A recursive online algorithm for the estimation of time-varying ARCH parameters*

EXTENDED VERSION †

Rainer Dahlhaus and Suhasini Subba Rao
Institut für Angewandte Mathematik
Universität Heidelberg
Im Neuenheimer Feld 294
69120 Heidelberg
Germany

November 19, 2006

Abstract

In this paper we propose an online recursive algorithm for estimating the parameters of a time-varying ARCH process. The estimation is done by updating the estimator at time point $(t-1)$ with observations about the time point t to yield an estimator of the parameter at time point t. The sampling properties of this estimator are studied in a nonstationary context, in particular asymptotic normality and an expression for the bias due to nonstationarity are established. The minimax risk for the tvARCH process is evaluated and compared with the mean squared error of the online recursive estimator. It is shown, if the time-varying parameters belong to a Hölder class of order less than or equal to one, then, with a suitable choice of step-size, the recursive online estimator attains the local minimax bound. On the other hand, if the order of the Hölder class is greater than one, the minimax bound for the recursive algorithm cannot be attained. However, by running two recursive online algorithms in parallel with different step-sizes and taking a linear combination of the estimators the minimax bound can be attained for Hölder classes of order between one and two.

*Supported by the Deutsche Forschungsgemeinschaft (DA 187/12-2).

†This technical report is the extended version of the paper "A recursive online algorithm for the estimation of time-varying ARCH parameters", to appear in ... In addition to the published paper this version contains in particular a minimax result in Theorem 2.4, Remark 2.2, Theorem A.1 (being part of the proof of Theorem 3.1) and the proof of Lemma A.6.

1 Introduction

The class of ARCH processes can be generalised to include nonstationary processes, by including models with parameters which are time dependent. More precisely $\{X_{t,N}\}$ is called a time-varying ARCH (tvARCH) process of order p if it satisfies the representation

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

where $\{Z_t\}$ are independent, identically (iid) distributed random variables with $\mathbb{E}(Z_0) = 0$ and $\mathbb{E}(Z_0^2) = 1$. This general class of time-varying ARCH (tvARCH) processes was investigated in detail in Dahlhaus and Subba Rao (2006). It was shown that the tvARCH process can locally be approximated by a stationary process, the details we summarise below. Furthermore, a local quasi-likelihood method was proposed to estimate the parameters of the tvARCH(p) model.

A potential application of the tvARCH process is to model long financial time series. The modelling of financial data using nonstationary time series models has recently attracted considerable attention. A justification for using such models can be found, for example, in Mikosch and Stărică (2000, 2004). However given that financial time series are often sampled at high frequency, evaluating the likelihood as each observation comes online can be computationally expensive. Thus an "online" method, which uses the previous estimate of the parameters at time point $(t-1)$ and the observation at time point t to estimate the parameter at time point t, would be ideal and cost effective. There exists a huge literature on recursive algorithms - mainly in the context of linear systems (cf. Ljung and Söderström (1983), Solo (1981, 1989) or in the context of neural networks (cf. White (1996), Chen and White (1998)). For a general overview see also Kushner and Yin (2003). Motivated by the least mean squares algorithm considered in Moulines et al. (2005) for time-varying AR processes, we consider in this paper the following online recursive algorithm for tvARCH models:

$$\underline{\hat{a}}_{t,N} = \underline{\hat{a}}_{t-1,N} + \lambda \{X_{t,N}^2 - \underline{\hat{a}}_{t-1,N}^T \mathcal{X}_{t-1,N}\} \frac{\mathcal{X}_{t-1,N}}{|\mathcal{X}_{t-1,N}|_1^2}, \quad t = (p+1), \ldots, N, \tag{2}$$

where $\mathcal{X}_{t-1,N}^T = (1, X_{t-1,N}^2, \ldots, X_{t-p,N}^2)$ and where $|\mathcal{X}_{t-1,N}|_1 = 1 + \sum_{j=1}^{p} X_{t-j,N}^2$ and we start with the initial conditions $\underline{\hat{a}}_{p,N} = (0, \ldots, 0)$. This algorithm is linear in the estimators, despite the nonlinearity of the tvARCH process. We call the stochastic algorithm defined in (2) the ARCH normalised recursive estimation (ANRE) algorithm. Let $\underline{a}(u)^T = (a_0(u), \ldots, a_p(u))$, then $\underline{\hat{a}}_{t,N}$ is regarded as an estimator of $\underline{a}(\frac{t}{N})$ or of $\underline{a}(u)$ if $|t/N - u| < 1/N$.

In this paper we will prove consistency and asymptotic normality of this recursive estimator. Furthermore, we will discuss the improvements of the estimator obtained by combining two estimates from (2) with different λ. Unlike most other work in the area of recursive estimation the properties of the estimator are proved under the assumptions that the true process is a process with time-varying coefficients, i.e. a nonstationary process. The rescaling of the coefficients in (1) to the unit interval corresponds to the 'infill asymptotics' in nonparametric regression: As $N \to \infty$ the system does not describe the asymptotic behavior of the system in a physical sense but is meant as a meaningful asymptotics to approximate e.g. the distribution of estimates based on a finite sample size. A similar approach has been used in Moulines et al. (2005) for tvAR-models. A more detailed discussion of the relevance of this approach and the relation to non-rescaled processes can be found in Section 3.

In fact the ANRE algorithm resembles the NLMS-algorithm investigated in Moulines et al. (2005). Rewriting (2) we have

$$\underline{\hat{a}}_{t,N} = (I - \lambda \frac{\mathcal{X}_{t-1,N}\mathcal{X}_{t-1,N}^T}{|\mathcal{X}_{t-1,N}|_1^2})\, \underline{\hat{a}}_{t-1,N} + \lambda \frac{X_{t,N}^2 \mathcal{X}_{t-1,N}}{|\mathcal{X}_{t-1,N}|_1^2}. \quad (3)$$

We can see from (3) that the convergence of the ANRE algorithm relies on showing some type of exponential decay of the past. In this paper we will show that for any $p > 0$

$$\mathbb{E}\left\|\prod_{i=1}^{k}(I - \lambda \frac{1}{|\mathcal{X}_{t-i-1,N}|_1^2}\mathcal{X}_{t-i-1,N}\mathcal{X}_{t-i-1,N}^T)\right\|^p \leq K(1-\lambda\delta)^k \text{ for some } \delta > 0, \quad (4)$$

where $\|\cdot\|$ denotes the spectral norm of a matrix and $\prod_{i=0}^{n} A_i = A_0 \cdots A_n$. Roughly speaking this means we have, on average, exponential decay of the past. Similar properties are often established in the control literature and it is often referred to as persistence of excitation of the stochastic matrix (see, for example, Guo (1994) and Aguech et al. (2000)), which in our case is the matrix $(1 - \frac{\lambda}{|\mathcal{X}_{t-1,N}|_1^2}\mathcal{X}_{t-1,N}\mathcal{X}_{t-1,N}^T)$. Persistence of excitation guarentees convergence of the algorithm, which we will use to prove the asymptotic properties of $\underline{\hat{a}}_{t,N}$.

In Section 2 we state all results on the asymptotic behaviour of $\underline{\hat{a}}_{t,N}$ including consistency, asymptotic normality and rate efficiency. Furthermore, we suggest a modified algorithm based on two parallel algorithms. In Section 3 we discuss practical implications. Sections 4 and 5 contain the proofs which in large parts are based on the pertubation technique. We also derive a lower bound for the optimal rate of such estimators and prove that it is achieved. Some technical methods are put in the appendix. We note that some of results in the appendix are of independent interest, as they deal with many of the probablistic properties of both ARCH and tvARCH processes and their vector representations.

2 The ANRE algorithm

We first review some properties of the tvARCH process. Dahlhaus and Subba Rao (2006) and Subba Rao (2004) have shown that the time varying ARCH process can locally be approximated by a stationary ARCH process. Let u be fixed and

$$X_t(u) = Z_t \tilde{\sigma}_t(u) \quad \sigma_t(u)^2 = a_0(u) + \sum_{j=1}^{p} a_j(u) X_{t-j}(u)^2, \quad (5)$$

where $\{Z_t\}$ are independent, identically distributed random variables with $\mathbb{E}(Z_0) = 0$ and $\mathbb{E}(Z_0^2) = 1$. We also set $\mathcal{X}_t(u)^T = (1, X_t(u)^2, \ldots, X_{t-p+1}(u)^2)$. In Lemma 4.1 we show that $X_t(u)^2$ can be regarded as the stationary approximation of $X_{t,N}^2$ around the time points $t/N \approx u$.

Let $\lambda_{min}(A)$ and $\lambda_{max}(A)$ denote the smallest and largest eigenvalues of the matrix A. We use the following assumptions.

Assumption 2.1 *Let $\{X_{t,N}\}$ and $\{X_t(u)\}$ be sequences of stochastic processes which satisfy (1) and (5) respectively. Furthermore*

(i) For some $r \in [1, \infty)$, there exists $\eta > 0$ with

$$\{\mathbb{E}(Z_0^{2r})\}^{1/r} \sup_u \{\sum_{j=1}^{p} a_j(u)\} < 1 - \eta.$$

(ii) There exists $0 < \rho_1 \leq \rho_2 < \infty$ such that for all $u \in (0, 1]$ $\rho_1 \leq a_0(u) \leq \rho_2$.

(iii) There exists $\beta \in (0, 1]$ and a constant K such that for $u, v \in (0, 1]$

$$|a_j(u) - a_j(v)| \leq K|u - v|^\beta \quad \text{for } j = 0, \ldots, p.$$

(iv) Let $Y_0(u) = a_0(u)Z_0^2$ and

$$\text{and } Y_t(u) = \left\{ a_0(u) + \sum_{j=1}^{t} a_j(u) Y_{t-j}(u) \right\} Z_t^2 \quad \text{for} \quad t = 1, \ldots, p.$$

Define $\underline{Y}_p(u) = (1, Y_1(u), \ldots, Y_p(u))^T$ and $\Sigma(u) = \mathbb{E}\big(\underline{Y}_p(u)\underline{Y}_p(u)^T\big)$. Then there exists a constant C such that

$$\inf_u \lambda_{min}\big\{\Sigma(u)\big\} \geq C.$$

Remark 2.1 It is clear that $\Sigma(u)$ is a positive semi-definite matrix, hence its smallest eigenvalue is greater than or equal to zero. It can be shown that if $p/\mathbb{E}(Z_t^4)^{1/2} < 1$ and $\sup_u a_0(u) > 0$, then $\lambda_{min}(\Sigma(u)) > (1 - p/\mathbb{E}(Z_t^4)^{1/2})/a_0(u)^{(2p+1)/(p+1)}$. However this condition is only sufficient and lower bounds can be obtained under much weaker conditions. □

We now investigate some of the asymptotic properties of the ANRE algorithm. Throughout this paper we assume that $\lambda \to 0$ and $\lambda N \to \infty$ as $N \to \infty$. We mention explicitly that λ does not depend on t, i.e. we are considering the fixed-stepsize case. The assumption $\lambda \to 0$ is possible in the triangle array framework of our model and the resulting assertions (for example Theorem 2.2) are meant as an approximation of the corresponding finite sample size distributions and not as the limit in any physical sense.

The following results are based on a representation proved at the end of Section 4.3. The difference $\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)$ is dominated by two terms, that is

$$\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0) \quad = \quad \mathcal{L}_{t_0}(u_0) + \mathcal{R}_{t_0,N}(u_0) + O_p(\delta_N), \tag{6}$$

where

$$\delta_N = (\frac{1}{(N\lambda)^{2\beta}} + \frac{\sqrt{\lambda}}{(N\lambda)^{\beta}} + \lambda + \frac{1}{N^{\beta}}), \tag{7}$$

$$\begin{aligned}
\mathcal{L}_{t_0}(u_0) \quad &= \quad \sum_{k=0}^{t_0-p-1} \lambda\{I - \lambda F(u_0)\}^k \mathcal{M}_{t_0-k}(u_0) \qquad (8)\\
\mathcal{R}_{t_0,N}(u_0) \quad &= \quad \sum_{k=0}^{t_0-p-1} \lambda\{I - \lambda F(u_0)\}^k \Big(\{\mathcal{M}_{t_0-k}(\frac{t_0-k}{N}) - \mathcal{M}_{t_0-k}(u_0)\} + \\
&\qquad F(u_0)\{\underline{a}(\frac{t_0-k}{N}) - \underline{a}(u_0)\}\Big),
\end{aligned}$$

with

$$\mathcal{M}_t(u) = (Z_t^2 - 1)\sigma_t(u)^2 \frac{\mathcal{X}_{t-1}(u)}{|\mathcal{X}_{t-1}(u)|_1^2} \tag{9}$$

and

$$F(u) = \mathbb{E}\left(\frac{\mathcal{X}_0(u)\mathcal{X}_0(u)^T}{|\mathcal{X}_0(u)|_1^2}\right). \tag{10}$$

We note that $\mathcal{L}_{t_0}(u_0)$ and $\mathcal{R}_{t_0,N}(u_0)$ play two different roles. $\mathcal{L}_{t_0}(u_0)$ is the weighted sum of the stationary random variables $\{X_t(u_0)\}_t$, which locally approximate the tvARCH process $\{X_{t,N}\}_t$, whereas $\mathcal{R}_{t_0,N}(u_0)$ is the (stochastic) bias due to nonstationarity; if the tvARCH process were stationary this term would be zero. It is clear from the above that the magnitude of $\mathcal{R}_{t_0,N}(u_0)$ depends on the regularity of the time-varying parameters $\underline{a}(u)$ e.g., the Hölder class that $\underline{a}(u)$ belongs to. By using (6) we are able to obtain a bound for the mean squared error of $\underline{\hat{a}}_{t_0,N}$. Let $|\cdot|$ denote the Euclidean norm of a vector.

Theorem 2.1 *Suppose Assumption 2.1 holds with $r > 4$ and $u_0 > 0$. Then if $|u_0 - t_0/N| < 1/N$ we have*

$$\mathbb{E}\left\{|\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)|^2\right\} = O(\lambda + \frac{1}{(N\lambda)^{2\beta}}), \tag{11}$$

where $\lambda \to 0$, as $N \to \infty$ and $N\lambda >> (\log N)^{1+\varepsilon}$ with some $\varepsilon > 0$.

The proof can be found at the end of Section 4.2. We note that an immediate consequence of Theorem 2.1 is that the estimator $\underline{\hat{a}}_{t_0,N} \xrightarrow{\mathcal{P}} \underline{a}(u_0)$.

We note that the condition $r > 4$, is quite strong. We use the condition $r = 4$, to prove persistence of excitation (see Section A.2), which is an integral part in the proof of the theorem. However, we believe that it is possible to prove the same result under lower moment conditions. In order to obtain the mean squared error we use the condition $r > 4$.

The stochastic term $\mathcal{L}_{t_0}(u_0)$ is the sum of martingale differences, which allows us to obtain the following central limit theorem, whose proof is at the end of Section 4.3.

Theorem 2.2 *Suppose Assumption 2.1 holds with $r > 4$ and $u_0 > 0$. Let $\mathcal{R}_{t_0,N}(u_0)$ be defined as in (8). If $|t_0/N - u_0| < 1/N$,*

(i) $\lambda >> N^{-4\beta/(4\beta+1)}$ and $\lambda >> N^{-2\beta}$ then

$$\lambda^{-1/2}\{\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)\} - \lambda^{-1/2}\mathcal{R}_{t_0,N}(u_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(u_0)), \tag{12}$$

(ii) and $\lambda >> N^{-\frac{2\beta}{2\beta+1}}$ then

$$\lambda^{-1/2}\{\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(u_0)), \tag{13}$$

where $\lambda \to 0$ as $N \to \infty$ and $N\lambda >> (\log N)^{1+\varepsilon}$, for some $\varepsilon > 0$, with

$$\Sigma(u) = \frac{\mu_4}{2} F(u)^{-1} \mathbb{E}\Big(\frac{\sigma_1(u)^4 \mathcal{X}_0(u)\mathcal{X}_0(u)^T}{|\mathcal{X}_0(u)|_1^4}\Big) \tag{14}$$

and $\mu_4 = \mathbb{E}(Z_0^4) - 1$.

