty$, which implies $\sup_{t,N} \mathbb{E}(X_{t,N}^{2n}) < \infty$.

We now define a stationary process, which locally approximates the tvGARCH process. Let us suppose the stationary GARCH process $\{\tilde{X}_t(u)\}$ satisfies

$$\tilde{X}_t(u) = \big\{ a_0(u) + \sum_{i=1}^{p} a_i(u) \tilde{X}_{t-i}(u)^2 + \sum_{j=1}^{q} b_j(u) \tilde{\sigma}_{t-j}(u)^2 \big\} Z_t^2. \tag{31}$$

It is straightforward to show that (31) has a unique, stationary solution and that $\{\tilde{X}_t(u)\}$ satisfies the state-space representation given in (2).

The conditions of Theorem 3.1 are fulfilled with the tvGARCH process defined as in (3) and the stationary process defined in (31). Also, by the above, $\sup_{t,N}\|\mathcal{X}_{t,N}\|_n^E < \infty$, *thus condition (16) holds. Therefore by using Theorem 3.1 and Proposition 3.2 we have*

$$\mathcal{X}_{t,N} = \tilde{\mathcal{X}}_t(u) + (|\frac{t}{N} - u|^\beta + \frac{1}{N^\beta})R_{t,N}, \qquad \text{where} \qquad \sup_{t,N}\|R_{t,N}\|_n^E < \infty.$$

Example 3.2 *The process* $\{X_{t,N}\}$ *is called a time-varying Autoregressive Conditional Heteroscedasticity (p) (tvARCH(p)) process if is satisfies*

$$X_{t,N} = Z_t\sigma_{t,N}, \quad \sigma_{t,N}^2 = a_0(\frac{t}{N}) + \sum_{i=1}^{p} a_i(\frac{t}{N})X_{t-i,N}^2,$$

where $\mathbb{E}(Z_t^2) = 1$. *It is clear that tvARCH processes are a subclass of tvGARCH processes, so they have the state space representation in Section 2, with parameters* $b_j(u) = 0$ *for all* j. *It also admits its own representation in (1), with* $\mathcal{X}_{t,N}^T = (X_{t,N}^2, \ldots, X_{t-p+1,N}^2)$, $\underline{b}_t(u)^T = (a_0(u)Z_t^2, 0, \ldots, 0) \in \mathbb{R}^p$ *and*

$$A_t(u) = \begin{pmatrix} \underline{a}_t(u) & a_p(u)Z_t^2 \\ I_{p-1} & O_{p-1} \end{pmatrix},$$

where $\underline{a}_t(u) = (a_1(u)Z_t^2, \ldots, a_{p-1}(u)Z_t^2) \in \mathbb{R}^{p-1}$. *Suppose* $n \in [0,\infty)$ *and let* $\mu_n = \mathbb{E}(Z_t^{2n})^{1/n}$. *If* $\sup_u \mu_n\{\sum_j a_j(u)\} < (1-\eta)$, *then by using the same arguments as given in Lemma 3.1, it can be shown that* $\sup_u \lambda_{spec}\{[A_t(u)]_n\} < (1-\eta)^{1/p}$. *By using Moulines et al. (2003), Lemma 15, for all* k, $u \in [0,1)$, *we have*

$$\left\|[\prod_{i=0}^{k-1} A_{t-i}(\frac{t-i}{N})]_n\right\|_{spec} \leq K(1-\delta)^k \quad \text{and} \quad \left\|[\prod_{i=0}^{k-1} A_{t-i}(u)]_n\right\|_{spec} \leq K(1-\delta)^k \tag{32}$$

for some $0 < \delta < \eta/2$.

We now return to the local average example in (23). Under the assumption that all the conditions in Proposition 3.2 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

$$\frac{1}{2M+1}\sum_{k=-M}^{M}\mathcal{X}_{t_0+k,N} = \frac{1}{2M+1}\sum_{k=-M}^{M}\tilde{\mathcal{X}}_{t_0+k}(u_0) + ((\frac{M}{N})^\beta)O_p(1) + \frac{1}{N^\beta}O_p(1). \tag{33}$$

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

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

where $M \to 0$, $M/N \to 0$ as $N \to \infty$. In the following section we derive the so called 'derivative' process which allows us to obtain a tighter bound for $\mathcal{B}_{t_0,N}$.

4 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 (22) and show that a Taylor series expansion of the time-varying process in terms of stationary processes, can be derived (Theorem 4.2). In order to do this we define the so called '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 $\{\frac{\partial A_t(u)}{\partial u}\rfloor_{u=t/N}\}_t$ 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

$$\frac{\partial A(u)}{\partial u} = \Big\{\frac{\partial A(u)_{i,j}}{\partial u} : i = 1, \ldots, p, j = 1, \ldots, q\Big\}.$$

We say that the function $f \in Lip(d + \beta')$ if there exists a K such that $\sup_{u,v} |f^{(d)}(u) - f^{(d)}(v)| \leq K|u - v|^{\beta'}$.

We make the following assumptions.

Assumption 4.1 *The sequence of stochastic processes $\{\mathcal{X}_{t,N}\}$ have a time-varying state-space representation defined in (1), $\{\mathcal{A}_t(r)\}$, $\{\tilde{\underline{b}}_t\}$ and $\{\mathcal{A}_t\}$ are defined as in Assumption 3.1, and $\{A_t(u)\}$ and $\{\underline{b}_t(u)\}$ satisfy the following properties:*

(i) The process $\{\mathcal{X}_{t,N}\}$ satisfies Assumption 3.1 with $\beta = 1$.

(ii) Let $d \in \mathbb{Z}^+$ and $\beta' > 0$. The matrices $\{A_t(\cdot)\}$ and $\{\underline{b}_t(\cdot)\}$ are assumed to satisfy for $n = 0, \ldots, d-1$:

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

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

$$\sup_u |\frac{\partial^n A_t(u)}{\partial u^n}|_{abs} \leq C\mathcal{A}_t(r) \quad \text{if } u \in (\frac{r-1}{M}, \frac{r}{M}] \text{ and } \sup_u |\frac{\partial^n \underline{b}_t(u)}{\partial u^n}|_{abs} \leq C\tilde{\underline{b}}_t,$$

where $C < \infty$. Therefore $\{A_t(u)\}_t$ and $\{\underline{b}_t(u)\}_t$ belong to the Lipschitz class Lip$(d+\beta')$. This is a kind of Hölder continuity of order $d + \beta'$ for random matrices.