Remark 2.2 (Stationary ARCH-models) We note that an analogous algorithm to the ANRE algorithm exists for stationary ARCH processes. Suppose the stationary ARCH(p) process $\{X_t\}_t$ satisfies

$$X_t = Z_t\sigma_t \quad \sigma_t^2 = a_0 + \sum_{j=1}^{p} a_j X_{t-j}^2,$$

where $\{Z_t\}_t$ are iid random variables with $\mathbb{E}(Z_t) = 0$, $\mathbb{E}(Z_t^2) = 1$, $a_0 > 0$, $a_j \geq 0$ if $1 \leq j \leq p$, and $\sum_{j=1}^{p} a_j < 1$. Let $\underline{a}^T = (a_0, \ldots, a_p)$ and

$$\underline{\hat{a}}_t = \underline{\hat{a}}_{t-1} + \frac{K}{t}\{X_t^2 - \underline{\hat{a}}_{t-1}^T \mathcal{X}_{t-1}\} \frac{\mathcal{X}_{t-1}}{|\mathcal{X}_{t-1}|_1^2}, \quad t \geq (p+1),$$

where $\mathcal{X}_{t-1}^T = (1, X_{t-1}^2, \ldots, X_{t-p}^2)$ and K is a finite positive constant. Then given the observations X_k ($k = 1, \ldots, t$), we use $\underline{\hat{a}}_t$ as an estimator of $\underline{a}$. Since the stepsize $\lambda = \frac{K}{t}$ does not fulfill our assumptions we cannot apply the above theorems directly. However, we believe that by using the same methodology it can be shown that

$$\underline{\hat{a}}_t \xrightarrow{\mathcal{P}} \underline{a},$$

and

$$t^{1/2}(\underline{\hat{a}}_t - \underline{a}) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma) \quad \text{with} \quad \Sigma = \frac{\mu_4}{2}\mathbb{E}\big(\frac{\mathcal{X}_0\mathcal{X}_0^T}{|\mathcal{X}_0|_1^2}\big)^{-1}\mathbb{E}\big(\frac{\sigma_1^4\mathcal{X}_0\mathcal{X}_0^T}{|\mathcal{X}_0|_1^4}\big).$$

This algorithm has the advantage that it is simple to evaluate and gives "stable" estimators, in the sense that the difference $|\underline{\hat{a}}_t - \underline{\hat{a}}_{t+1}|_1$ is small. □

Until now we have assumed that $\underline{a}(u) \in Lip(\beta)$, where $\beta \leq 1$. Let $\dot{f}(u)$ denote the derivative of the vector or matrix $f(\cdot)$ with respect to u, Suppose $0 < \beta' \leq 1$ and $\underline{\dot{a}}(u) \in Lip(\beta')$, then we say $\underline{a}(u) \in Lip(1+\beta')$. We now show that an exact expression for the bias can be obtained if $\underline{a}(u) \in Lip(1+\beta')$ and $\beta' > 0$.

We make the following assumptions.

Assumption 2.2 *Let $\{X_{t,N}\}$ be a sequence of stochastic processes which satisfy (1). Suppose*

(i) Assumption 2.1 holds.

(ii) For some $\beta' > 0$

$$|\dot{a}_i(u) - \dot{a}_i(v)| \leq K|u-v|^{\beta'}, \quad \text{for } i = 0,\ldots,p.$$

Under the above assumptions we show in Lemma 5.3 that

$$\mathbb{E}\{\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)\} = -\frac{1}{N\lambda}F(u_0)^{-1}\underline{\dot{a}}(u_0) + O(\frac{1}{(N\lambda)^{1+\beta'}}). \tag{15}$$

We note that $\underline{\hat{a}}_{t_0,N}$ is not a typical parameter estimator of an ARCH process, where usually it is not possible to obtain an exact expression for the bias.

By using the bias above we obtain the following theorem, whose proof is at the end of Section 5. Let $tr(\cdot)$ denote the trace of the matrix.

Theorem 2.3 *Suppose Assumption 2.2 holds with $r > 4$ and $u_0 > 0$. Then if $|t_0/N - u_0| < 1/N$, we have*

$$\begin{aligned}\mathbb{E}|\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)|^2 \quad &= \quad \lambda\, tr\{\Sigma(u_0)\} + \frac{1}{(N\lambda)^2}|F(u_0)^{-1}\underline{\dot{a}}(u_0)|^2 + \\ & \qquad O\left\{\frac{1}{(N\lambda)^{2+\beta'}} + \frac{\lambda^{1/2}}{(N\lambda)^{1+\beta'}} + \frac{1}{(N\lambda)^2}\right\},\end{aligned} \tag{16}$$

and if λ is such that $\lambda^{-1/2}/(N\lambda)^{1+\beta'} \to 0$, then

$$\lambda^{-1/2}(\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)) + \lambda^{-1/2}\frac{1}{N\lambda}F(u_0)^{-1}\underline{\dot{a}}(u_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,\Sigma(u_0)), \tag{17}$$

where $\lambda \to 0$ as $N \to \infty$ and $\lambda N >> (\log N)^{1+\varepsilon}$, for some $\varepsilon > 0$.

We now consider the minimax risk for estimators of tvARCH processes. Let $\lfloor\beta\rfloor$ denote the largest integer strictly less than β and define the norm $\|\cdot\|_\beta$ on $\underline{f} : [0,1] \to \mathbb{R}^{p+1}$ as

$$\|\underline{f}\|_\beta = \sup_{u,v} \frac{|\underline{f}^{\lfloor\beta\rfloor}(u) - \underline{f}^{\lfloor\beta\rfloor}(v)|_1}{|u-v|^{\beta-\lfloor\beta\rfloor}}.$$

Let $\mathcal{C}^\beta(L)$ define the class

$$\mathcal{C}^\beta(L) = \Big\{ \underline{f}(\cdot) = \big(f_0(\cdot), \ldots, f_{p+1}(\cdot)\big) \,\Big|\, \underline{f} : [0,1] \to (\mathbb{R}^+)^{p+1}, \|\underline{f}\|_\beta \leq L, \quad (18)$$
$$\sup_u \sum_{i=1}^{p+1} f_i(u) < 1-\delta,\, 0 < \rho_1 < \inf_u f_0(u) < \rho_2 < \infty \Big\}.$$

It is clear if $\underline{a}(\cdot) \in \mathcal{C}^\beta(L)$ are the parameters of the tvARCH process $\{X_{t,N}\}$, then Assumption 2.1(i,ii,iii) is satisfied. In the following theorem we evaluate a lower bound for the minimax risk.

Theorem 2.4 *Let f_Z denote the density function of Z_k. Suppose $\mathbb{E}(Z_t^4) < \infty$ and*

$$\mathcal{Z} = \mathbb{E}\left\{1 + Z_k \frac{d\log f_Z(x)}{dx}\rfloor_{x=Z_k}\right\}^2 < \infty.$$

Let $\mathcal{C}^\beta(L)$ be defined as in (18) and If $|t_0/N - u_0| < 1/N$, then for any $\|\underline{x}\| \leq 1$, we have

$$\inf_{\underline{a}_{t_0,N} \in \sigma(X_{t,N}; t=1,\ldots,N)} \sup_{\underline{a}(u) \in \mathcal{C}^\beta(L)} \mathbb{E}[\underline{x}^T(\underline{a}_{t_0,N} - \underline{a}(u_0))(\underline{a}_{t_0,N} - \underline{a}(u_0))^T\underline{x}]$$
$$\geq C(\mathcal{Z}, \mathbb{E}(Z_0^4), \rho_1))\, N^{-2\beta/(2\beta+1)}, \quad (19)$$

where $\{X_{t,N}\}$ is a tvARCH process defined by the parameters $a(\cdot)$ and $C(\mathcal{Z}, \mathbb{E}(Z_0^4, \rho_1))$ is a constant which only depends on $\mathcal{Z}, \mathbb{E}(Z_0^4)$ and ρ_1.

Comparing this bound with (11) it is straightforward to show that the ANRE algorithm attains the optimal rate if $\underline{a}(\cdot) \in Lip(\nu)$ with $\nu \leq 1$ (with $\lambda = N^{-\frac{2\nu}{1+2\nu}}$). The story is different when $1 < \nu < 2$. If $\underline{a}(\cdot) \in Lip(1+\beta')$, $\beta' > 0$, the mean squared error of the ANRE estimator in (16) gets minimal for $\lambda \approx N^{-2/3}$ with minimum rate $\mathbb{E}|\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)|^2 = O(N^{-2/3})$. However, in (19) the minimax rate in this case is $N^{-\frac{2(1+\beta')}{1+2(1+\beta')}}$ which is smaller than $N^{-2/3}$. We now present a recursive method which attains the optimal rate even in this case.

Remark 2.3 (Bias correction, rate optimality) The idea here is to achieve a bias correction and the optimal rate by running two ANRE algorithms with different stepsizes λ_1 and λ_2 in parallel. Let $\underline{\hat{a}}_{t,N}(\lambda_1)$ and $\underline{\hat{a}}_{t,N}(\lambda_2)$ be the ANRE algorithms with stepsize λ_1 and λ_2 respectively, and assume that $\lambda_1 > \lambda_2$. By using (15) for $i = 1,2$ we have

$$\mathbb{E}\{\underline{\hat{a}}_{t_0,N}(\lambda_i)\} = \underline{a}(u_0) - \frac{1}{N\lambda_i} F(u_0)^{-1}\underline{\dot{a}}(u_0) + O\big(\frac{1}{(N\lambda_i)^{1+\beta'}}\big). \quad (20)$$

Since $\underline{a}(u_0) - \frac{1}{N\lambda_i} F(u_0)^{-1} \underline{\dot{a}}(u_0) \approx \underline{a}\left(u_0 - \frac{1}{N\lambda_i} F(u_0)^{-1}\right)$ we heuristically estimate $\underline{a}(u_0 - \frac{1}{N\lambda_i} F(u_0)^{-1})$ instead of $\underline{a}(u_0)$ by the algorithm. By using two different λ_i we can find a linear combination of the corresponding estimates such that we "extrapolate" the two values $\underline{a}(u_0) - \frac{1}{N\lambda_i} F(u_0)^{-1} \underline{\dot{a}}(u_0)$ $(i = 1, 2)$ to $\underline{a}(u_0)$. Formally let $0 < w < 1$, $\lambda_2 = w\lambda_1$ and

$$\underline{\check{a}}_{t_0,N}(w) = \frac{1}{1-w} \underline{\hat{a}}_{t_0,N}(\lambda_1) - \frac{w}{1-w} \underline{\hat{a}}_{t_0,N}(\lambda_2).$$

If $|t_0/N - u_0| < 1/N$, then by using (20) we have

$$\mathbb{E}\{\underline{\check{a}}_{t_0,N}(w)\} = \underline{a}(u_0) + O\big(\frac{1}{(N\lambda)^{1+\beta'}}\big).$$

By using Propostion 4.3 we have

$$\mathbb{E}|\underline{\check{a}}_{t_0,N} - \underline{a}(u_0))|^2 = O(\lambda + \frac{1}{(N\lambda)^{2(1+\beta')}}),$$

and choosing $\lambda = \text{const} \times \mathrm{N}^{-(2+2\beta')/(3+2\beta')}$ gives the optimal rate. It remains the problem of choosing λ (and w). It is obvious that λ should be chosen adaptively to the degree of nonstationarity. That is λ should be large if the characteristics of the process are changing more rapidly. However, a more specific suggestion would require more investigations - both theoretically and by simulations.

The above method can not be extended to higher order derivatives since then the other remainders of order $(N\lambda)^{-2}$ (see (16) and the proof of Theorem 2.3).

Finally we mention that choosíng $\lambda_2 < w\lambda_1$ will lead to an estimator of $\underline{a}(u_0 + \Delta)$ with some $\Delta > 0$ (with rate as above). This could be the basis for the prediction of volatility of time varying ARCH-processes. □

3 Practical implications

Suppose that we observe data from a (non-rescaled) time-varying ARCH-process in discrete time

$$X_t = Z_t \sigma_t, \quad \sigma_t^2 = a_0(t) + \sum_{j=1}^{p} a_j(t) X_{t-j}^2, \quad t \in \mathbb{Z}. \tag{21}$$

In order to estimate $\underline{a}(t)$ we use the estimator $\underline{\hat{a}}_t$ as given in (2) (with all subscripts N dropped). An approximation for the distribution of the estimator is given by Theorem 2.2. Theorem 2.2(ii) can be used directly since it is completely formulated without N. The matrices $F(u_0)$ and $\Sigma(u_0)$ depend on the unknown stationary approximation $\mathcal{X}_t(u_0)$ of the process at $u_0 = t_0/N$, i.e. at time t_0 in non-rescaled time. Since this approximation is unknown we may use instead the process itself in a small neighborhood of t_0, i.e. we may estimate for example $F(u_0)$ by $\frac{1}{m}\sum_{j=0}^{m-1} \frac{\mathcal{X}_{t_0-j}\mathcal{X}_{t_0-j}^T}{|\mathcal{X}_{t_0-j}|_1^2}$ with m small and

$\mathcal{X}_{t-1}^T = (1, X_{t-1}^2, \ldots, X_{t-p}^2)$. An estimator which fits better to the recursive algorithm is $[1-(1-\lambda)^{t_0-p+1}]^{-1} \sum_{j=0}^{t_0-p} \lambda(1-\lambda)^j \frac{\mathcal{X}_{t_0-j}\mathcal{X}_{t_0-j}^T}{|\mathcal{X}_{t_0-j}|_1^2}$. In the same way we can estimate $\Sigma(u_0)$ which altogether leads e.g. to an approximate confidence interval for $\underline{\hat{a}}_t$. In a similar way Theorem 2.2(i) can be used.

The situation is more difficult with Theorem 2.1 and Theorem 2.3, since here the results depend (at first sight) on N. Suppose that we have parameter functions $\tilde{a}_j(\cdot)$ and some $N > t_0$ with $\tilde{a}_j(\frac{t_0}{N}) = a_j(t_0)$ (i.e. the original function has been rescaled to the unit interval). Consider Theorem 2.3 with the functions $\tilde{a}_j(\cdot)$. The bias in (15) and (16) contains the term

$$\frac{1}{N}\,\dot{\tilde{a}}_j(u_0) \approx \frac{1}{N}\frac{\tilde{a}_j(\frac{t_0}{N}) - \tilde{a}_j(\frac{t_0-1}{N})}{\frac{1}{N}} = a_j(t_0) - a_j(t_0-1)$$

which again is independent of N. To avoid confusion we mention that $\frac{1}{N}\,\dot{\tilde{a}}_j(u_0)$ of course depends on N once the function $\tilde{a}_j(\cdot)$ has been fixed (as in the asymptotic approach of this paper) but it does not depend on N when it is used to approximate the function $a_j(t)$ since then the function $\tilde{a}_j(\cdot)$ is a different one for each N. In the spirit of the remarks above we would e.g. use as an estimator of $\frac{1}{N}\,\dot{\tilde{a}}_j(u_0)$ in (15) and (16) the expression $[1-(1-\lambda)^{t_0-p+1}]^{-1} \sum_{j=0}^{t_0-p} \lambda(1-\lambda)^j \big[a_j(t_0) - a_j(t_0-j)\big]$.