Example 4.1 *We now consider the tvGARCH, tv Random coefficient GARCH and tv random coefficient AR(1) processes. Suppose the parameters of each of the above processes belong to Lip$(d+\beta')$, then it is straightforward to show that Assumption 4.1 is satisfied for each of the above mentioned processes.*

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

$$\frac{\partial \tilde{\mathcal{X}}_t(u)}{\partial u} = \frac{\partial A_t(u)}{\partial u}\tilde{\mathcal{X}}_{t-1}(u) + A_t(u)\frac{\partial \tilde{\mathcal{X}}_{t-1}(u)}{\partial u} + \frac{\partial \underline{b}_t(u)}{\partial u}. \tag{34}$$

An interesting feature about the derivative process is that its existence requires only weak assumptions on the derivative matrix $\frac{\partial A_t(u)}{\partial u}$. In other words, given that $\mathcal{X}_{t,N}$ is well defined, the existence of $\frac{\partial A_t(u)}{\partial u}$ is sufficient for the derivative process to be well defined.

We will show in Corollary 4.3 that $\frac{\partial \tilde{\mathcal{X}}_t(u)}{\partial u}$ is a well defined, stationary, random process. To prove this assertion we use the following representation. Let $\tilde{\mathcal{X}}_t(2,u)^T = (\frac{\partial \tilde{\mathcal{X}}_t(u)^T}{\partial u}, \tilde{\mathcal{X}}_t(u)^T)$. Then $\{\mathcal{X}_t(2,u)\}$ satisfies the representation

$$\tilde{\mathcal{X}}_t(2,u) = A_t(2,u)\tilde{\mathcal{X}}_{t-1}(2,u) + \underline{b}_t(2,u), \tag{35}$$

where

$$A_t(2,u) = \begin{pmatrix} A_t(u) & \frac{\partial A_t(u)}{\partial u} \\ 0 & A_t(u) \end{pmatrix} \text{ and } \underline{b}_t(2,u) = \begin{pmatrix} \underline{b}_t'(u) \\ \underline{b}_t(u) \end{pmatrix}. \tag{36}$$

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

$$\mathcal{X}_{t,N}(2) = A_t(2, \frac{t}{N})\mathcal{X}_{t-1,N}(2) + \underline{b}_t(2, \frac{t}{N}), \tag{37}$$

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

We now consider the processes $\{\tilde{\mathcal{X}}_t(2,u)\}_t$ and $\{\mathcal{X}_{t,N}(2)\}_t$. We need the following definitions and lemma. Let $\underline{\tilde{b}}_t(2)^T = (C\underline{\tilde{b}}_t^T, \underline{\tilde{b}}_t^T)$,

$$\mathcal{A}_t(2,r) = \begin{pmatrix} \mathcal{A}_t(r) & C\mathcal{A}_t(r) \\ 0 & \mathcal{A}_t(r) \end{pmatrix}, \quad \mathcal{A}_t(2) = \begin{pmatrix} \mathcal{A}_t & C\mathcal{A}_t \\ 0 & \mathcal{A}_t \end{pmatrix} \tag{38}$$

where C is defined as in Assumption 4.1.

We now show that the random matrices $\{A_t(2,u)\}_t$ have similar properties to that of $\{A_t(u)\}_t$, which implies that the results in Section 3 can be applied to the processes $\{\tilde{\mathcal{X}}_t(2,u)\}_t$, $\{\mathcal{X}_{t,N}(2)\}_t$ and the derivative process $\{\frac{\partial \tilde{\mathcal{X}}_t(u)}{\partial u}\}_t$. In the following lemma we exploit the block triangular structure of $\{A_t(2,u)\}_t$ to show that $\{A_t(2,u)\}_t$ has a negative Lyapunov exponent.

Lemma 4.1 *Suppose Assumption 4.1 holds with $d = 1$. Then for $r = 1, \ldots, M$, we have*

$$\inf\{\frac{1}{n}\mathbb{E}(\log \|A_t(2,u) \ldots A_{t-n+1}(2,u)\|_{spec})\} < 0, \tag{39}$$

$$\inf\{\frac{1}{n}\mathbb{E}(\log \|\mathcal{A}_t(2,r) \ldots \mathcal{A}_{t-n+1}(2,r)\|_{spec})\} < 0. \tag{40}$$

PROOF. In the Appendix $\square$

Theorem 4.1 *Suppose Assumption 4.1 holds with $d = 1$ and let $\{\mathcal{X}_{t,N}(2)\}$ be defined as in (37). Then the process $\{\mathcal{X}_{t,N}(2)\}$ satisfies Assumption 3.1 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 4.1 we have

$$|A_t(2,u)|_{abs} \leq \mathcal{A}_t(2,r) \text{ if } u \in (\frac{r-1}{M}, \frac{r}{M}] \text{ and } \sup_u |\underline{b}_t(2,u)|_{abs} \leq \underline{\tilde{b}}_t(2).$$

Furthermore, by using Lemma 4.1, the random matrix sequences $\{A_t(2,u)\}_t$ and $\{\mathcal{A}_t(2,r)\}_t$ have negative Lyapunov exponents. This implies that the conditions in Assumption 3.1(i) are verified. Finally Assumption 4.1(ii) implies that Assumption 3.1(ii) holds with $A_t(2,u)$ and $\mathcal{A}_t(2,r)$, thus we obtain the result. $\square$

By using Theorem 4.1 we obtain the following corollaries.

Corollary 4.1 *Suppose Assumption 3.1 holds and let $\{\tilde{\mathcal{X}}_t(2,u)\}_t$ and $\{\mathcal{X}_{t,N}(2)\}_t$ be defined as in (35) and (37). Then we have*

$$\tilde{\mathcal{X}}_t(2,u) = \sum_{k=0}^{\infty} \Big(\prod_{j=0}^{k-1} A_t(2,u)\Big) \underline{b}_{t-k}(2,u) \quad \textit{which} \tag{41}$$

converge almost surely and are the unique solutions of (35). Furthermore

$$\mathcal{X}_{t,N}(2) = \sum_{k=0}^{\infty} A_t(2,\frac{t}{N}) \ldots A_{t-k+1}(2,\frac{t-k+1}{N}) \underline{b}_{t-k}(2,\frac{t-k}{N}) \tag{42}$$

converges almost surely and is the unique solution of (37).

PROOF. We can use Theorem 4.1, Proposition 3.1 and Theorem A.1 to prove the result. $\square$

We now show that the time-varying derivative process can locally be approximated by the stationary derivative process. Let

$$\underline{Y}_t(2) = \sum_{k=1}^{\infty} \Big\{ \sum_{r=1}^{M} \prod_{j=0}^{k-1} \mathcal{A}_{t-j}(2,r) \Big\} \underline{\tilde{b}}_{2,t-k}. \tag{43}$$

Corollary 4.2 *Suppose Assumption 4.1 holds with $d=1$. Let $\mathcal{X}_{t,N}(2)$, $\tilde{\mathcal{X}}_t(2,u)$, $\mathcal{A}_t(2,r)$, $\mathcal{A}_t(2)$, $\underline{\tilde{b}}_t(2)$, $\underline{W}_t$ and $Y_t(2)$ be defined as in (37), (35), (38), (21), (43) 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 for any $t \in \{1,\ldots,N\}$

$$|\mathcal{X}_{t,N}(2) - \tilde{\mathcal{X}}_t(2,\frac{t}{N})|_{abs} \leq \frac{1}{N^{\beta'}} \underline{V}_{t,N}(2), \tag{44}$$

$$|\tilde{\mathcal{X}}_t(2,u) - \tilde{\mathcal{X}}_t(2,w)|_{abs} \leq |u-w|^{\beta'} \underline{W}_t(2) \tag{45}$$

$$|\mathcal{X}_{t,N}(2) - \tilde{\mathcal{X}}_t(2,u)|_{abs} \leq |\frac{t}{N} - u|^{\beta'} \underline{W}_t(2) + \frac{1}{N^{\beta'}} \underline{V}_{t,N}(2), \tag{46}$$

and

$$\|\underline{W}_t\|_1 \leq \|\underline{W}_t(2)\|_1, \tag{47}$$

where $\{\underline{V}_{t,N}(2)\}$ *is similar to* $\{\underline{V}_{t,N}\}$ *defined in (20), 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*

$$\underline{W}_t(2) = C\sum_{r=1}^{M}\sum_{k=1}^{\infty}\big\{\prod_{j=0}^{k-1}\mathcal{A}_{t-j}(2,r_1)\big\}\big\{\mathcal{A}_{t-k}(2)|\underline{Y}_{t-k-1}(2)|_{abs} + \underline{\tilde{b}}_{t-k}(2)\big\}. \tag{48}$$

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 4.1 and by using Theorem 4.1 all the conditions in Theorem 3.1 are satisfied, therefore we obtain (44), (45) and (46). Since $\underline{W}_t$ is a sub-vector of $\underline{W}_t(2)$ we obtain (47). $\square$

Suppose the assumptions of Corollary 4.2 hold, then by using Corollary 3.1 with $\beta = 1$ we have

$$\|\mathcal{X}_{t,N} - \tilde{\mathcal{X}}_t(u)\|_1 \leq (|\frac{t}{N} - u| + \frac{1}{N})\,(\|\underline{W}_t\|_1 + \|V_{t,N}\|_1)\,.$$

Our object is to use the derivative process to obtain an exact expression for the difference above. To do this we bound the derivative process.

Corollary 4.3 *Suppose Assumption 3.1 holds with* $d = 1$. *Let* $\{\tilde{\mathcal{X}}_t(2,u)\}$ *and* $\{\underline{W}_t(2)\}_t$ *be defined as in (35) and (48) respectively. Then*

$$\begin{aligned}\frac{\partial\tilde{\mathcal{X}}_t(u)}{\partial u} \quad &= \quad \sum_{k=0}^{\infty}\sum_{r=0}^{k-1}\Big\{[\prod_{i=0}^{r-1}A_{t-i}(u)]\frac{\partial A_{t-r}(u)}{\partial u}[\prod_{i=r+1}^{k-1}A_{t-i}(u)]\Big\}\underline{b}_{t-k}(u) + \\ &\quad + \sum_{k=1}^{\infty}\{\prod_{i=0}^{k-1}A_{t-i}(u)\}\frac{\partial\underline{b}_{t-k}(u)}{\partial u},\end{aligned} \tag{49}$$

is almost surely the unique well defined solution of (34). Furthermore,

$$\sup_{u,v}\Big|\frac{\partial\tilde{\mathcal{X}}_t(u)}{\partial u} - \frac{\partial\tilde{\mathcal{X}}_t(v)}{\partial v}\Big|_{abs} \leq |u-v|^{\beta'}\{W_t(2)\}_{1,\ldots,p}, \tag{50}$$

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

PROOF. Let

$$\begin{aligned}\underline{Y}_t(u) &= \sum_{k=0}^{\infty}\sum_{r=0}^{k-1}\left\{[\prod_{i=0}^{k-1}A_{t-i}(u)]A'_{t-r}(u)[\prod_{i=r+1}^{k-1}A_{t-i}(u)]\right\}\underline{b}_{t-k}(u)+\\ &\quad+\sum_{k=1}^{\infty}\{\prod_{i=0}^{k-1}A_{t-i}(u)\}\underline{b}'_{t-k}(u),\end{aligned}$$

By formally expanding (34) and using Lemma A.2 we have that $\underline{Y}_t(u)$ is the unique well defined solution of the difference differential equation (34).