These considerations also demonstrate the need for the asymptotic approach of this paper: While it is not possible to set down a meaningful asymptotic theory for the model (21) and to derive e.g. a central limit theorem for the estimator $\underline{\hat{a}}_t$, the approach of the present paper for the rescaled model (1) leads to such results. This is achieved by the 'infill asymptotics' where more and more data become available of each local structure (e.g. about time u_0) as $N \to \infty$. The results can then be used also for approximations in the model (21) - e.g. for confidence intervals.

4 Proofs

4.1 Some preliminary results

In the next lemma we give a bound for the approximation error between $X_{t,N}^2$ and $X_t(u)^2$. The proofs of these results and further details can be found in Dahlhaus and Subba Rao (2006) (see also Subba Rao (2004)).

Lemma 4.1 *Suppose Assumption 2.1 holds with some $r \geq 1$. Let $\{X_{t,N}\}$ and $\{X_t(u)\}$ be defined as in (1) and (5). Then we have*

(i) $\{X_t(u)^2\}_t$ is a stationary ergodic sequence. Furthermore there exists a stochastic process $\{V_{t,N}\}_t$ and a stationary ergodic process $\{W_t\}_t$ with $\sup_{t,N} \mathbb{E}(|V_{t,N}|^r) < \infty$ and $\mathbb{E}(|W_t|^r) < \infty$, such that

$$|X_{t,N}^2 - X_t(u)^2| \leq \frac{1}{N^\beta} V_{t,N} + |\frac{t}{N} - u|^\beta W_t \tag{22}$$

$$|X_t(u)^2 - X_t(v)^2| \leq |u-v|^{\beta} W_t. \qquad (23)$$

(ii) $\sup_{t,N} \mathbb{E}(X_{t,N}^{2r}) < \infty$ *and* $\sup_u \mathbb{E}(X_t(u)^{2r}) < \infty$.

We now define the derivative process by $\{\dot{X}_{tt}^2(u)\}_t$.

Lemma 4.2 *Suppose Assumption 2.2 holds with some $r \geq 1$. Then the derivative $\{\dot{X}_t^2(u)\}_t$ is a well defined process and satisfies the representation*

$$\dot{X}_t^2(u) = \left\{ \dot{a}_0(u) + \sum_{j=1}^{p} a_j(u) \dot{X}_{t-j}^2(u) + \dot{a}_j(u) X_{t-j}(u)^2 \right\} Z_t^2.$$

$\{\dot{X}_t(u)^2\}_t$ *is a stationary, ergodic process with*

$$|\dot{X}_t^2(u) - \dot{X}_t^2(v)| \leq |u-v|^{\beta'} W_t, \qquad (24)$$

where W_t is the same W_t used in Lemma 4.1. Almost surely all paths of $\{X_t(u)^2\}_u$ belong to $Lip(1)$ and we have the Taylor series expansion

$$X_{t,N}^2 = X_t(u)^2 + (\frac{t}{N} - u)\dot{X}_t^2(u) + (|\frac{t}{N} - u|^{1+\beta'} + \frac{1}{N}) R_{t,N},$$

where $|R_{t,N}|_1 \leq (V_{t,N} + W_t)$. Furthermore, $\sup_N X_{t,N}^2 \leq W_t$, $\sup_u X_t(u)^2 \leq W_t$ and $\sup_u |\dot{X}_t^2(u)| < W_t$. And the norms are bounded: $\sup_u \mathbb{E}(|\dot{X}_t^2(u)|^r) < \infty$ and $\sup_{t,N} \mathbb{E}(|R_{t,N}|^r) < \infty$.

Define $\mathcal{F}_t = \sigma(Z_t^2, Z_{t-1}^2, \ldots)$, (is clear that $\mathcal{F}_t = \sigma(X_{t,N}^2, X_{t-1,N}^2, \ldots) = \sigma(X_t(u)^2, X_{t-1}(u)^2, \ldots)$, since a Volterra expansion gives $X_{t,N}$ in terms of $\{Z_t^2\}_t$ and the ARCH equations give Z_t in terms of $\{X_t\}_t$). We now consider the mixing properties of functions of the processes $\{\mathcal{X}_{t,N}\}_t$ and $\{\mathcal{X}_t(u)\}_t$. The proof of the proposition below can be found in Section A.1.

Proposition 4.1 *Suppose Assumption 2.1 holds with $r = 1$. Let $\{X_{t,N}\}$ and $\{X_t(u)\}$ be defined as in (1) and (5) respectively. Then there exists a $(1-\eta) \leq \rho < 1$ such that for any $\phi \in Lip(1)$*

$$\big|\mathbb{E}\big[\phi(\mathcal{X}_{t,N})|\mathcal{F}_{t-k}\big] - \mathbb{E}\big[\phi(\mathcal{X}_{t,N})\big]\big|_1 \leq K\rho^k(1 + |\mathcal{X}_{t-k,N}|_1), \qquad (25)$$

$$\big|\mathbb{E}\big[\phi(\mathcal{X}_t(u))|\mathcal{F}_{t-k}\big] - \mathbb{E}\big[\phi(\mathcal{X}_t(u))\big]\big|_1 \leq K\rho^k(1 + |\mathcal{X}_{t-k}(u)|_1), \qquad (26)$$

$$\big|\mathbb{E}\big[\phi(\mathcal{X}_{t,N})|\mathcal{F}_{t-k}\big]\big|_1 \leq K(1 + \rho^k|\mathcal{X}_{t-k,N}|_1), \qquad (27)$$

if $j, k > 0$ then

$$\big|\mathbb{E}\big[\phi(\mathcal{X}_{t,N})|\mathcal{F}_{t-k}\big] - \mathbb{E}\big[\phi(\mathcal{X}_{t,N})|\mathcal{F}_{t-k-j}\big]\big|_1 \leq K\rho^k\left(|\mathcal{X}_{t-k,N}|_1 + |\mathcal{X}_{t-k-j,N})|_1\right), \qquad (28)$$

and if Assumption 2.1 holds with some $r > 1$ then for $1 \leq q \leq r$ we have

$$\mathbb{E}\big(|\mathcal{X}_{t,N}|_1^q \big| \mathcal{F}_{t-k}) \leq K|\mathcal{X}_{t-k,N}|_1^q \qquad (29)$$

where the constant K is independent of t, k, j and N.

The corollary below follows by using (25) and (26).

Corollary 4.1 *Suppose Assumption 2.1 holds. Let* $\{X_{t,N}\}$ *and* $\{X_t(u)\}$ *be defined as in (1) and (5) respectively and* $\phi \in Lip(1)$*. Then* $\{\phi(\mathcal{X}_{t,N})\}$ *and* $\{\phi(\mathcal{X}_t(u))\}$ *are* L_q*-mixingales of size* $-\infty$*.*

4.2 The pertubation technique

In this section we use the pertubation technique, introduced in Aguech et al. (2000), to study the ANRE estimator, which we use to show consistency. To analyse the algorithm we compare it with a similar algorithm, which is driven with the true parameters and where $X_{t,N}$ has been replaced by the stationary process $X_t(u)$. Let $\delta_{t,N}(u) = \hat{\underline{a}}_{t,N} - \underline{a}(u)$,

$$\mathcal{M}_{t,N} = (Z_t^2 - 1)\sigma_{t,N}^2 \frac{\mathcal{X}_{t-1,N}}{|\mathcal{X}_{t-1,N}|_1^2}, \qquad \mathcal{M}_t(u) = (Z_t^2 - 1)\sigma_t(u)^2 \frac{\mathcal{X}_{t-1}(u)}{|\mathcal{X}_{t-1}(u)|_1^2}, \tag{30}$$

$$F_{t,N} = \frac{\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T}{|\mathcal{X}_{t,N}|_1^2}, \quad F_t(u) = \frac{\mathcal{X}_t(u)\mathcal{X}_t(u)^T}{|\mathcal{X}_t(u)|_1^2}, \tag{31}$$

and $F(u)$ be defined as in (10). The 'true algorithm' is

$$\underline{a}(u) = \underline{a}(u) + \lambda\{X_t(u)^2 - \underline{a}(u)^T\mathcal{X}_{t-1}(u)\}\frac{\mathcal{X}_{t-1}(u)}{|\mathcal{X}_{t-1}(u)|_1^2} - \lambda\mathcal{M}_t(u). \tag{32}$$

An advantage about the specific form of the random matrices $F_{t,N}$ is that $|F_{t,N}|_1 \leq (p+1)^2$. This upper bound will make some of the analysis easier to handle.

By subtracting (32) from (2) we get

$$\delta_{t,N}(u) = \big(I - \lambda F_{t-1,N}\big)\delta_{t-1,N}(u) + \lambda\mathcal{B}_{t,N}(u) + \lambda\mathcal{M}_t(u), \tag{33}$$

where

$$\mathcal{B}_{t,N}(u) = \{\mathcal{M}_{t,N} - \mathcal{M}_t(u)\} + F_{t-1,N}\{\underline{a}(\frac{t}{N}) - \underline{a}(u)\}. \tag{34}$$

We note that (33) can also be considered as a recursive algorithm, thus we call (33) the error algorithm. There are two terms driving the error algorithm: the bias $\mathcal{B}_{t,N}(u)$ and the stochastic term, $\mathcal{M}_t(u)$. Because the error algorithm is linear with respect to the estimators we can separate $\delta_{t,N}(u)$ in terms of the bias and the stochastic terms:

$$\delta_{t,N}(u) = \delta_{t,N}^B(u) + \delta_{t,N}^M(u) + \delta_{t,N}^R(u),$$

where

$$\delta_{t,N}^B(u) = (I - \lambda F_{t-1,N})\delta_{t-1,N}^B(u) + \lambda\mathcal{B}_{t,N}(u), \quad \delta_{p,N}^B(u) = 0 \tag{35}$$

$$\delta_{t,N}^M(u) = (I - \lambda F_{t-1,N})\delta_{t-1,N}^M(u) + \lambda\mathcal{M}_t(u), \quad \delta_{p,N}^M(u) = 0 \tag{36}$$

$$\text{and } \delta_{t,N}^R(u) = (I - \lambda F_{t-1,N})\delta_{t-1,N}^R(u), \quad \delta_{p,N}^R(u) = -\underline{a}(u). \tag{37}$$

We have for $y \in \{B, M, R\}$

$$\delta^{y}_{t,N}(u) \;=\; \sum_{k=0}^{t-p-1} \Big\{ \prod_{i=0}^{k-1} \big(I - \lambda F_{t-i-1,N}\big) \Big\} \mathcal{D}^{y}_{t-k,N}(u), \tag{38}$$

where $\mathcal{D}^{B}_{t,N}(u) = \lambda \mathcal{B}_{t,N}(u)$, $\mathcal{D}^{M}_{t,N}(u) = \lambda \mathcal{M}_t(u)$, $\mathcal{D}^{R}_{t,N}(u) = 0$ if $t > p$ and $\mathcal{D}^{R}_{p,N}(u) = -\underline{a}(u)$.

It is clear for $\delta^{y}_{t,N}(u)$ to converge, that the random product $\prod_{i=0}^{t-1}(I - \lambda F_{t-i-1,N})$ must decay. Technically this is one of the main results. It is stated in the following theorem.

Theorem 4.1 *Suppose the Assumption 2.1 holds with $r = 4$ and N is sufficiently large. Then for all $q \geq 1$ and $(p+1) \leq t \leq N$ there exists $M > 0, \delta > 0$ such that*

$$\mathbb{E} \left\| \prod_{i=0}^{k-1} \big(I - \lambda \frac{\mathcal{X}_{t-i-1,N} \mathcal{X}^{T}_{t-i-1,N}}{\|\mathcal{X}_{t-i-1,N}\|_1^2}\big) \right\|^{q} \leq M \exp\{-\delta \lambda k\} \tag{39}$$

Under Assumption 2.1 and by using the proposition above, there exists a $\delta > 0$ such that

$$\big(\mathbb{E}|\delta^{R}_{t,N}|^{q}\big)^{1/q} \leq M \exp\{-\delta \lambda t\}, \tag{40}$$

for $1 \leq q \leq r$ and $t = p+1, \ldots, N$. Therefore this term decays exponentially and, as we shall see below, is of lower order than $\delta^{B}_{t,N}(u)$ and $\delta^{M}_{t,N}(u)$.

We now study the stochastic bias at t_0 with $|t_0/N - u_0| < 1/N$. It is straightforward to see that $\delta^{B}_{t,N}(u) = \delta^{1,B}_{t,N}(u) + \delta^{2,B}_{t,N}(u) + \delta^{3,B}_{t,N}(u)$, where

$$\begin{aligned}
\delta^{1,B}_{t,N}(u) &= (I - \lambda F_{t-1,N}) \delta^{1,B}_{t-1,N}(u) + \lambda \{\mathcal{M}_{t,N} - \mathcal{M}_t(u)\}, \\
\delta^{2,B}_{t,N}(u) &= (I - \lambda F_{t-1,N}) \delta^{2,B}_{t-1,N}(u) + \lambda F(u) \{\underline{a}(\frac{t}{N}) - \underline{a}(u)\}, \\
\delta^{3,B}_{t,N}(u) &= (I - \lambda F_{t-1,N}) \delta^{3,B}_{t-1,N}(u) + \lambda (F_{t-1,N} - F(u)) \{\underline{a}(\frac{t}{N}) - \underline{a}(u)\},
\end{aligned} \tag{41}$$

$\delta^{1,B}_{p,N} = 0$, $\delta^{2,B}_{p,N} = 0$ and $\delta^{3,B}_{p,N} = 0$.

In order to obtain asymptotic expressions for the expectation of each component $\delta^{1,B}_{t_0,N}$, $\delta^{2,B}_{t_0,N}$, $\delta^{2,B}_{t_0,N}$ and $\delta^{M}_{t_0,N}$ we use the pertubation technique proposed by Aguech et al. (2000). Roughly speaking this means substituting the product of random matrices by a deterministic product and showing that the difference is of a lower order. We can decompose for $x = M, (1,B), (2,B)$ the stochastic and bias terms as follows

$$\delta^{x}_{t_0,N}(u_0) = J^{x,1}_{t_0,N}(u_0) + J^{x,2}_{t_0,N}(u_0) + H^{x}_{t_0,N}(u_0), \tag{42}$$

where

$$\begin{aligned}
J^{x,1}_{t,N}(u_0) &= (I - \lambda F(u_0)) J^{x,1}_{t-1,N}(u_0) + G^{x}_{t,N}, \qquad (43)\\
J^{x,2}_{t,N}(u_0) &= (I - \lambda F(u_0)) J^{x,2}_{t-1,N}(u_0) - \lambda \{F_{t-1,N} - F(u_0)\} J^{x,1}_{t-1,N}(u_0), \\
\text{and } H^{x}_{t,N}(u_0) &= (I - \lambda F_{t-1,N}) H^{x}_{t-1,N}(u_0) - \lambda \{F_{t-1,N} - F(u_0)\} J^{x,2}_{t-1,N}(u_0),
\end{aligned}$$

with

$$G_{t,N}^{M} = \lambda \mathcal{M}_t(u_0), \qquad G_{t,N}^{1,B} = \lambda[\mathcal{M}_{t,N} - \mathcal{M}_t(u_0)],$$
$$G_{t,N}^{2,B} = F(u_0)[\underline{a}(\tfrac{t}{N}) - \underline{a}(u)], \quad G_{t,N}^{3,B} = (F_{t-1,N} - F(u_0))[\underline{a}(\frac{t}{N}) - \underline{a}(u)] \quad \text{for } t < t_0. \tag{44}$$

Furthermore, $J_{p,N}^{x,1}(u_0) = J_{p,N}^{x,2}(u_0) = H_{p,N}^{x}(u_0) = 0$. (42) can easily be derived by taking the sum on both sides of the three equations in (43). In the proposition below we will show that $J_{t_0,N}^{x,1}(u_0)$ is the principle term in the expansion of $\delta_{t_0,N}^{x}(u_0)$, i.e. with $\delta_{t_0,N}^{x}(u_0) \approx J_{t_0,N}^{x,1}(u_0)$. Substituting (44) into (43) we have

$$\begin{aligned}
J_{t_0,N}^{(1,B),1}(u_0) &= \sum_{k=0}^{t_0-p-1} \lambda\{I - \lambda F(u_0)\}^k \{\mathcal{M}_{t_0-k,N} - \mathcal{M}_{t_0-k}(u_0)\} \\
J_{t_0,N}^{(2,B),1}(u_0) &= \sum_{k=1}^{t_0-p-1} \{I - \lambda F(u_0)\}^k F(u_0)\{\underline{a}(\frac{t_0-k}{N}) - \underline{a}(u_0)\} \\
J_{t_0,N}^{(3,B),1}(u_0) &= \sum_{k=1}^{t_0-p-1} \{I - \lambda F(u_0)\}^k (F_{t-k-1,N} - F(u_0))\{\underline{a}(\frac{t_0-k}{N}) - \underline{a}(u_0)\}
\end{aligned} \tag{45}$$

In the proof below we require the following definition

$$D_{t,N} = \sum_{i=0}^{p-1}[V_{t-i,N} + W_{t-i}], \text{ for } t \geq p \tag{46}$$

where $V_{t,N}$ and W_t are defined as in Lemma 4.1.