Let $\mathcal{N}$ be a set of measure zero such that for all $\omega \in \mathcal{N}^c$ $\underline{Y}_t(2,\omega)$ converges. Therefore for $\omega \in \mathcal{N}^c$ we have

$$\sup_u \|\tilde{\mathcal{X}}_t(u,\omega)\|_1 \leq \|\underline{Y}_t(2,\omega)\|_1 < \infty, \quad \sup_u \|\underline{Y}_t(u,\omega)\|_1 \leq \sum_{r=1}^{M}\|\underline{Y}_t(2,\omega)\|_1 < \infty,$$

hence almost surely we can interchange the derivative with the sum in $\tilde{\mathcal{X}}_t(u)$. This implies that $\underline{Y}_t(u)$ is almost surely the derivative of $\tilde{\mathcal{X}}_t(u)$ ($\frac{\partial \tilde{\mathcal{X}}_t(u)}{\partial u} = \underline{Y}_t(u)$).

Since $\frac{\partial \tilde{\mathcal{X}}_t(u)}{\partial u} = \tilde{\mathcal{X}}_t(2,u)_{1,\ldots,p}$, by using Corollary 4.2 we have that

$$|\frac{\partial \tilde{\mathcal{X}}_t(u,\omega)}{\partial u} - \frac{\partial \tilde{\mathcal{X}}_t(v,\omega)}{\partial v}|_{abs} \leq |u-v|^{\beta'}\{\underline{W}_t(2,\omega)\}_{1,\ldots,p}. \quad (51)$$

From the above it immediately follows that $\tilde{\mathcal{X}}_t(u,\omega) \in \mathrm{Lip}(1+\beta')$ for all $\omega \in \mathcal{N}^c$. □

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

Theorem 4.2 *Let $\mathcal{X}_{t,N}(2)$ and $\tilde{\mathcal{X}}_t(2,u)$ be defined as in (37) and (35) respectively. Suppose the assumptions of Corollary 4.2 hold. Then for any $t \in \{1,\ldots,N\}$*

(i)

$$\mathcal{X}_{t,N} = \tilde{\mathcal{X}}_t(u) + (\frac{t}{N}-u)\frac{\partial \tilde{\mathcal{X}}_t(u)}{\partial u} + |\frac{t}{N}-u|^{\beta'+1}O_p(1) + \frac{1}{N}O_p(1); \quad (52)$$

(ii)

$$\mathcal{X}_{t,N} = \tilde{\mathcal{X}}_t(u) + (\frac{t}{N}-u)d\mathcal{X}_{t,N}(2) + |\frac{t}{N}-u|^{\beta'+1}O_p(1) + \frac{1}{N}O_p(1). \quad (53)$$

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 4.3) and $\underline{Y}_t(2,\omega)$ converge for all $\omega \in \mathcal{N}_1^c$. Then by using Corollary 4.3 we have that $\tilde{\mathcal{X}}_t(u,\omega) \in Lip(1+\beta')$ if $\omega \in \mathcal{N}_1^c$. By using (17), making a Taylor series expansion of $\tilde{\mathcal{X}}_t(\frac{t}{N},\omega)$ about u and using the mean value theorem we obtain

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

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 (52). We use (46) and the method given above to prove (53). $\square$

Remark 4.1 *Under Assumption 4.1, it is easy to extend all the above results upto the dth derivative of the process. By taking the sth derivative of $\tilde{\mathcal{X}}_t(u)$ we have*

$$\frac{\partial^s \tilde{\mathcal{X}}_t(u)}{\partial u^s} = \sum_{j=0}^{s} \binom{s}{j} \frac{\partial^j A(u)}{\partial u^j} \frac{\partial^{s-j} \tilde{\mathcal{X}}_{t-1}(u)}{\partial u^{s-j}} + \frac{\partial \underline{b}_t^s(u)}{\partial u^s}, \qquad \textit{for } 1 \leq s \leq d,$$

which has the state-space representation

$$\tilde{\mathcal{X}}_t(s+1,u) = A_t(s+1,u)\tilde{\mathcal{X}}_{t-1}(s+1,u) + \underline{b}_t(s+1,u) \tag{54}$$

where,

$$\tilde{\mathcal{X}}_t(s+1,u)^T = (\frac{\partial^s \tilde{\mathcal{X}}_t(u)^T}{\partial u^s}, \ldots, \tilde{\mathcal{X}}_t(u)^T), \quad \underline{b}_t(s+1,u)^T = (\frac{\partial^s \underline{b}_t(u)^T}{\partial u^s}, \ldots, \underline{b}_t(u)^T),$$

$A_t(s+1,u)$ is a $p(s+1) \times p(s+1)$-dimensional block upper triangular random matrix defined by the recursion

$$A_t(s+1,u) = \begin{pmatrix} A_t(u) & B_t(s) \\ 0_{ps\times p} & A_t(s,u) \end{pmatrix},$$

where

$$B_t(s) = \left\{ \binom{s}{1} \frac{\partial A_t(u)}{\partial u}, \quad \ldots \quad , \binom{s}{s} \frac{\partial^s A_t(u)}{\partial u^s} \right\}$$

and $A_t(1,u) = A_t(u)$.

Since $A_t(s+1,u)$ is an upper triangular matrix, it can be shown that $\{A_t(s+1,u)\}_t$ has a negative Lyapunov exponent, thus $\{\mathcal{X}_t(s+1,u)\}_t$ is a well defined stationary process. Consequently we have the Taylor series expansion: for any $t \in \{1,\ldots,N\}$

$$\mathcal{X}_{t,N} = \tilde{\mathcal{X}}_t(u) + \sum_{s=1}^{d} \frac{1}{s!}(\frac{t}{N} - u)^s \frac{\partial^s \tilde{\mathcal{X}}_t(u)}{\partial u^s} + \big(|\frac{t}{N} - u|^{d+\beta'} O_p(1) + \frac{1}{N} O_p(1)\big). \tag{55}$$

We observe that (55) implies that the nonstationary process $\{\mathcal{X}_{t,N}\}$ is a linear combination of stationary processes, that is $\{\tilde{\mathcal{X}}_t(u)\}$ and $\{\frac{\partial^s \tilde{\mathcal{X}}_t(u)}{\partial u^s} : s = 1, \ldots, d\}$. The nonstationarity in the representation (55) comes from the coefficients $\{\frac{1}{s!}(\frac{t}{N} - u)^s : s = 1, \ldots, d\}$, which we recall measure the deviation from stationarity.

We now show that under Assumption 4.1 and the conditions in Proposition 3.2, the moments of the derivative process are uniformly bounded.

Proposition 4.1 *Suppose Assumption 4.1 holds with $d = 1$. Let $\tilde{\mathcal{X}}_t(2, u)$, $\mathcal{X}_{t,N}(2)$, $\underline{V}_{t,N}(2)$ and $\underline{W}_t(2)$ be defined as in Corollary 4.2. Suppose for each $r \in \{1, \ldots, M\}$ and for some $n \in [1, \infty)$, that $\|\tilde{\underline{b}}_t\|_n^E < \infty$ and for some $\delta > 0$ $\sup_{1 \leq r \leq M} \lambda_{spec}\{[\mathcal{A}_t(r)]_n\} < 1 - \delta$.*

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

PROOF. Since for each $r \in \{1, \ldots, M\}$, $\mathcal{A}_t(2, r)$ 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 3.2 can be used to prove the result. □

It is straightforward to generalise the proposition above to $\tilde{\mathcal{X}}_t(s + 1, u)$, where $s \geq 1$. This leads immediately to the corollary below.

Corollary 4.4 *Suppose Assumption 4.1 holds for $d = d_1$. Suppose for each $r \in \{1, \ldots, M\}$ and for some $n \in [1, \infty)$, that $\|\tilde{\underline{b}}_t\|_n^E < \infty$ and for some $\delta > 0$ $\sup_{1 \leq r \leq M} \lambda_{spec}\{[A_t(r)]_n\} < 1 - \delta$. Then for $1 \leq s \leq d_1$ we have*

$$\sup_u \| \frac{\partial^s \tilde{\mathcal{X}}_t(u)}{\partial u^s} \|_n^E < \infty.$$

We now return to the local average example in (23), and obtain a tighter bound for the remainder $\mathcal{B}_{t_0,N}$. By using (52) we have

$$\frac{1}{2M+1} \sum_{k=-M}^{M} \left(\mathcal{X}_{t_0+k,N} - \tilde{\mathcal{X}}_{t_0+k}(u_0) \right) = \frac{1}{2M+1} \sum_{k=-M}^{M} \left(\frac{k}{N} \right) \frac{\partial \tilde{\mathcal{X}}_{t_0+k}(u)}{\partial u} \rfloor_{u=u_0} + R_{t_0,N} \quad (56)$$

where

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

It follows from (56) that the size of the remainder or bias due to nonstationarity depends on the magnitude of the derivative processes $\{\frac{\partial \tilde{\mathcal{X}}_t(u)}{\partial u}\}_t$ at the point u_0. Furthermore, if the the conditions in Proposition 4.1 are satisfied with $n = 2$, then $\|\mathcal{B}_{t_0,N}\|_2^E = O(\frac{M}{N} + \frac{1}{N})$. However we can reduce this bound by assuming that the derivative process satisfies some mixing conditions. We note that conditions are given in Section 5 which guarentee the derivative process is strongly mixing. Let us suppose $\{\frac{\partial \tilde{X}_t(u)}{\partial u}\}_t$ is a short memory process. Then if $\sqrt{M}/N << (M/N)^{1+\beta'}$ we have

$$\frac{1}{2M+1}\sum_{k=-M}^{M}\left(\frac{k}{N}\right)\frac{\partial^2 \tilde{\mathcal{X}}_{t_0+k}(u)}{\partial u^2}\rfloor_{u=u_0} = \frac{\sqrt{M}}{N}O_p(1)$$

and

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

It can be shown that the bound above can be tightened further if we assume that Assumption 4.1 is satisfied with $d = 2$. In this case, if $\{\mathcal{X}_t(3, u)\}$ is a short memory process, then we have

$$\|\mathcal{B}_{t_0,N}\|_2^E = \frac{(M+1)(2M+1)}{6N^2}\left\|\mathbb{E}\left(\frac{\partial \mathcal{X}_t(u)}{\partial u}\rfloor_{u=u_0}\right)\right\|_2 + O\left((\frac{M}{N})^{2+\beta'}\right).$$

From the above we can see that the dominating term in the remainder $\mathcal{B}_{t_0,N}$ is the sum of second derivatives. By using $\|\mathcal{B}_{t_0,N}\|_2^E$ and $\mathrm{var}(\frac{1}{2M+1}\sum_{k=-M}^{M}\tilde{\mathcal{X}}_t(u))$ we are able to evaluate the mean square error of the local average and thus obtain the optimal segment length M.

Example 4.2 *(i) Let us consider the tvGARCH(p, q) process, where the conditions in Example 4.1(ii) are satisfied. We have shown in Example 4.1(i), that Assumption 4.1 (with d = 1) is satisfied. We now consider the stationary derivative process associated with the tvGARCH. Formally differentiating (31) gives*

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

which we have shown earlier admits a state space representation.

Moreover we have the Taylor series expansion

$$X_{t,N} = \tilde{X}_t(u) + (\frac{t}{N} - u)\frac{\partial \tilde{X}_t(u)}{\partial u} + (|\frac{t}{N} - u|^{1+\beta'} + \frac{1}{N}). \tag{57}$$

Finally the conditions of Proposition 4.1 are satisfied, therefore $\sup_{t,N}\|\underline{X}_{t,N}(2)\|_n^E < \infty$ *and* $\sup_u \|\frac{\partial \tilde{X}_t(u)}{\partial u}\|_n^E < \infty$.

(ii) Let us consider the tv-random coefficients GARCH. If the derivatives $\{a_i(u)\}$ and $\{b_j(u)\}$ are β'-Lipschitz continuous, then the process which satisfies

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

has a unique well defined solution. Furthermore the tv-random coefficient GARCH process admits the Taylor series expansion given in (57).