Proposition 4.2 *Suppose Assumption 2.1 holds with $r > 4$ and let $\delta_{t_0,N}^{B}(u_0)$ and $J_{t_0,N}^{(1,B),1}(u_0)$, $J_{t_0,N}^{(2,B),1}(u_0)$ and $J_{t_0,N}^{(3,B),1}(u_0)$, be defined as in (35) and (45). Then for $|t_0/N - u_0| < 1/N$ and $\lambda N \geq (\log N)^{1+\varepsilon}$, for some $\varepsilon > 0$, we have*

(i)

$$\Big(\mathbb{E}|J_{t_0,N}^{(1,B),1}(u_0) + J_{t_0,N}^{(2,B),1}(u_0)|^r\Big)^{1/r} = O(\frac{1}{(N\lambda)^{\beta}}) \tag{47}$$

(ii) $(\mathbb{E}|J_{t_0,N}^{(3,B),1}(u_0)|^r)^{1/r} = O(\frac{1}{(N\lambda)^{2\beta}} + \lambda^{1/2}(N\lambda)^{-\beta})$

(iii) and

$$\Big(\mathbb{E}|\delta_{t_0,N}^{B}(u_0) - J_{t_0,N}^{(1,B),1}(u_0) - J_{t_0,N}^{(2,B),1}(u_0)|^2\Big)^{1/2} = O(\frac{1}{(N\lambda)^{2\beta}} + \frac{\sqrt{\lambda}}{(N\lambda)^{\beta}} + \lambda). \tag{48}$$

PROOF. We first prove (i). Let us consider $J_{t_0,N}^{(1,B),1}(u_0)$. By using (117) we have

$$|\mathcal{M}_{t,N} - \mathcal{M}_t(u_0)|_1 \leq |Z_t^2 + 1|\{\frac{1}{N^{\beta}} + (\frac{p}{N})^{\beta} + |\frac{t}{N} - u_0|^{\beta}\}D_{t,N}.$$

Therefore by substituting the above bound into $J^{(1,B),1}_{t_0,N}(u_0)$ and by using (110) we have

$$\begin{aligned}\left(\mathbb{E}|J^{(1,B),1}_{t_0,N}(u_0)|^r\right)^{1/r} &\leq \sum_{k=0}^{t_0-p-1}\lambda\|(I-\lambda F(u_0))^k\|\ (\mathbb{E}|\mathcal{M}_{t_0-k,N}-\mathcal{M}_{t_0-k}(u_0)|^r)^{1/r} \\ &\leq \frac{K}{N^\beta}\sum_{k=0}^{t_0-p-1}\lambda(1-\delta\lambda)^k\{1+p^\beta+k^\beta\}(\mathbb{E}(Z_0^2+1)^r)^{1/r}(\mathbb{E}|D_{t_0-k-1,N}|^r)^{1/r}.\end{aligned}$$

Now by using Lemma 4.1(i) we have that $\sup_{t,N}\|D_{t,N}\|_r^E<\infty$. Furthermore, from Moulines et al. (2005), Lemma C.3 we have

$$\sum_{k=1}^{N}(1-\lambda)^k k^\beta \leq \lambda^{-1-\beta}. \tag{49}$$

By using the above we obtain

$$(\mathbb{E}|J^{(1,B),1}_{t_0,N}(u_0)|^r)^{1/r} \leq K\sup_{t,N}(\mathbb{E}|D_{t,N}|^r)^{1/r}(\mathbb{E}|Z_0^2+1|^r)^{1/r}\frac{1}{(N\lambda)^\beta}=O(\frac{1}{(N\lambda)^\beta}). \tag{50}$$

We now bound $(\mathbb{E}|J^{(2,B),1}_{t_0,N}(u_0)|^r)^{1/r}$. Since $\underline{a}(u)\in Lip(\beta)$, by using (110) and (49) we have

$$(\mathbb{E}|J^{(2,B),1}_{t_0,N}(u_0)|^r)^{1/r} \leq C\sum_{k=1}^{t_0-p-1}\lambda(1-\delta\lambda)^k\big(\frac{k}{N}\big)^\beta=O(\frac{1}{(\lambda N)^\beta}). \tag{51}$$

Thus (50) and (51) give the bound (47), which completes the proof of (i).

To prove (ii), we now bound $(\mathbb{E}|J^{(3,B),1}_{t_0,N}(u_0)|^r)^{1/2}$. We first observe that $J^{(3,B),1}_{t_0,N}(u_0)$ can be written as

$$\begin{aligned}J^{(3,B),1}_{t_0,N}(u_0) &= I_{t_0,N}+II_{t_0,N} \\ \text{where}\quad I_{t_0,N} &= \sum_{k=1}^{t_0-p-1}\{I-\lambda F(u_0)\}^k(F_{t-k-1,N}-F_{t-k-1}(u_0))\{\underline{a}(\frac{t_0-k}{N})-\underline{a}(u_0)\} \\ II_{t_0,N} &= \sum_{k=1}^{t_0-p-1}\{I-\lambda F(u_0)\}^k(F_{t-k-1}(u_0)-F(u_0))\{\underline{a}(\frac{t_0-k}{N})-\underline{a}(u_0)\}.\end{aligned}$$

Using the same proof to prove $J^{(1,B),1}_{t_0,N}(u_0)$, it is straightforward to show that $(\mathbb{E}|I_{t_0,N}|^r)^{1/r}=O((N\lambda)^{-2\beta})$. In order to bound $II_{t_0,N}$, let $\bar{F}_k(u_0)=F_k(u_0)-F(u_0)$ and $\phi(\underline{x})=\underline{x}\underline{x}^T/|\underline{x}|_1^2$ where $\underline{x}=(1,x_1,\ldots,x_p)$, then $\phi(\underline{x})\in Lip(1)$ and $F_{t,N}=\phi(\mathcal{X}_{t,N})$. Since $\phi\in Lip(1)$, by using Corollary 4.1 we have that $F_t(u_0)-F(u_0)\}$ can be written as the sum of martingale differences:

$$F_t(u_0)-F(u_0)=\sum_{\ell=0}^{\infty}m_t(\ell), \tag{52}$$

where $m_t(\ell)$ is a $(p+1)\times(p+1)$ matrix, defined by $m_t(\ell) = \{\mathbb{E}(F_{t,N}|\mathcal{F}_{t-\ell}) - \mathbb{E}(F_{t,N}|\mathcal{F}_{t-\ell-1})\}$. By using (28) we have $(\mathbb{E}|m_t(\ell)|^{r/2})^{2/r} \leq K\rho^{\ell}$. Substituting the above into $B^{2,B}_{t_0,N}$ we have

$$\begin{aligned} II_{t_0,N} &= \sum_{\ell=0}^{\infty}\sum_{k=0}^{t_0-p-1} \lambda\{I - \lambda F(u_0)\}^k m_{t_0-k-1}(\ell)\{\underline{a}(\frac{t_0-k}{N}) - \underline{a}(u_0)\} \\ \{II_{t_0,N}\}_i &= \sum_{j=1}^{p+1}\sum_{\ell=0}^{\infty}\sum_{k=0}^{t_0-p-1} \lambda\{I - \lambda F(u_0)\}^k\{m_{t_0-k-1}(\ell)\}_{ij}\{a_j(\frac{t_0-k}{N}) - a_j(u_0)\}. \end{aligned}$$

We note that $\|\{I - \lambda F(u_0)\}^k\| \leq K(1-\lambda\delta)^k$ (see (110)). Furthermore, if $t_1 < t_2$ then $\mathbb{E}\{m_{t_1}(\ell)m_{t_2}(\ell)\} = \mathbb{E}\{m_{t_1}(\ell)\mathbb{E}(m_{t_2}(\ell)|\mathcal{F}_{t_1-\ell})\} = 0$, therefore $\{m_t(\ell)\}_t$ is a sequence of martingale differences. Therefore, since $J^{(2,B),1}_{t_0-k-1,N}$ is deterministic we can use Burkhölder's inequality (cf. (Hall & Heyde, 1980), Theorem 12.2) to obtain

$$\begin{aligned} (\mathbb{E}|II_{t_0,N}|^r)^{1/r} &\leq K\lambda\sum_{\ell=0}^{\infty}\sum_{i,j=1}^{p+1}\left(\mathbb{E}|\sum_{k=0}^{t_0-p-1}\{I - \lambda F(u_0)\}^k\{m_{t_0-k-1}(\ell)\}_{ij}\{\underline{a}(\frac{t_0-k}{N})_j - \underline{a}(u_0)_j\}|^r\right)^{1/r} \\ &\leq K\lambda\sum_{\ell=0}^{\infty}\sum_{i,j=1}^{p+1}\left\{2r\sum_{k=0}^{t_0-p-1}(1-\lambda\delta)^{2k}\big(\mathbb{E}|\{m_k(\ell)\}_{ij}|^r\big)^{2/r}(\frac{k}{N})^{\beta}\right\} \\ &\leq \frac{K}{(N\lambda)^{\beta}}\sum_{\ell=0}^{\infty}\rho^{\ell}\lambda^{1/2} = O(\frac{\lambda^{1/2}}{(N\lambda)^{\beta}}). \end{aligned} \quad (53)$$

Thus we have proved (ii).

We now prove (iii). By using (42) we have

$$\begin{aligned} &(\mathbb{E}|\delta^B_{t_0,N}(u_0) - J^{(1,B),1}_{t_0,N}(u_0) - J^{(2,B),1}_{t_0,N}(u_0)|^{r/2})^{2/r} \\ =\ &(\mathbb{E}|J^{(3,B),1}_{t_0,N}(u_0)|^{r/2})^{2/r} + (\mathbb{E}|\sum_{i=1}^{3}J^{(i,B),2}_{t_0,N}(u_0)|^{r/2})^{2/r} + (\mathbb{E}|\sum_{i=1}^{3}H^{(i,B)}_{t_0,N}(u_0)|^{r/2})^{2/r} + O(\frac{1}{N^{\beta}} + \exp\{-\lambda\delta t_0\}). \end{aligned}$$

We partition $\sum_{i=1}^{3} J^{(i,B),2}_{t_0,N}$ into four terms:

$$\sum_{i=1}^{3} J^{(i,B),2}_{t_0,N} = A^B_{t_0,N} + B^{1,B}_{t_0,N} + B^{2,B}_{t_0,N} + B^{3,B}_{t_0,N} \quad (54)$$

where,

$$A^B_{t_0,N} = -\sum_{k=0}^{t_0-p-1}\lambda\{I - \lambda F(u_0)\}^k[F_{t_0-k-1,N} - F_{t_0-k-1}(u_0)] \times \left(\sum_{i=1}^{3} J^{(i,B),1}_{t_0-k-1,N}\right)$$

and, for $y \in \{1,2,3\}$,

$$B^{y,B}_{t_0,N} = -\sum_{k=0}^{t_0-p-1}\lambda\{I - \lambda F(u_0)\}^k[F_{t_0-k-1}(u_0) - F(u_0)]J^{(y,B),1}_{t_0-k-1,N}(u_0).$$

We first bound $A^B_{t_0,N}$. By using (50), (51), (53) and (119) we have

$$(\mathbb{E}|A^B_{t_0,N}|^{r/2})^{2/r} \leq \frac{K}{(N\lambda)^{2\beta}}.$$

By using (134) and (135) we have

$$\begin{aligned}(\mathbb{E}|B^{1,B}_{t_0,N}|^{r/2})^{r/2} &= \mathbb{E}\Big| - \sum_{k=0}^{t_0-p-1}\sum_{i=0}^{t_0-p-k-1} \lambda^2\{I-\lambda F(u_0)\}^{k+i}\bar{F}_{t_0-k-1}(u_0) \times \\ & \quad \{\mathcal{M}_{t_0-k-i-1,N} - \mathcal{M}_{t_0-k-i-1}(u_0)\}\Big|^{r/2})^{2/r} \leq K\lambda.\end{aligned}$$

Using a similar proof to the above to bound $(\mathbb{E}|J^{(3,B),1}_{t_0-k-1,N}(u_0)|^r)^{1/r}$ we can show that $\|B^{2,B}_{t_0,N}\|^E_{r/2} = O(\lambda^{1/2}(N\lambda)^{-\beta})$. By using (53) it is straightforward to show that $\|B^{3,B}_{t_0,N}\|^E_{r/2} = O((N\lambda)^{-2\beta}+\lambda^{1/2}(N\lambda)^{-\beta})$. Substituting the above bounds for $(\mathbb{E}|A^B_{t_0,N}|^{r/2})^{2/r}$, $(\mathbb{E}|B^{1,B}_{t_0,N}|^{r/2})^{2/r}$, $(\mathbb{E}|B^{2,B}_{t_0,N}|^{r/2})^{2/r}$ and $(\mathbb{E}|B^{3,B}_{t_0,N}|^{r/2})^{2/r}$ into (54) we obtain

$$(\mathbb{E}|J^{(1,B),2}_{t_0,N}(u_0) + J^{(2,B),2}_{t_0,N}(u_0) + J^{(3,B),2}_{t_0,N}(u_0)|^{r/2})^{2/r} = O(\delta_N) \quad (55)$$

We now prove (iii) by bounding for $y = 1,2,3$, $\|H^{(y,B)}_{t_0,N}(u_0)\|^E_2$. By using Hölder's inequality, (55), Theorem 4.1 and that $\{F_{t,N}\}$ are bounded random matrices we have

$$\begin{aligned}&(\mathbb{E}|H^{(1,B)}_{t_0,N}(u_0) + H^{(2,B)}_{t_0,N}(u_0) + H^{(3,B)}_{t_0,N}(u_0)|^2)^{1/2} \leq \sum_{k=0}^{t_0-p-1}\lambda\left(\mathbb{E}\|\big(\prod_{i=0}^{k-1}(I-\lambda F_{t_0-i,N})\big)\|^{2r/(r-4)}\right)^{(r-4)/2r} \\ &(\mathbb{E}|[F_{t_0-k-1,N} - F(u_0)]\{J^{(1,B),2}_{t_0-k-1,N}(u_0) + J^{(2,B),2}_{t_0-k-1,N}(u_0) + J^{(3,B),2}_{t_0-k-1,N}\}|^{r/2})^{2/r} \\ &\leq K\big(\frac{1}{(N\lambda)^{2\beta}} + \frac{\lambda^{1/2}}{(N\lambda)^{\beta}} + \lambda\big).\end{aligned} \quad (56)$$

Since $(\mathbb{E}|J^{(i,B),2}_{t_0,N}(u_0)|_2)^{1/2} \leq \mathbb{E}|J^{(i,B),2}_{t_0,N}(u_0)|^{r/2})^{2/r}$, by substituting (55) and (56) into (54) we have

$$\begin{aligned}&(\mathbb{E}|\delta^B_{t_0,N}(u_0) - J^{(1,B),1}_{t_0,N}(u_0) - J^{(2,B),1}_{t_0,N}(u_0|^2)^{1/2} \\ =\ & O(\frac{1}{(N\lambda)^{2\beta}} + \frac{\sqrt{\lambda}}{(N\lambda)^{\beta}} + \lambda).\end{aligned}$$

and we obtain (48). □

In the following lemma we show that $\delta^M_{t_0,N}(u_0)$ is dominated by $J^{M,1}_{t_0,N}(u_0)$, where

$$J^{M,1}_{t_0,N}(u_0) = \sum_{k=0}^{t_0-p-1}\lambda(I-\lambda F(u_0))^k\mathcal{M}_{t_0-k}(u_0). \quad (57)$$

We observe that $J^{M,1}_{t_0,N}(u_0)$ and $\mathcal{L}_{t_0}(u_0)$ (defined (8)) are the same.