(iii) Suppose $\{X_{t,N}\}$ is a tv-random coefficient AR(1) process. If the parameters $a'(\cdot)$ and $\sigma'(\cdot)$ are β'-Lipschitz continuous, then Assumption 4.1 is satisfied with $d = 1$ and we can show that similar results to those given in (ii) also hold for the tv-random coefficient AR(1) process.

5 Mixing properties of derivative processes

We now consider the mixing properties for the stationary derivative process $\{\tilde{\mathcal{X}}_t(2,u)\}$. To establish geometric mixing of $\{\tilde{\mathcal{X}}_t(2,u)\}$ 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 5.1 (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 (61) and is ϕ-irreducible, then it is geometrically ergodic, hence strongly mixing with a geometric rate.*

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

Lemma 5.2 *Suppose Assumption 3.1 is satisfied and $A_t(2,u)$ is defined as in (36). Then*

$$\mathbb{E}\|A_t(2,u)\|_{spec}^{\varepsilon} < \infty \tag{58}$$

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

PROOF. Under Assumption 3.1 and by using (71) with $n = 1$ we have $\|A_t(2,u)\|_{spec} \leq C\|A_t(u)\|_{spec} \leq C'\|A_t(u)\|_1$, thus $\mathbb{E}\|A_t(2,u)\|_{spec}^{\varepsilon} \leq 2C'\mathbb{E}\|A_t(u)\|_{spec}^{\varepsilon} < \infty$, which gives us (58). From (39) we have that $\{A_t(2,u)\}$ has a negative Lyapunov exponent. □

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

Theorem 5.1 *Suppose Assumption 4.1 (with* $d = 1$*) is satisfied and let the process* $\{\tilde{\mathcal{X}}_t(2, u)\}_t$ *defined in (35) be* ϕ*-irreducible. Then the process* $\{\tilde{\mathcal{X}}_t(2, u)\}$ *is geometrically ergodic and thus strongly mixing with a geometric rate.*

PROOF. We now show that the conditions of Lemma 5.1 are satisfied, the result then follows.

By using Lemma 5.2 there exists an $m > 0$ and $\delta < 0$ such that

$$\frac{1}{m}\mathbb{E}\log\|A_t(2, u)\dots A_{t-m+1}(2, u)\|_{spec} \leq \delta. \tag{59}$$

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

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

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

$$\underline{d}_{m,t}(2, u) = \sum_{k=1}^{m-1} A_t(2, u)\dots A_{t-k+1}(2, u)\underline{b}_{t-k}(2, u) + \underline{b}_t(2, u). \tag{60}$$

By using (58) we have that $\mathbb{E}\|C_{m,t}(2, u)\|_{spec} < \infty$.

From the above it is clear $\{\tilde{\mathcal{X}}_{m,t}(2, u)\}_t$ satisfies the conditions in Lemma 5.1 and is therefore geometrically ergodic. It then follows that $\{\tilde{\mathcal{X}}_t(2, u)\}_t$ is geometrically ergodic. □

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

Corollary 5.1 *Let* $\{\tilde{\mathcal{X}}_t(2, u)\}$ *be defined as in (35). Suppose Assumption 4.1 holds (with* $d = 1$*),* $\{\tilde{\mathcal{X}}_t(2, u)\}$ *is* ϕ*-irreducible and* $\mathbb{E}\|\tilde{\mathcal{X}}_t(2, u)\|^2 < \infty$*. Then we have*

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

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

PROOF. By using Theorem 5.1 we have that $\{\tilde{\mathcal{X}}_t(2, u)\}$ is geometrically ergodic, therefore by using Davidson (1994), Corollary 14.3 we have the result. □

It follows from Corollary 5.1 that $\{\tilde{\mathcal{X}}_t(u)\}_t$ and $\{\frac{\partial \tilde{\mathcal{X}}_t(u)}{\partial u}\}_t$ are short memory processes.

6 Applications

The notion of the stationary approximations and the derivative process can fruitfully be used in many applications. The key is the representation (55) 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 (2003), Theorem 3.2, 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 (55) 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 give proofs of the results stated earlier.

Lemma A.1 *Suppose A is a positive matrix and $|B|_{abs} \leq A$ then*

$$\|B\|_{lub,m} \leq \|A\|_{lub,m}.$$

PROOF. Let $\underline{x}$ be a p-dimensional vector. We observe that $|B\underline{x}|_{abs} \leq |B|_{abs}|\underline{x}|_{abs} \leq A|\underline{x}|_{abs}$. Therefore we have

$$\|B\|_{lub,m} = \sup_{\|\underline{x}\|_m=1} \|B\underline{x}\|_m \leq \sup_{\|\underline{x}\|_m=1} \||B|_{abs}|\underline{x}|_{abs}\|_m \leq \sup_{\|\underline{x}\|_m=1} \|A|\underline{x}|_{abs}\|_m = \|A\|_{lub,m},$$

and hence the result. $\square$

Below we consider solutions of time-varying state space models. We use a theorem of Brandt (1986), which we now state for convenience.

Theorem A.1 (**Brandt (1986)**) *Let $\{A_t, b_t : t \in \mathbb{Z}\}$ be strictly stationary ergodic processes such that both $\mathbb{E}(\log^+ \|A_t\|_2)$, $\mathbb{E}(\log^+ \|b_t\|_2)$ are finite and*

$$\mathcal{X}_t = A_t \mathcal{X}_{t-1} + \underline{b}_t. \tag{61}$$

Suppose the Lyapunov exponent of $\{A_t : t \in \mathbb{Z}\}$ is γ, where $\gamma < 0$. Then for all $t \in \mathbb{Z}$ the series

$$\mathcal{X}_t = \sum_{k=1}^{\infty} A_t \dots A_{t-k+1} \underline{b}_{t-k} + \underline{b}_t$$

converges almost surely and the process is the unique strictly stationary solution of (61).