Proposition 4.3 *Suppose Assumption 2.1 holds with $r > 4$. Let $\delta^{M}_{t_0,N}(u_0)$ and $J^{M,1}_{t_0,N}(u_0)$ be defined as in (36) and (57). Then for $|t_0/N - u_0| < 1/N$ and $\lambda N \geq (\log N)^{1+\varepsilon}$, for some $\varepsilon > 0$, we have*

$$(\mathbb{E}|J^{M,1}_{t_0,N}(u_0)|^r)^{1/r} = O(\sqrt{\lambda}) \tag{58}$$

$$(\mathbb{E}|\delta^{M}_{t_0,N}(u_0) - J^{M,1}_{t_0,N}(u_0)|_2)^{1/2} = O(\lambda + \frac{\sqrt{\lambda}}{(N\lambda)^{\beta}}). \tag{59}$$

PROOF. Since each component of the vector sequence $\{\mathcal{M}_t(u_0)\}$ is a martingale difference we can use the Burkhölder inequality componentwise and (110) to obtain

$$\begin{aligned}(\mathbb{E}|J^{M,1}_{t_0,N}(u_0)|^r)^{1/r} &\leq \lambda\big\{2r\sum_{k=0}^{t_0-p-1}(1-\delta\lambda)^{2k}(\mathbb{E}|Z^2_{t_0-k}-1|^r)^{2/r}(\mathbb{E}|\frac{\sigma_{t_0-k}(u_0)^2\mathcal{X}_{t_0-k-1}(u_0)}{|\mathcal{X}_{t_0-k-1}(u_0)|_1^2}|^r)^{2/r}\big\}^{1/2} \\ &\leq K\sqrt{\lambda}. \end{aligned} \tag{60}$$

Since $\delta^{M}_{t_0,N}(u_0) - J^{M,1}_{t_0,N}(u_0) = J^{M,2}_{t_0,N}(u_0) + H^{M}_{t_0,N}(u_0)$, to prove (59) we bound $J^{M,2}_{t_0,N}(u_0)$ and $H^{M}_{t_0,N}(u_0)$. By using the same arguments given in Proposition 4.2 and (60) we have

$$(\mathbb{E}|J^{M,2}_{t_0,N}(u_0)|^{r/2})^{2/r} \leq K\big(\lambda + \frac{\sqrt{\lambda}}{(N\lambda)^{\beta}}\big). \tag{61}$$

Finally, we obtain a bound for $H^{M}_{t_0,N}(u_0)$. By using Hölder's inequality, Theorem 4.1 and $(\mathbb{E}\|F_{t_0-k-1,N} - F(u_0)\|^{r/2})^{2/r} \leq 2(p+1)^2$ we have

$$\begin{aligned}(\mathbb{E}|H^{M}_{t_0,N}(u_0)|^2)^{1/2} &\leq 2\sum_{k=0}^{t_0-p-1}\lambda\|\big(\prod_{i=0}^{k-1}(I-\lambda F_{t-i-1,N})\big)\|^{2r/(r-4)})^{(r-4)/2r} \times \\ &\quad (\mathbb{E}|\{F_{t_0-k-1,N} - F(u_0)\}\{J^{M,2}_{t_0-k-1,N}(u_0)\}|^{r/2})^{2/r} \\ &\leq K\big(\lambda + \frac{\sqrt{\lambda}}{(N\lambda)^{\beta}}\big). \end{aligned} \tag{62}$$

Thus we have shown (59). □

By using Propositions 4.2 and 4.3 we have established that $J^{(1,B),1}_{t_0,N}(u_0)$, $J^{(2,B),1}_{t_0,N}(u_0)$ and $J^{M,1}_{t_0,N}(u_0)$ are the principle terms in $\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)$. More precisely we have shown

$$\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0) = J^{(1,B),1}_{t_0,N}(u_0) + J^{(2,B),1}_{t_0,N}(u_0) + J^{M,1}_{t_0,N}(u_0) + R^{(1)}_N \tag{63}$$

where $\|R^{(1)}_N\|^E_2 = O(\delta_N)$ (δ_N is defined in (7)). Furthermore, if $\lambda \to 0$, $\lambda N \to \infty$ as $N \to \infty$, then this leads to consistency of $\underline{\hat{a}}_{t_0,N}$. By using (63), we show below that an upper bound for the mean squared error can be immediately obtained.

PROOF of Theorem 2.1 By substituting the bounds for $(\mathbb{E}|J^{(1,B),1}_{t_0,N}(u_0)+J^{(2,B),1}_{t_0,N}(u_0)|^2)^{1/2}$ and $(\mathbb{E}|J^{M,1}_{t_0,N}(u_0)|^2)^{1/2}$ (in Propositions 4.2 and 4.3) into (63) we have $\mathbb{E}(|\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)|^2) = O(\frac{1}{(\lambda N)^{\beta}} + \lambda^{1/2} + \delta_N)^2$. Thus we have (11). □

4.3 Proof of Theorem 2.2

We can see from Propositions 4.2 and 4.3 that the $J_{t_0,N}^{(1,B),1}(u_0) + J_{t_0,N}^{(1,B),2}(u_0)$ and $J_{t_0,N}^{M,1}(u_0)$ are leading terms in the expansion of $\delta_{t_0,N}$. For this reason to prove Theorem 2.2 we need only to consider these terms. We observe that

$$J_{t_0,N}^{(1,B),1}(u_0) + J_{t_0,N}^{(2,B),1}(u_0) = \mathcal{R}_{t_0,N}(u_0) + R_N^{(2)}, \tag{64}$$

where have replaced $\mathcal{M}_{t,N}$ by $\mathcal{M}_t(\frac{t}{N})$ leading to the remainder

$$R_N^{(2)} = \sum_{k=0}^{t_0-p-1} \lambda\{I - \lambda F(u_0)\}^k \left\{\mathcal{M}_{t_0-k,N} - \mathcal{M}_{t_0-k}(\frac{t_0-k}{N})\right\}.$$

By using (22) we have $(\mathbb{E}|R_N^{(2)}|^r)^{1/r} \leq \frac{K}{N^\beta}$, for all t, N. Therefore under Assumption 2.1 with $r > 2$, if $\lambda N \geq (\log N)^{1+\varepsilon}$ for some $\varepsilon > 0$, then by substituting (64) into (63) we have

$$\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0) = \mathcal{L}_{t_0}(u_0) + \mathcal{R}_{t_0,N}(u_0) + R_N^{(3)}, \tag{65}$$

where $(\mathbb{E}|R_N^{(3)}|^2)^{1/2} = O(\delta_N)$. In the proposition below we show asymptotic normality of $\mathcal{L}_{t_0}(u_0)$, which we use together with (65) to obtain asymptotic normality of $\underline{\hat{a}}_{t_0,N}$

Proposition 4.4 *Suppose Assumption 2.1 holds with $r = 1$ and for some $\delta > 0$, $\mathbb{E}(Z_0^{4+\delta}) < \infty$. Let $\mathcal{L}_{t_0}(u_0)$ and $\Sigma(u)$ be defined as in (8) and (14) respectively. If $|t_0/N - u_0| < 1/N$, then we have*

$$\lambda^{-1/2}\mathcal{L}_{t_0}(u_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma(u_0)), \tag{66}$$

where $\lambda \to 0$ as $N \to \infty$ and $\lambda N \geq (\log N)^{1+\varepsilon}$, for some $\varepsilon > 0$.

PROOF. Since $\mathcal{L}_{t_0}(u_0)$ is the weighted sum of martingale differences the result follows from the martingale central limit theorem and the Cramér-Wold device. It is straightforward to check the conditional Lindeberg condition, hence we omit this verification. We simply note that by using (111) we obtain the limit of the conditional variance of $\mathcal{L}_{t_0}(u_0)$:

$$\mu_4 \sum_{k=0}^{t_0-p-1} \lambda^2\{I - \lambda F(u_0)\}^{2k}\sigma_{t_0-k}(u_0)^4 \frac{\mathcal{X}_{t_0-k-1}(u_0)\mathcal{X}_{t_0-k-1}(u_0)^T}{|\mathcal{X}_{t_0-k-1}(u_0)|_1^4} \xrightarrow{\mathcal{P}} \Sigma(u_0). \tag{67}$$

□

We use the proposition above to prove asymptotic normality of $\underline{\hat{a}}_{t_0,N}$ (after a bias correction).

PROOF of Theorem 2.2. It follows from (65) that

$$\lambda^{-1/2}\{\underline{a}_{t_0,N} - \underline{a}(u_0)\} \quad = \quad \lambda^{-1/2}\mathcal{R}_{t_0,N}(u_0) + \lambda^{-1/2}\mathcal{L}_{t_0}(u_0) + O_p(\lambda^{1/2}\delta_N).$$

Therefore, if $\lambda >> N^{-4\beta/(4\beta+1)}$ and $\lambda >> N^{-2\beta}$ then $\frac{\lambda^{-1/2}}{(N\lambda)^{2\beta}} \to 0$ and $\frac{\lambda^{-1/2}}{N^{\beta}} \to 0$ respectively, and we have

$$\lambda^{-1/2}\{\underline{a}_{t_0,N} - \underline{a}(u_0)\} \;=\; \lambda^{-1/2}\mathcal{R}_{t_0,N}(u_0) + \lambda^{-1/2}\mathcal{L}_{t_0}(u_0) + o_p(1).$$

By using the above and Proposition 4.4 we obtain (12). Finally, since $|\mathcal{R}_{t_0,N}(u_0)|_1 = O_p(\frac{1}{(N\lambda)^{\beta}})$, if $\lambda >> N^{-\frac{2\beta}{2\beta+1}}$ then $\lambda^{-1/2}\mathcal{R}_{t_0,N}(u_0) \xrightarrow{\mathcal{P}} 0$ and we have (13). □

5 An asymptotic expression for the bias under additional regularity

In this section, under additional assumptions on the smoothness of the parameters $\underline{a}(u)$, we obtain an exact expression for the bias, which we use to prove Theorem 2.3.

In the section above we have shown that $\delta^{B}_{t_0,N} \approx \mathcal{R}_{t_0,N}(u_0)$. Since $\mathcal{M}_t(u)$ is a function on $X_t(u)^2$ whose derivative exists, the derivative of $\mathcal{M}_t(u)$ also exists and is given by

$$\dot{\mathcal{M}}_t(u) = (Z_t^2 - 1)\big\{F_{t-1}(u)\dot{\underline{a}}(u) + \dot{F}_{t-1}(u)\underline{a}(u)\big\}, \tag{68}$$

where

$$\dot{F}_{t-1}(u) = \underline{Y}_{t-1}(u)\dot{\underline{Y}}_{t-1}(u)^T + \dot{\underline{Y}}_{t-1}(u)\underline{Y}_{t-1}(u)^T, \tag{69}$$

with $\underline{Y}_{t-1}(u) = \frac{1}{|\mathcal{X}_{t-1}(u)|_1}\mathcal{X}_{t-1}(u)$ and

$$\dot{\underline{Y}}_{t-1}(u) = \frac{-\sum_{j=1}^{p}\dot{X}^2_{t-j}(u)}{[1+\sum_{j=1}^{p}X_{t-j}(u)^2]^2}\begin{pmatrix}1\\ X_{t-1}(u)^2\\ \dots\\ X_{t-p}(u)^2\end{pmatrix} + \frac{1}{1+\sum_{j=1}^{p}X_{t-j}(u)^2}\begin{pmatrix}0\\ \dot{X}^2_{t-1}(u)\\ \dots\\ \dot{X}^2_{t-p}(u)\end{pmatrix}$$

It is interesting to note that, like $\{\mathcal{M}_t(u)\}_t$, $\{\dot{\mathcal{M}}_t(u)\}_t$ is a sequence of vector martingale differences. We will use $\dot{\mathcal{M}}_t(u)$ to refine the approximation $\mathcal{R}_{t_0,N}(u_0)$ and show $\mathcal{R}_{t_0,N}(u_0) \approx \tilde{\mathcal{R}}_{t_0,N}(u_0)$, where

$$\tilde{\mathcal{R}}_{t_0,N}(u_0) = \sum_{k=0}^{t_0-1}\lambda\,(1-\lambda F(u_0))\,(\frac{t_0-k}{N} - u_0)\left\{\dot{\mathcal{M}}_{t_0-k}(u_0) + F(u_0)\dot{\underline{a}}(u_0)\right\}. \tag{70}$$

We use this to obtain Theorem 2.3.

Lemma 5.1 *Suppose Assumption 2.2 holds with some $r \geq 2$. Then we have*

$$\big|\dot{\mathcal{M}}_t(u) - \dot{\mathcal{M}}_t(v)\big|_1 \leq K|u-v|^{\beta'}|Z_t^2+1|L_t, \tag{71}$$

where $L_t = \big\{1+\sum_{k=1}^{p}W_{t-k}\big\}^2$, with $(\mathbb{E}(L_t^{r/2})^{2/r} < \infty$. and almost surely

$$\mathcal{M}_t(v) = \mathcal{M}_t(u) + (v-u)\dot{\mathcal{M}}_t(u) + (u-v)^{1+\beta'}R_t(u,v) \tag{72}$$

where $|R_t(u,v)| \leq L_t$.

PROOF. By using (68), $\underline{a}(u) \in Lip(1+\beta')$ and that $|F_{t-1}(u)|_1 \leq (p+1)^2$ we have

$$\begin{aligned}
& |\dot{\mathcal{M}}_t(u) - \dot{\mathcal{M}}_t(v)|_1 \\
\leq \quad & K|Z_t^2+1|\Big\{|F_{t-1}(u) - F_{t-1}(v)|_1 + |\dot{F}_{t-1}(u) - \dot{F}_{t-1}(v)|_1 + \sup_u |\dot{F}_{t-1}(u)|_1 |u-v|\Big\}.
\end{aligned} \tag{73}$$

In order to bound (73) we consider $F_t(u)$ and its derivatives. We see that $|F_{t-1}(u) - F_{t-1}(v)|_1 \leq K|u-v|L_t$, thus bounding the first bound on the right hand side of (73). To obtain the other bounds we use (69). Now by using (23) and Lemma 4.2 we have that $\sup_u |\dot{F}_{t-1}(u)|_1 \leq KL_t$ and $|\dot{F}_{t-1}(u) - \dot{F}_{t-1}(v)|_1 \leq KL_t^2|u-v|^{\beta}$. Altogether this verifies (71). By using the Cauchy-Schwartz inequality we have that $(\mathbb{E}(L_t^{r/2})^{2/r} < \infty$.

Finally we prove (73). Since L_t is a well defined random variable almost surely all the paths of $\mathcal{M}_t(u) \in Lip(1+\beta')$. Therefore, there exists a set $\mathcal{N}$, such that $P(\mathcal{N}) = 0$ and for every $\omega \in \mathcal{N}^c$ we can make a Taylor expansion of $\mathcal{M}_t(u, \omega)$ about w and obtain (72). □

We now use (72) to show that $\mathcal{R}_{t_0,N}(u_0) \approx \tilde{\mathcal{R}}_{t_0,N}(u_0)$.

Lemma 5.2 *Suppose Assumption 2.2 holds with $r \geq 2$. Then if $|t_0/N - u_0| < 1/N$ we have*

$$\mathcal{R}_{t_0,N}(u_0) = \tilde{\mathcal{R}}_{t_0,N}(u_0) + R_N^{(4)}, \tag{74}$$

where $\mathbb{E}(|R_N^{(4)}|^r)^{1/r} = O(\frac{1}{(N\lambda)^{1+\beta'}})$.

PROOF. To obtain the result we make a Taylor expansion of $\underline{a}(\frac{t-k}{N})$ and $\mathcal{M}_{t-k}(\frac{t-k}{N})$ about u_0, and substitute this into $\mathcal{R}_{t_0,N}$. By using Lemma 5.1 we obtain the desired result. □

In the lemma below we consider the mean and variance of the bias and stochastic terms $\mathcal{R}_{t_0,N}(u_0)$ and $\mathcal{L}_{t_0}(u_0)$, which we use to obtain an asymptotic expression for the bias. We will use the following results. Since $\inf_u \lambda_{min}(F(u)) > C > 0$ (see (108)), we have

$$\sum_{k=0}^{t-1} \lambda\{I - \lambda F(u)\}^k F(u) \quad \to \quad I, \tag{75}$$

$$\Big\|\sum_{k=0}^{t-1} \lambda^2\{I - \lambda F(u)\}^{2k} (\frac{k}{N})^2\Big\| \quad = \quad O(\frac{1}{N^2\lambda}), \tag{76}$$

if $\lambda \to 0$, $t\lambda \to \infty$ as $t \to \infty$.

Lemma 5.3 *Suppose Assumption 2.2 holds with $r > 4$. Let $\mathcal{R}_{t_0,N}(u_0)$, $\mathcal{L}_{t_0}(u_0)$ and $\Sigma(u)$ be defined as in (8) and (14) respectively. Then if $|t_0/N - u_0| < 1/N$ we have*

$$\mathbb{E}\left(\mathcal{R}_{t_0,N}(u_0)\right) \quad = \quad -\frac{1}{N\lambda}F(u_0)^{-1}\dot{\underline{a}}(u) + O(\frac{1}{(N\lambda)^{1+\beta'}}), \tag{77}$$

$$var(\mathcal{R}_{t_0,N}(u_0)) \quad = \quad O(\frac{1}{N^2\lambda} + \frac{1}{N}), \tag{78}$$

$$var(\mathcal{L}_{t_0}(u_0)) = \lambda\Sigma(u_0) + o(\lambda) \tag{79}$$

and $\mathbb{E}\left(\mathcal{L}_{t_0}(u_0)\right) = 0$.

PROOF. Since $\{\frac{\partial \mathcal{M}_k(u)}{\partial u}\}$ are martingale differences, by using appling (74) to (75) we have (77). We now show (78). By using (76), $\sup_u |\frac{\partial \underline{a}(u)}{\partial u}| < \infty$ and $\sup_u |F_t(u)| < (p+1)^2$ we have

$$\begin{aligned}
\text{var}\big\{\mathcal{R}_{t_0,N}(u_0)\big\} \quad &= \quad \mu_4 \sum_{k=0}^{t_0-p-1} \lambda^2\{I - \lambda F(u_0)\}^{2k}(\frac{k}{N})^2 \text{var}\big\{F_{t_0-k-1}(u_0)\underline{\dot{a}}(u_0) + \dot{F}_{t_0-k-1}(u_0)\underline{a}(u_0)\big\} + O(\frac{1}{N^2}) \\
&= \quad O(\frac{1}{N^2\lambda} + \frac{1}{N^2}).
\end{aligned}$$

It is straightforward to show (79) using (111). Finally, since $\mathcal{L}_{t_0}(u_0)$ is the sum of martingale differences, $\mathbb{E}(\mathcal{L}_{t_0}(u_0)) = 0$. □

From the lemma above it immediately follows that

$$\mathcal{R}_{t_0,N}(u_0) = -\frac{1}{N\lambda}F(u_0)^{-1}\underline{\dot{a}}(u_0) + O_p\left\{\frac{1}{(N\lambda)^{1+\beta'}} + \frac{1}{N\lambda^{1/2}}\right\} \tag{80}$$

and $(N\lambda)\mathcal{R}_{t_0,N}(u_0) \xrightarrow{\ell_2} -F(u_0)^{-1}\underline{\dot{a}}(u_0)$. We now use the above prove Theorem 2.3.

PROOF of Theorem 2.3 By substituting (80) into (65) (with $\beta = 1$) we have

$$\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0) = \frac{-1}{N\lambda}F(u_0)^{-1}\underline{\dot{a}}(u_0) + \mathcal{L}_{t_0}(u_0) + \delta'_N R_N^{(5)}, \tag{81}$$

where $(\mathbb{E}|R_N^{(5)}|^2)^{1/2} < \infty$ and $\delta'_N = \frac{1}{(N\lambda)^{1+\beta'}} + \frac{1}{(N\lambda)^2} + \frac{1}{\lambda^{1/2}N} + \lambda + \frac{1}{N^{1/2}}$. By using the above and $\text{var}(\mathcal{L}_{t_0}(u_0)) = \lambda\Sigma(u_0)$ we have

$$\mathbb{E}|\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)|^2 \quad = \quad \lambda\, tr\{\Sigma(u_0)\} + \frac{1}{(N\lambda)^2}\big|F(u_0)^{-1}\underline{\dot{a}}(u_0)\big|^2 + O\left(\left[(N\lambda)^{-1} + \sqrt{\lambda} + \delta'_N\right]\delta'_N\right),$$

which gives (16). In order to prove (17) we use (81) and Proposition 4.4. We first note that if $\lambda^{-1/2}/(N\lambda)^{1+\beta'} \to 0$, then by using (81) we have

$$\lambda^{-1/2}\{\underline{\hat{a}}_{t_0,N} - \underline{a}(u_0)\} + \lambda^{-1/2}\frac{1}{\lambda N}F(u_0)^{-1}\underline{\dot{a}}(u_0) = \lambda^{-1/2}\mathcal{L}_{t_0}(u_0) + o_p(1).$$

Therefore by using Proposition 4.4 we have (17). □

6 A lower bound for the minimax rate

PROOF of Theorem 2.4 We derive a lower bound for the minimax rate using the van Trees inequality (cf. Gill and Levit (1995) and Moulines et al. (2005)). In order to do this we define a prior distribution on a subset of $\mathcal{C}^{\beta}(L)$, which we then go on to bound using the van Trees inequality.

Let us suppose the function $\phi : [0,1] \to \mathbb{R}^{+}$ is such that $\|\phi\|_{\beta} \leq L$, $\sup_u |\phi(u)| = 1$ and $\int_0^1 \phi(u)^2 du = 1$. Let $\underline{I}^T = (1,\ldots,1) \in \mathbb{R}^{p+1}$, $\underline{J}^T = (\rho_1, 0, \ldots, 0) \in \mathbb{R}^{p+1}$ and define the set

$$\mathcal{C}_{w_N} = \left\{ \underline{J} + \eta\phi(w_n(\cdot - u_0))\underline{I};\ \eta \in [0, \min(w_N^{-\beta}, p^{-1}(1-\delta))] \right\}.$$

It is straightforward to show for a large enough w_N that $\mathcal{C}_{w_N} \subseteq \mathcal{C}^{\beta}(L)$. Hence if $\underline{a}(\cdot) \in \mathcal{C}_{w_N}$, and are the parameters of the tvARCH process $\{X_{t,N}\}$, then Assumption 2.1(i,ii,iii) is satisfied and $\mathbb{E}(X_{t,N}^2) < (\rho_1+1)/(1-\delta)$.

Let $\lambda : [0,1] \to \mathbb{R}^{+}$ be a density which is continuous with respect to the Lebesgue measure and define $\lambda_N(\cdot) = w_N^{\beta}\lambda(w_N^{\beta}\cdot)$. It is clear that we can use $\lambda_N(\cdot)$ as a density on the set $\mathcal{C}_{w_N}$. For each $\underline{a}(\cdot) = \underline{J} + \eta\phi(w_n(\cdot - u_0))\underline{I} \in \mathcal{C}_{w_N}$ we define the tvARCH process

$$X_{t,N} = \sigma_{t,N} Z_t \qquad \sigma_{t,N}^2 = \rho_1 + \eta\phi(w_n(\frac{t}{N} - u_0))\mathcal{X}_{t-1,N},$$

where $\mathcal{X}_{t-1,N}^T = (1, X_{t-1,N}^2, \ldots, X_{t-p,N}^2)$. Let $\mathbb{E}_{\eta}$ be the expectation under the measure generated by the parameters $\underline{a}(\cdot) = (\underline{J} + \eta\phi(w_n(\cdot - u_0))\underline{I})$ and $\mathbb{E}_{\lambda_N}$ the expectation under the measure $\lambda_N(\cdot)$. Since $\mathcal{C}_{w_N} \subseteq \mathcal{C}^{\beta}(L)$ it can be shown that

$$\begin{aligned}
& \inf_{\underline{a}_{t_0,N} \in \sigma(X_{t,N};t=1,\ldots,N)} \ \sup_{\underline{a}(u) \in \mathcal{C}^{\beta}(L)} \mathbb{E}[\underline{x}^T(\underline{a}_{t_0,N} - \underline{a}(u_0))(\underline{a}_{t_0,N} - \underline{a}(u_0))^T \underline{x}] \\
\geq \quad & \mathbb{E}_{\lambda_N} \inf_{\underline{a}_{t_0,N} \in \sigma(X_{t,N};t=1,\ldots,N)} \{\mathbb{E}_{\eta}[\underline{x}^T(\underline{a}_{t_0,N} - \underline{a}(u_0))(\underline{a}_{t_0,N} - \underline{a}(u_0))^T \underline{x}]\}. \qquad (82)
\end{aligned}$$

We will assume that $\phi(\cdot)$ and ρ_1 are known. Now by using the van Trees inequality we have for all $|\underline{x}| \leq 1$ and $\underline{a}_{t_0,N} \in \sigma(X_{t,N}; t = 1,\ldots,N)$, that

$$\mathbb{E}_{\lambda_N}\{\mathbb{E}_{\eta}[\underline{x}^T(\underline{a}_{t_0,N} - \underline{a}(u_0))(\underline{a}_{t_0,N} - \underline{a}(u_0))^T \underline{x}]\} \geq \frac{1}{I(\lambda_N) + \mathbb{E}_{\lambda_N}(I(\eta))}, \qquad (83)$$

where

$$\mathcal{I}(\lambda_N) = \int \lambda_N(x) \left(\frac{d \log \lambda_N(x)}{dx} \right)^2 dx$$

and

$$\mathcal{I}_N(\eta) = \int \left(\frac{d \log f_{\eta}(x_{1,N},\ldots,x_{N,N})}{d\eta} \rfloor_{\eta=\eta} \right)^2 f_{\eta}(x_{1,N},\ldots,x_{N,N}) dy_{1,N} \ldots, dy_{1,N}$$

with $f_{\eta}(x_{1,N},\ldots,x_{N,N})$ the density of $\{X_{t,N}\}_{t=1}^{N}$. We now derive asymptotic expressions for $I(\lambda_N)$ and $I(\eta)$.

Using that $P(X_{t,N} \leq y | \mathcal{X}_{t-1,N}) = \mathbb{P}_Z(y/\sigma_{t,N}^2)$, where $\mathbb{P}_Z$ is the distribution function of Z_t. The log density of $\{X_{t,N}\}_{t=1}^N$ is

$$\log f_\eta\{x_{N,N}, \ldots, x_{1,N}\} = \sum_{t=p+1}^{N} \log \frac{1}{\sigma_{t,N}} f_Z\{\frac{x_{t,N}}{\sigma_{t,N}}\} + \log f_{\mathcal{X}_{p,N}}(x_{1,N}, \ldots, x_{p,N}),$$

where f_Z is the density function of Z_t, $\sigma_{t,N} = [1 + \eta\phi(w_N(\frac{t}{N} - u_0))|\underline{x}_{t-1,N}|_1]^{1/2}$, with $\underline{x}_{t-1,N} = (1, x_{t-1,N}^2, \ldots, x_{t-p,N}^2)$ and $f_{\mathcal{X}_{p,N}}$ is the density of $\mathcal{X}_{p,N}$. Define

$$\ell_{t,N}(\zeta; x_{t,N}, \underline{x}_{t-1,N}) = -\frac{1}{2}\log[\rho_1 + \zeta\phi(w_N(\frac{t}{N} - u_0))|\underline{x}_{t-1,N}|_1] - \frac{x_{t,N}}{2[\rho_1 + \zeta\phi(w_N(\frac{t}{N} - u_0))|\underline{x}_{t-1,N}|_1]^{1/2}}.$$

Differentiating $\ell_{t,N}(\zeta; x_{t,N}, \underline{x}_{t-1,N})$ wrt. ζ and evaluating it at η, it is straightforward to see that

$$2\dot{\ell}_{t,N}(\eta\,; x_{t,N}, \underline{x}_{t-1,N}) = -\frac{\phi(w_N(\frac{t}{N} - u_0))|\underline{x}_{t-1,N}|_1}{\sigma_{t,N}^2}\left(\rho_1 + z_t \frac{d\log f_Z(z_t)}{dz_t}\right),$$

where $z_t = x_{t,N}/\sigma_{t,N}$. Now let us suppose

$$\mathcal{A}_{p,N} = \mathbb{E}_\eta(\frac{\partial \log f_{\mathcal{X}_{p,N}}\{X_{1,N}, \ldots, X_{p,N}\}}{\partial\eta})^2 < \infty.$$

It is well known that the derivative of the log conditional densities are a sequence of martingale differences; that is $\{\dot{\ell}_{t,N}(\eta; X_{t,N}, \mathcal{X}_{t-1,N}) : t = p+1, \ldots, N\}$, are martingale differences. From this it follows that for $t > p$, we have

$$\mathbb{E}_\eta\left(\frac{\partial \log f_{\mathcal{X}_{p,N}}(X_{1,N}, \ldots, X_{p,N})}{\partial\eta}\dot{\ell}_{t,N}(\eta; X_{t,N}, \mathcal{X}_{t-1,N})\right) = 0.$$