By using Lemma A.1 we have the following corollary:

Corollary A.1 *Let $\{A_t, b_t : t \in \mathbb{Z}\}$ be a sequence of positive matrices and vectors, $\{B_t, c_t : t \in \mathbb{Z}\}$ be strictly, stationary processes and let $\{\mathcal{X}_t\}_t$ satisfy*

$$\mathcal{X}_t = B_t \mathcal{X}_{t-1} + \underline{c}_t. \tag{62}$$

Suppose $|B_t|_{abs} \leq A_t$, $|c_t|_{abs} \leq b_t$ and $\{A_t, b_t : t \in \mathbb{Z}\}$ satisfy the conditions of Theorem A.1. Then for all $t \in \mathbb{Z}$ the series

$$\mathcal{X}_t = \sum_{k=1}^{\infty} B_t \dots B_{t-k+1} \underline{c}_{t-k} + \underline{c}_t,$$

converges almost surely and the process is the unique, strictly stationary solution of (62).

PROOF. The proof follows from the observation $|B_t \dots B_{t-m}|_{abs} \leq A_t \dots A_{t-m}$. Therefore by using Lemma A.1 we can show that $\{B_t\}$ has a negative Lyapunov exponent. Thus the result follows from Theorem A.1. $\square$

We now show that under suitable conditions certain nonstationary state-space models have a unique solution. This is a nonstationary version of Brandt's Theorem.

Lemma A.2 *Suppose $\{\mathcal{A}_t(i) : i = 1, \dots, M\}$ satisfy Assumption 3.1(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 $n_0 \leq n_1 \leq \dots \leq n_M$, $(s, t]$ denote the integer sequence $(s, t] = \{s+1, s+2, \dots, t\}$ and $J^t_{r,k} = [t-k, t] \cap [n_{r-1}, n_r]$. Then*

$$\underline{Y}_t = \sum_{k \geq 1} \{ \prod_{r=1}^{M} \prod_{i \in J^t_{r,k}} \mathcal{A}_{t-i}(r) \} \underline{d}_{t-k} + \underline{d}_t \tag{63}$$

and

$$\underline{Y}_t = \sum_{k\geq 1} k\{\prod_{r=1}^{M}\prod_{i\in J^t_{r,k}} \mathcal{A}_{t-i}(r)\}\underline{d}_{t-k} \tag{64}$$

converges almost surely.

PROOF. We now prove (63). Since for each $r \in \{1,\ldots,M\}$ the Lyapunov exponent of $\{\mathcal{A}_t(r)\}$ is less than some δ, where $\delta < 0$, we have

$$\limsup_{n\to\infty} \frac{1}{k}\log\|\prod_{r=1}^{M}\prod_{i\in J^t_{r,k}} \mathcal{A}_{t-i}(r)\|_{spec} \leq \sum_{r=1}^{M}\limsup_{n\to\infty}\frac{1}{k}\log\|\prod_{i\in J^t_{r,k}}\mathcal{A}_{t-i}(r)\|_{spec} \leq \frac{\delta}{2}. \tag{65}$$

We now consider the tail of the random vectors $\underline{d}_t$. Using Chebyshev's inequality we have

$$\begin{aligned}\sum_k \mathbb{P}(\log^+\|\underline{d}_k\|_1 > \frac{-k\delta}{4}) &= \sum_k \mathbb{P}\{\|\underline{d}_k\|_1^{\varepsilon} > \exp(\frac{-k\delta\varepsilon}{4})\} \\ &\leq \sum_k \exp(\frac{k\delta\varepsilon}{4})\mathbb{E}(\|\underline{d}_k\|_1^{\varepsilon}) < \infty\end{aligned}$$

Therefore by Kolmogorov's zero one law the event $\log^+\|d_k\| > \frac{-k\delta}{4}$ occurs only finitely often with probability one. Thus almost surely $\lim_{k\to\infty}\sup\frac{1}{k}\log^+\|d_k\|_1 \leq -\delta/4$. Therefore by using (65) we have

$$\begin{aligned}&\limsup_{k\to\infty}\frac{1}{k}\log\|\{\prod_{r=1}^{M}\prod_{i\in J^t_{r,k}}\mathcal{A}_{t-i}(r)\}\underline{d}_{t-k}\| \leq \\ &\limsup_{k\to\infty}\big(\frac{1}{k}\log\|\prod_{r=1}^{M}\prod_{i\in J^t_{r,k}}\mathcal{A}_{t-i}(r)\| + \frac{1}{k}\log\|\underline{d}_{t-k}\|_1\big) \leq \delta/4.\end{aligned}$$

Now by using the above we have that (63) converges almost surely.

To prove (64) we use (65) to obtain

$$\begin{aligned}&\limsup_{k\to\infty}\frac{1}{k}\log\|k\prod_{r=1}^{M}\{\prod_{i\in J^t_{r,k}}\mathcal{A}_i(r)\}\|_{spec} \\ =\ &\limsup_{k\to\infty}\big(\frac{\log k}{k} + \frac{1}{k}\log\|\prod_{r=1}^{M}\prod_{i\in J^t_{r,k}}\mathcal{A}_i(r)\|_{spec}\big) < \frac{\delta}{3}.\end{aligned} \tag{66}$$

Now by using the arguments given to prove (63) we obtain (64). $\square$

PROOF of Theorem 3.1 By the triangular inequality we have

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

We first derive a bound for $|\mathcal{X}_{t,N} - \tilde{\mathcal{X}}_t(\frac{t}{N})|_{abs}$. By expanding $\mathcal{X}_{t,N}$ and $\tilde{\mathcal{X}}_t(\frac{t}{N})$ and under Assumption 3.1 we have

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

Now by continuing the iteration above we obtain

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

where $\underline{V}_{t,N}$ is defined in (20). Under Assumption 3.1 and by using (16) we have that $\mathbb{E}(\log \|\mathcal{A}_t\mathcal{X}_{t-1,N}\|) < \infty$. Therefore by using (64) with $d_t := (\mathcal{A}_t\mathcal{X}_{t-1,N} + \underline{\tilde{b}}_t)$, we have that $\underline{V}_{t,N}$ converges almost surely.