Therefore, the corresponding Cramér-Rao bound is

$$\begin{aligned}
\mathcal{I}_N(\eta) &= \mathbb{E}_\eta\left\{\frac{\partial \log f_{\mathcal{X}_{p,N}}(X_{1,N}, \ldots, X_{p,N})}{\partial\eta} + \sum_{t=p+1}^{N} \dot{\ell}_{t,N}(\eta; X_{t,N}, \mathcal{X}_{t-1,N})\right\}^2 \\
&= \frac{1}{4}\sum_{t=p+1}^{N} \phi\{w_N(\frac{t}{N} - u_0)\}^2 \mathbb{E}\left[\frac{|\mathcal{X}_{t-1,N}|_1}{\sigma_{t,N}^2}\left(\rho_1 + Z_t\frac{d\log f_Z(Z_t)}{dZ_t}\right)\right]^2 + \mathcal{A}_{p,N}.
\end{aligned}$$

Since $\eta \to 0$, as $N \to \infty$, for large enough w_N we have that $w_N^{-\beta} p\mathbb{E}(Z_t^4)^{1/2} < 1$, which implies $\sup_t \mathbb{E}_\eta(X_{t,N}^4) < [(\rho_1 + w_N^{-\beta})\mathbb{E}(Z_t^4)^{1/2}/(1 - pw_N^{-\beta}\mathbb{E}(Z_t^4)^{1/2})]^2$. Therefore, since $|\underline{x}_{t-1,N}|/\sigma_{t,N}^2 < |\underline{x}_{t-1,N}|/\rho_1$ we obtain

$$\begin{aligned}
\mathcal{I}_N(\eta) &\leq \frac{1}{4}\sup_t \mathbb{E}_\eta(|\mathcal{X}_{t-1,N}|_1^2)\sum_{t=p+1}^{N} \phi\{w_N(u_0 - \frac{t}{N})\}^2 \mathcal{Z} + \mathcal{A}_{p,N} \\
&\leq \mathcal{Z}(p+1)\frac{(\rho_1 + w_N^{-\beta})^2\mathbb{E}(Z_t^4)}{4\rho_1[(1 - pw_N^{-\beta}\mathbb{E}(Z_t^4)^{1/2})]^2}\sum_{t=p+1}^{N} \phi\{w_N(u_0 - \frac{t}{N})\}^2 + \mathcal{A}_{p,N} \\
&\leq K(\mathcal{Z}, \mathbb{E}(Z_0^4), \rho_1)\frac{N}{w_N}\int_0^1 \phi(x)^2 dx + o(\frac{N}{w_N}), \qquad (84)
\end{aligned}$$

where $\mathcal{Z} = \mathbb{E}(\rho_1 + Z_t \frac{d\log f_Z(x)}{dx}\rfloor_{x=Z_t})^2$ and for a large enough w_N, $K(\mathcal{Z}, \mathbb{E}(Z_0^4), \rho_1) = \mathcal{Z}(p+1)\frac{(2\rho_1)^2\mathbb{E}(Z_t^4)}{4\rho_1[(1-\mathbb{E}(Z_t^4)^{-1/2})]^2}$. We now bound $I(\lambda_N)$. It is clear that

$$I(\lambda_N) = w_N^{2\beta} I(\lambda) \quad \text{where} \quad I(\lambda) \;=\; \int_0^1 \{\frac{\partial \log \lambda(u)}{\partial u}\}^2 \lambda(u)\mathrm{d}u. \tag{85}$$

Substituting (85) and (84) into (82) and (83) we obtain

$$\begin{aligned} & \inf_{\underline{a}_{t_0,N}\in\sigma(X_{t,N};t=1,\ldots,N)} \sup_{\underline{a}(u)\in\mathcal{C}^\beta(L)} \mathbb{E}[\underline{x}^T(\underline{a}_{t_0,N} - \underline{a}(u_0))(\underline{a}_{t_0,N} - \underline{a}(u_0))^T\underline{x}] \\ \geq & \quad \frac{1}{K(\mathcal{Z}, \mathbb{E}(Z_0^4), \rho_1)\frac{N}{w_N}\int_0^1 \phi(x)^2 dx + w_N^{2\beta} I(\lambda)} \end{aligned}$$

By letting $w_N = O(N^{1/(1+2\beta)})$, the above is minimised and we obtain

$$\inf_{\underline{a}_{t_0,N}\in\sigma(X_{t,N};t=1,\ldots,N)} \sup_{\underline{a}(u)\in\mathcal{C}^\beta(L)} \mathbb{E}[\underline{x}^T(\underline{a}_{t_0,N} - \underline{a}(u_0))(\underline{a}_{t_0,N} - \underline{a}(u_0))^T\underline{x}] \geq K(\mathcal{Z}, \mathbb{E}(Z_0^4), \rho_1) N^{-2\beta/(2\beta+1)},$$

where $C(\mathcal{Z}, \mathbb{E}(Z_0^4), \rho_1)$ is a constant which depends only on $\mathcal{Z}, \mathbb{E}(Z_0^4), \rho_1$, thus giving the required result. □

A Appendix

A.1 Mixingale properties of $\phi(\mathcal{X}_{t,N})$

Our object in this section is to prove Proposition 4.1. We do this by using the random vector representation of the tvARCH process. Let

$$\begin{aligned} \breve{A}_t(u) &= \begin{pmatrix} a_1(u)Z_t^2 & a_2(u)Z_t^2 & \ldots & a_{p-1}(u)Z_t^2 & a_p(u)Z_t^2 \\ 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \\ A_t(u) &= \begin{pmatrix} 0 & \underline{0}_p^T \\ \underline{0}_p & \breve{A}_t(u) \end{pmatrix}, \end{aligned} \tag{86}$$

$\underline{b}_t(u)^T = (1, a_0(u)Z_t^2, \underline{0}_{p-1}^T)$. By using the definition of the tvARCH process given in (1) we have that the tvARCH vectors $\{\mathcal{X}_{t,N}\}_t$ satisfy the representation

$$\mathcal{X}_{t,N} = A_t(\frac{t}{N})\mathcal{X}_{t-1,N} + \underline{b}_t(\frac{t}{N}). \tag{87}$$

We note that (87) looks like a vector AR process, the difference is that $A_t(\frac{t}{N})$ is a random matrix. However, similar to the vector AR case, it can be shown that the product $\prod_{k=0}^t A_k(\frac{k}{N})$ decays exponentially. It is this property which we use to prove Proposition 4.1.

Let $\mathcal{A}_N(t,j) = \big\{ \prod_{i=0}^{j-1} A_{t-i}(\frac{t-i}{N})\big\}$, $\mathcal{A}(u,t,j) = \big\{ \prod_{i=0}^{j-1} A_{t-i}(u)\big\}$, $\mathcal{A}_N(t,0) = I_{p+1}$ and $\mathcal{A}(u,t,0) = I_{p+1}$. By expanding (87) we have

$$\mathcal{X}_{t,N} \quad = \quad \mathcal{A}_N(t,k)\mathcal{X}_{t-k,N} + \sum_{j=0}^{k-1} \mathcal{A}_N(t,j)\, \underline{b}_{t-j}(\frac{t-j}{N}). \tag{88}$$

Suppose A is a $(p+1)\times(p+1)$ dimensional random matrix, with $(i,j)th$ element A_{ij}. Let $[A]_q$ be a $(p+1)\times(p+1)$ dimensional matrix where

$$[A]_q = \big\{ \big(\mathbb{E}|A_{ij}^q|\big)^{1/q}; i,j = 1,\ldots,p+1\big\}.$$

We will make frequent use of the following inequalities, which can easily be derived by repeatedly applying Hölder's and Minkowski's inequalities. Suppose A and $\underline{X}$ are independent random matrices and vectors, respectively, then $\big(\mathbb{E}\big|A\underline{X}\big|^q\big)^{1/q} \leq \big|[A]_q[\underline{X}]_q\big|_1$, and if $\underline{X}$ has only positive elements then $\mathbb{E}\left(|A\underline{X}|_1\big|\,\underline{X}\right) \leq \mathbb{E}\left(|[A]_1\underline{X}\,|_1\big|\,\underline{X}\right)$. Now by using Subba Rao (2004), Example 2.3 and Corollary A.2, we have

$$\|[\mathcal{A}_N(t,k)]_q\| \leq K\,\rho^k \quad \text{and} \quad \|[\mathcal{A}(u,t,k)]_q\| \leq K\rho^k, \tag{89}$$

where K is a finite constant.

PROOF of Proposition 4.1 We first prove (25). Let $\mathcal{C}_N(t,k) := \sum_{j=0}^{k-1} \mathcal{A}_N(t,j)\, \underline{b}_{t-j}(\frac{t-j}{N})$, that is

$$\mathcal{X}_{t,N} = \mathcal{A}_N(t,k)\mathcal{X}_{t-k,N} + \mathcal{C}_N(t,k) \tag{90}$$

with $\mathcal{A}_N(t,k), \mathcal{C}_N(t,k) \in \sigma(Z_t,\ldots,Z_{t-k+1})$ and $\mathcal{X}_{t-k,N} \in \mathcal{F}_{t-k}$. In particular we have

$$\mathbb{E}\{\phi\big(\mathcal{C}_N(t,k)\big)|\mathcal{F}_{t-k})\} = \mathbb{E}\{\phi\big(\mathcal{C}_N(t,k)\big)\}. \tag{91}$$

Furthermore, by using Minkowski's inequality it can be shown that

$$\{\mathbb{E}(|\mathcal{A}_N(t,k)\mathcal{X}_{t-k,N}|^q|\mathcal{F}_{t-k})\}^{1/q} \leq K|[\mathcal{A}_N(t,k)]_q\mathcal{X}_{t-k,N}|_1.$$

The Lipschitz continuity of ϕ and (88) now imply $|\phi\big(\mathcal{X}_{t,N}\big) - \phi\big(\mathcal{C}_N(t,k)\big)| \leq K|\mathcal{A}_N(t,k)\mathcal{X}_{t-k,N}|$. Therefore, by using the above we obtain

$$\begin{aligned}
&|\mathbb{E}\{\phi(\mathcal{X}_{t,N})|\mathcal{F}_{t-k})\} - \mathbb{E}\{\phi(\mathcal{X}_{t,N})\}|_1 \\
&= |\mathbb{E}_{t-k}\{\phi(\mathcal{X}_{t,N}) - \phi\big(\mathcal{C}_N(t,k)\big)|\mathcal{F}_{t-k})\} - \mathbb{E}\{\phi(\mathcal{X}_{t,N}) - \phi\big(\mathcal{C}_N(t,k)\big)\}|_1 \\
&\leq K\big(\,|[\mathcal{A}_N(t,k)]_1\mathcal{X}_{t-k,N}|_1 + \mathbb{E}\{|[\mathcal{A}_N(t,k)]_1\mathcal{X}_{t-k,N}|_1\,\big) \\
&\leq K\,\|[\mathcal{A}_N(t,k)]_1\||\mathcal{X}_{t-k,N}|_1 + \mathbb{E}\big(\|[\mathcal{A}_N(t,k)]_1\||\mathcal{X}_{t-k,N}|_1\big) \\
&\leq K\,\|[\mathcal{A}_N(t,k)]_1\|\big(|\mathcal{X}_{t-k,N}|_1 + \mathbb{E}|\mathcal{X}_{t-k,N}|_1\big) \\
&\leq K\rho^k\left(1 + |\mathcal{X}_{t-k,N}|_1\right)
\end{aligned} \tag{92}$$

since $\mathbb{E}|\mathcal{X}_{t-k,N}|_1$ is uniformly bounded thus giving (25). The proof of (26) is the same as the proof above, so we omit details. (27) follows from (25) with the triangular inequality.

To prove (28) we use (90). Since

$$\mathbb{E}\{\phi(\mathcal{C}_N(t,k))|\mathcal{F}_{t-k}\} = \mathbb{E}\{\phi(\mathcal{C}_N(t,k))\} = \mathbb{E}\{\phi(\mathcal{C}_N(t,k))|\mathcal{F}_{t-k-j}\}$$

we obtain as in (92)

$$\begin{aligned}&|\mathbb{E}\{\phi(\mathcal{X}_{t,N})|\mathcal{F}_{t-k}\} - \mathbb{E}\{\phi(\mathcal{X}_{t,N})|\mathcal{F}_{t-k-j}\}|_1 \\ &\quad \le K\,\mathbb{E}\{|[\mathcal{A}_N(t,k)]_1\mathcal{X}_{t-k,N}|_1|\mathcal{F}_{t-k}\} + \mathbb{E}\{|[\mathcal{A}_N(t,k)]_1\mathcal{X}_{t-k,N}|_1\big|\mathcal{F}_{t-k-j}\} \\ &\quad \le K\rho^k\left(|\mathcal{X}_{t-k,N}|_1 + |\mathcal{X}_{t-k,N}|_1\right)\end{aligned}$$

which gives (28). To prove (29) we use (92). We first note that

$$\left[\mathbb{E}|C_N(t,k)|^q\big|\mathcal{F}_{t-k}\right]^{1/q} \quad \le \quad K\sum_{j=0}^{k-1}\left|[\mathcal{A}_N(t,j)]_q[\underline{b}_{t-j}(\frac{t-j}{N})]_q\right|_1 \le \frac{K}{1-\rho}$$

and by using Minkowski's inequality and the equivalence of norms, there exists a constant K independent of $\mathcal{X}_{t,N}$ such that

$$\left\{\mathbb{E}\left(|\mathcal{X}_{t,N}|^q\big|\mathcal{F}_{t-k}\right)\right\}^{1/q} \le \left\{\mathbb{E}\left(|\mathcal{A}_N(t,k)\mathcal{X}_{t-k,N}|^q\big|\mathcal{F}_{t-k}\right)\right\}^{1/q} + \left\{\mathbb{E}\left(|C_N(t,k)|^q\big|\mathcal{F}_{t-k}\right)\right\}^{1/q}.$$

Now by using $\left\{\mathbb{E}\left(|\mathcal{A}_N(t,k)\mathcal{X}_{t-k,N}|^q\big|\mathcal{F}_{t-k}\right)\right\}^{1/q} \le K\rho^k|\mathcal{X}_{t-k,N}|_1$ we have

$$\mathbb{E}\big\{|\mathcal{X}_{t,N}|^q\big|\mathcal{F}_{t-k}\big\} \le \big\{K\rho^k|\mathcal{X}_{t-k,N}|_1 + \frac{K}{1-\rho}\big\}^q \le K|\mathcal{X}_{t-k,N}|_1^q,$$

hence we obtain (29) □

We use the corollary below to prove Lemma A.7.

Corollary A.1 *Suppose Assumption 2.1 holds with* $r = q_0$ *and* $\mathbb{E}(Z_0^{4q_0}) < \infty$. *Let* $f, g : \mathbb{R}^{p+1} \to \mathbb{R}^{(p+1)\times(p+1)}$ *be Lipschitz continuous functions, such that for all positive* $\underline{x} \in \mathbb{R}^{p+1}$, $|f(\underline{x})|_{abs} \le \mathbb{I}_{p+1}$ *and* $|g(\underline{x})|_{abs} \le \mathbb{I}_{p+1}$, *where* $\mathbb{I}_{p+1}$ *is a* $p+1\times p+1$*-dimensional matrix with one in all the elements, and* $|A|_{abs} = (|A_{i,j}| : i,j = 1,\dots,p+1)$. *Then for* $q \le q_0$ *we have*

$$\begin{aligned}&\left[\mathbb{E}\left(\mathbb{E}\big\{(Z_{t-k+1}^2-1)f(\mathcal{X}_{t-k,N})g(\mathcal{X}_t(u))\big|\mathcal{F}_{t-k-j}\big\} - \mathbb{E}\big\{(Z_{t-k+1}^2-1)f(\mathcal{X}_{t-k,N})g(\mathcal{X}_t(u))\big\}\right)^q\right]^{1/q} \\ \le\ & K\{\mathbb{E}(Z_0^4)+1\}\rho^j\left((\mathbb{E}|\mathcal{X}_{t-k-j,N}|^q)^{1/q} + (\mathbb{E}|\mathcal{X}_{t-k-j}(u)|^q)^{1/q}\right), \qquad (93)\end{aligned}$$

and

$$\begin{aligned}&\left[\mathbb{E}\left(\mathbb{E}\big\{(Z_{t-k+1}^2-1)f(\mathcal{X}_{t-k}(u))g(\mathcal{X}_t(u))\big|\mathcal{F}_{t-k-j}\big\} - \mathbb{E}\big\{(Z_{t-k+1}^2-1)f(\mathcal{X}_{t-k}(u))g(\mathcal{X}_t(u))\big\}\right)^q\right]^{1/q} \\ \le\ & K\{\mathbb{E}(Z_0^4)+1\}\rho^j\left((\mathbb{E}|\mathcal{X}_{t-k-j}(u)|^q)^{1/q} + (\mathbb{E}|\mathcal{X}_{t-k-j}(u)|^q)^{1/q}\right), \qquad (94)\end{aligned}$$

for $j,k \ge 0$, *where* ρ *is such that* $(1-\eta) < \rho < 1$.