Using a similar method to the above we can show

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

where $\underline{W}_t$ is defined as in (21). By using Lemma A.2 we have that $\{\underline{W}_t\}_t$ converges almost surely.

(19) follows from (17) and (18). $\square$

Lemma A.3 *Suppose A is a $p \times p$-dimensional random matrix, $\mathcal{X}$ a p-dimensional random vector and A and $\mathcal{X}$ are independent. Let $[\cdot]_n$ be defined as in (24) then*

$$\|A\mathcal{X}\|_n^E \leq K\|[A]_n\|_{spec}\|\mathcal{X}\|_n^E, \tag{67}$$

for some finite constant K independent of A.

Let $\{A_i\}_i$ be a sequence of $p \times p$-dimensional, independent, random matrices. The we have

$$\|[A_1 \ldots A_n]_n\|_{spec} \leq K\|[A_1]_m \ldots [A_m]_n\|_{spec}. \tag{68}$$

PROOF. Let $\underline{x}$ be a q-dimensional deterministic vector and A_i denote the ith row of the matrix A. By using the Minkowski inequality we have

$$\begin{aligned}
\|A\mathcal{X}\|_n^E &\leq \{\sum_{i=1}^{q}\mathbb{E}(|A_i^T\mathcal{X})^n|\}^{1/n} \leq \sum_{i=1}^{q}\sum_{j=1}^{q}\{\mathbb{E}|A_{ij}\mathcal{X}_j^n|\}^{1/n} \\
&\leq \sum_{i=1}^{q}\sum_{j=1}^{q}\{\mathbb{E}|A_{ij}^n|\}^{1/n}\{\mathbb{E}|\mathcal{X}_j^n|\}^{1/n} \leq \|[A]_n[\mathcal{X}]_n\|_1 \leq K\|[A]_n\|_{spec}\|[\mathcal{X}]_n\|_2 \\
&\leq K\|[A]_n\|_{spec}\|\mathcal{X}\|_n^E,
\end{aligned}$$

where K is a finite constant. Thus we obtain (67)

We consider $[A_1 \ldots A_n]_n$. By using Minkowski's inequality and the independence of the matrices $\{A_i\}$ we have

$$[A_1 \ldots A_n]_n \leq [A_1]_n \ldots [A_m]_n. \tag{69}$$

Therefore, we obtain the result. □

PROOF of Proposition 3.2 We now show that $\|\mathcal{X}_{t,N}\|_m^E$ is uniformly bounded over t and N. Since $\{\mathcal{A}_t(i)\}_t$ and $\{\underline{\tilde{b}}_t\}_t$ are independent by using (67) and (68) we have

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

By using (14) and the above we have

$$\begin{aligned}
\|\mathcal{X}_{t,N}\|_m^E &\leq \sum_{k=0}^{\infty}\|(\prod_{r=1}^{M}\prod_{i\in J_{r,k}}\mathcal{A}_{t-i}(r))\underline{\tilde{b}}_{t-k}\|_m^E \\
&\leq \sum_{k=0}^{\infty}\prod_{r=1}^{M}\|[\prod_{i\in J_{r,k}}\mathcal{A}_{t-i}(r)]_m\|_{spec}\sup_t\|\underline{\tilde{b}}_t\|_m^E \\
&\leq \sum_{k=0}^{\infty}\prod_{r=1}^{M}\|[\mathcal{A}_0(r)]_m^{\#(J_{r,k})}\|_{spec}\sup_t\|\underline{\tilde{b}}_t\|_m^E
\end{aligned} \tag{70}$$

where $\#(J_{r,k})$=number of elements in the set $J_{r,k}$. Since $\lambda([A_0(i)]_m) \leq 1-\delta$ for $i = 1,\ldots,M$ and by using Moulines et al. (2003), Lemma 15, there exists a K independent of $\mathcal{A}_0(r)$ and n such that $\|[\mathcal{A}_0(r)]_m^n\|_{spec} \leq K(1-\delta/2)^n$. Therefore

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

By substituting the above into (70) and by using $\sum_{r=1}^{M}\#(J_{r,k}) = k$ we get

$$\|\mathcal{X}_{t,N}\|_m^E \;\leq\; \sum_{k=0}^{\infty} K^M(1-\frac{\delta}{2})^k \sup_t \|\underline{\tilde{b}}_t\|_m^E < \infty$$

Therefore $\sup_{t,N}\|\mathcal{X}_{t,N}\|_m^E \leq K\sup_t \|\underline{\tilde{b}}_t\|_m^E$, thus giving us (25).

Using the same method as above we can prove $\sup_{t,N}\|V_{t,N}\|_m^E < \infty$, (26) and (27). □

PROOF of Lemma 4.1 We now prove (39). Under Assumption 4.1 (with $d=1$) 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 a 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 Lemma A.1 and the above we have

$$\|A_t(2,u)\ldots A_{t-n+1}(2,u)\|_{spec} \leq \|B_k(t,n)\|_{spec} \leq C_1(n^2+1)\|\mathcal{A}_t(r)\ldots\mathcal{A}_{t-n+1}(r)\|_{spec}. \quad (71)$$

Thus

$$\frac{1}{n}\mathbb{E}\log\|A_t(2,u)\ldots A_{t-n+1}(2,u)\|_{spec} \leq \frac{\log(n^2+1)}{n} + \frac{1}{n}\mathbb{E}(\log\|\mathcal{A}_t(r)\ldots\mathcal{A}_{t-n+1}(r)\|_{spec}).$$

Since $\log(n^2+1)/n \to 0$, from the above we obtain (39). The proof of (40) is similar, hence we omit this proof. □

References

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

Bibi, A. (2003). On the covariance structure of time-varying bilinear models. *Stochastic Analysis and applications*, *21*, 25-60.

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. (2003). Statistical inference of time varying ARCH processes. *Preprint.*

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. (2003). On recursive estimation for locally stationary time varying autoregressive processes. *To appear in Ann. Statist.*

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. (2003). Quasi-maximum likelihood estimation in conditionally hetroscedastic time series: a stochastic recurrence equation approach. *To appear in Ann. Statist.*

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.