PROOF. We give the proof of (93) only, the proof of (94) is the same. We will use the same notation introduced in the proof of Proposition 4.1 and let $\mathcal{C}(u,t,k) = \sum_{j=0}^{k-1} \mathcal{A}(u,t,j)\, \underline{b}_{t-j}(u)$, that is $\mathcal{X}_t(u) = \mathcal{A}(u,t,k+j)\mathcal{X}_{t-k-j}(u) + \mathcal{C}(u,t,k+j)$.

We will use (91). Since $|f(\underline{x})|_{abs} \leq \mathbb{I}_{p+1}$ and $|g(\underline{x})|_{abs} \leq \mathbb{I}_{p+1}$ we have

$$\begin{aligned} & |(Z_{t-k+1}^2 - 1)\big(f(\mathcal{X}_{t-k,N})g(\mathcal{X}_t(u)) - f(\mathcal{C}_N(t-k,j))g(\mathcal{C}(u,t,k+j))\big)| \\ \leq \; & |Z_{t-k+1}^2 + 1|\big(|\mathcal{A}_N(t-k,j)\mathcal{X}_{t-k-j,N}|_1 + |\mathcal{A}(u,t,k+j)\mathcal{X}_{t-k-j}(u)|_1\big). \qquad (95) \end{aligned}$$

Since $\mathcal{A}(u,t,k-1)$, $(Z_{t-k}^2 + 1)A_{t-k-1}(u)$ and $\mathcal{A}(u,t-k,j)$ are independent matrices and by using (89) we have

$$\begin{aligned} \mathbb{E}\big\|[(Z_{t-k+1}^2 + 1)\mathcal{A}(u,t,k+j)]_1\big\| \leq \mathbb{E}\big\|[\mathcal{A}(u,t,k-1)]_1\big\| \times \\ \times \big\|[(Z_{t-k+1}^2 + 1)A_{t-k+1}(u)]_1\big\|_1 \big\|[\mathcal{A}(u,t-k,j)]_1\big\| \leq K\mathbb{E}(Z_{t-k+1}^4 + 1)\rho^{k+j-1}. \end{aligned} \qquad (96)$$

Considering the conditional expectation of (95), by using (96) and $|\mathbb{E}_{t-k-j}\{(Z_{t-k+1}^2 - 1)\mathcal{A}_N(t-k,j)\mathcal{X}_{t-k-j,N}\}| \leq K\rho^j |\mathcal{X}_{t-k-j,N}|_1$ we have

$$\begin{aligned} & |\mathbb{E}\big[(Z_{t-k+1}^2 - 1)\big\{f(\mathcal{X}_{t-k,N})g(\mathcal{X}_t(u)) - f(\mathcal{C}_N(t-k,j))g(\mathcal{C}(u,t,k+j)\big\}|\mathcal{F}_{t-k-1}\big]| \\ \leq \; & K\rho^j |\mathcal{X}_{t-k-j,N}|_1 \mathbb{E}(Z_{t-k+1}^2 + 1) + K\mathbb{E}(Z_{t-k+1}^4 + 1)\rho^{k+j-1}|\mathcal{X}_{t-k-j}(u)|_1 \\ \leq \; & K\mathbb{E}(Z_0^4 + 1)\rho^j \big(|\mathcal{X}_{t-k-j,N}|_1 + |\mathcal{X}_{t-k-j}(u)|_1\big). \end{aligned}$$

Using similar arguments we obtain the bound:

$$\begin{aligned} & \big(\mathbb{E}\,\big|\mathbb{E}\big[(Z_{t-k+1}^2 - 1)\big\{f(\mathcal{X}_{t-k,N})g(\mathcal{X}_t(u)) - f(\mathcal{C}_N(t-k,j))g(\mathcal{C}(u,t,k+j)\big\}\big]\big|^r\big)^{1/r} \\ \leq \; & K\mathbb{E}(Z_0^4 + 1)\rho^j \Big(\mathbb{E}(|\mathcal{X}_{t-k-j,N}|^r)^{1/r} + (\mathbb{E}|\mathcal{X}_{t-k-j}(u)|^r)^{1/r}\Big) \end{aligned}$$

leading to the result. □

A.2 Persistence of excitation

As we mentioned in Section 1, a crucial component in the analysis of many stochastic, recursive algorithms is to establish persistence of excitation of the transition random matrix in the algorithm: in our case this implies showing Theorem 4.1. Intuitively, it is clear that the verification of this result depends on showing that the smallest eigenvalue of the conditional expectation of of the semi-positive definite matrix $\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T/\|\mathcal{X}_{t,N}\|_1^2$ is bounded away from zero. In particular this is one of the major conditions given in Moulines et al. (2005), Theorem 16, where conditions were given under which persistence of excitation can be established. Their result is for a quite general class of algorithms which satisfy

$$\underline{Y}_t = \{I - \lambda A_t(\lambda)\}\underline{Y}_{t-1} + \underline{B}_t.$$

In our case the random matrix is not a function of λ, so to reduce notation we state the theorem below for the case where $A_t(\lambda) = A_t$. Let $\mathbb{M}_d^+$ be the space of all positive semi-definite symmetric matrices.

Theorem A.1 *(***Moulines, Priouret and Roueff***) Let $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F} : n \in \mathbb{N}\})$ be a filtered space. Let $\{\phi_t, t \geq 0\}$ and $\{V_t, t \geq 0\}$ be non-negative adaptive processes such that $V_t \geq 1$ for all $t \geq 1$. Let $A := \{A_t : t \geq 0\}$ be an adapted $\mathbb{M}_d^+$ valued process. Let s be a positive integer, and let R_1 and μ_1 be positive constants. Assume that*

(a) there exist $\mu < 1$ and $B \geq 1$ such that for all $t \in \mathbb{N}$,

$$\mathbb{E}(V_{t+s}|\mathcal{F}_t) \leq \mu V_t I(\phi_t > R_1) + B V_t I(\phi_t \leq R_1); \tag{97}$$

(b) there exists an $\alpha_1 > 0$ such that for all $t \in \mathbb{N}$;

$$\mathbb{E}\big\{\lambda_{min}\big(\sum_{l=t+1}^{t+s} A_l\big)|\mathcal{F}_t\big\} \geq \alpha_1 I(\phi_t \leq R_1); \tag{98}$$

(c) there exists a λ_1 such that for all $\lambda \in [0, \lambda_1]$ and $t \in \mathbb{N}$; $\|\lambda A_t\|_{lub,2} \leq 1$;

(d) there exist a $q > 1$ and $C_1 > 0$ such that for all $t \in \mathbb{N}$ and $\lambda \in [0, \lambda_1]$

$$I(\phi_t \leq R_1) \sum_{l=t+1}^{t+s} \mathbb{E}\big(\|A_l\|^q|\mathcal{F}_t\big) \leq C_1.$$

Then for any $p \geq 1$ there exist C_0, δ_0 and $\mu_0 > 0$ such that, for all $\lambda \in [0, \lambda_1]$ and $n \geq 1$, we have

$$\mathbb{E}\big(\|\prod_{i=1}^{n}(I - \lambda A_i)\|^p|\mathcal{F}_0\big) \leq C_0 e^{-\delta_0 \lambda n} V_0.$$

We use this theorem to prove Theorem 4.1, where $A_t := F_{t,N}$, $\phi_t := |\mathcal{X}_{t,N}|_1$ and $V_t := |\mathcal{X}_{t,N}|_1$. (97) is often referred to as the drift condition, and concerns the actual process $\{X_{t,N}\}$. Roughly speaking, under this condition if $|\mathcal{X}_{t,N}|_1$ is large, then after some time it should 'drift' back to the interval $[0, R_1]$. On the other hand (98), places conditions on the algorithm, in particular the random matrices $\{F_{t,N}\}_t$. This is when excitation occurs; loosely this means if $|\mathcal{X}_{t,N}|_1 \in [0, R_1]$ then $\| \prod_{i=0}^{s-1}(I - \lambda F_{t+i})\| \leq (1 - \delta\lambda)$. The analysis can then be conducted for $|\mathcal{X}_{t,N}|_1$ in the interval $[0, R_1]$, the drift condition then is used to extend the results over the whole real line.

Suppose X is a random variable and define $\mathbb{E}_t(X) = \mathbb{E}(X|\mathcal{F}_t)$.

Lemma A.1 *Suppose that Assumption 2.1 holds with $r = 2$. Then, we have*

$$\lambda_{min}\big(\mathbb{E}\{\mathcal{X}_t(u)\mathcal{X}_t(u)^T|\mathcal{F}_{t-k}\}\big) > C, \tag{99}$$

and for N large enough

$$\lambda_{min}\big(\mathbb{E}\{\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T|\mathcal{F}_{t-k}\}\big) > \frac{C}{2}, \tag{100}$$

for $k \geq (p+1)$, where C is a finite constant independent of t, N and u.

PROOF. We will prove (100). The proof of (99) is similar.

Our object is to partition $\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T$ as follows

$$\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T = \Delta_1 + \Delta_2, \tag{101}$$

where Δ_1 and Δ_2 are positive-definite matrices, $\Delta_1 \in \sigma(Z_t, \ldots, Z_{t-p})$ and $\Delta_2 \in \mathcal{F}_{t,N}$. This implies $\lambda_{min}\{\mathbb{E}_{t-k}(\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T)\} \geq \lambda_{min}\{\mathbb{E}(\Delta_1)\}$, if $k \geq p+1$, which allows us to obtain a uniform lower bound for $\lambda_{min}\{\mathbb{E}_{t-k}(\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T)\}$. To facilitate this we represent $\mathcal{X}_{t,N}$ in terms of martingale vectors.

By using (1) we have

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

and

$$\mathcal{X}_{t,N} = \Theta(\frac{t}{N})\mathcal{X}_{t-1,N} + (Z_t^2 - 1)\sigma_{t,N}^2 D \tag{102}$$

where D is a $(p+1)$-dimensional vector with $D_i = 0$ if $i \neq 2$ and $D_2 = 1$ and

$$\Theta(u) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ a_0(u) & a_1(u) & a_2(u) & \ldots & a_{p-1}(u) & a_p(u) \\ 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \\ 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix}.$$

Iterating (102) $(p+1)$-times we have

$$\mathcal{X}_{t,N} = \sum_{i=0}^{p}(Z_{t-i}^2 - 1)\sigma_{t-i,N}^2\{\prod_{j=0}^{i-1}\Theta(\frac{t-j}{N})\}D + \{\prod_{j=0}^{p}\Theta(\frac{t-j}{N})\}\mathcal{X}_{t-p-1,N}.$$

Let $\mu_4 = \mathbb{E}(Z_t^2 - 1)^2$, $\Theta_{t,N}(i) = \prod_{j=0}^{i-1}\Theta(\frac{t-j}{N})$ and $B_{t,N}(i) = \Theta_{t,N}(i)DD^T\Theta_{t,N}(i)^T$. Since $\{(Z_t^2 - 1)\sigma_{t,N}^2\}_t$ are martingale differences we have

$$\begin{aligned} \mathbb{E}(\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T|\mathcal{F}_{t-k,N}) &= \mu_4\sum_{i=0}^{p}\mathbb{E}_{t-k}(\sigma_{t-i,N}^4)\Theta_{t,N}(i)DD^T\Theta_{t,N}(i)^T + \\ &\quad +\Theta_{t,N}(p+1)\mathbb{E}_{t-k}(\mathcal{X}_{t-p-1,N}\mathcal{X}_{t-p-1,N}^T)\Theta_{t,N}(p+1)^T, \text{ for } k \geq p+1. \end{aligned} \tag{103}$$

Since the matrices above are non-negative definite, we have that

$$\lambda_{min}\{\mathbb{E}(\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T|\mathcal{F}_{t-k})\} \geq \lambda_{min}\{\mu_4\sum_{i=0}^{p}\mathbb{E}_{t-k}(\sigma_{t-i,N}^4)B_{t,N}(i)\}. \tag{104}$$

We now refine $\mu_4 \sum_{i=0}^{p} \mathbb{E}_{t-k}(\sigma_{t-i,N}^4) B_{t,N}(i)$ to obtain Δ_1. By using (88) we have

$$\mathcal{X}_{t,N} = \sum_{j=0}^{p} \mathcal{A}_N(t,j) \underline{b}_{t-j}(\frac{t-j}{N}) + \mathcal{A}_N(t,p+1)\mathcal{X}_{t-p-1,N}.$$

Therefore by using (1) and $X_{t-i,N}^2 = (\mathcal{X}_{t,N})_{i+1}$ we have

$$\begin{aligned} \sigma_{t-i,N}^2 &= \frac{1}{Z_{t-i}^2} \sum_{j=0}^{p} \{\mathcal{A}_N(t,j)\underline{b}_{t-j}(\frac{t-j}{N})\}_{i+1} + \frac{1}{Z_{t-i}^2} \{\mathcal{A}_N(t,p+1)\mathcal{X}_{t-p-1,N}\}_{i+1} \\ &= H_{t-i,N}(t) + G_{t-i,N}(t), \end{aligned} \tag{105}$$

where $H_{t-i,N}(t) \in \sigma(Z_{t-i}, \ldots, Z_{t-p})$ and $G_{t-i,N}(t) \in \mathcal{F}_{t-i}$. Since $H_{t-i,N}(t)$ and $G_{t-i,N}(t)$ are positive this implies with (104)

$$\lambda_{min}\{\mathbb{E}(\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T | \mathcal{F}_{t-k})\} \geq \lambda_{min}\{P_{t,N}\},$$

with

$$P_{t,N} = \mu_4 \sum_{i=0}^{p} \mathbb{E}(H_{t-i,N}(t)^2) B_{t,N}(i).$$

To bound this we define the corresponding terms for the stationary approximation $X_t(u)$. We set

$$P(u) = \mu_4 \sum_{i=0}^{p} \mathbb{E}(H_{t-i}(u,t)^2) B(u,i),$$

where

$$H_{t-i}(u,t) = \frac{1}{Z_{t-i}^2} \sum_{j=0}^{p} \{\mathcal{A}(u,t,j)\underline{b}_{t-j}(u)\}_{i+1}$$

and $B(u,i) = \Theta(u)^i DD^T (\Theta(u)^i)^T$. A close inspection of the above calculation steps reveals that

$$P(u) = \mathbb{E}\{\underline{Y}_p(u)\underline{Y}_p(u)^T\}$$

with $\underline{Y}_p(u)$ from Assumption 2.1(iv). Therefore,

$$\lambda_{min}\{P(\frac{t}{N})\} \geq \inf_u \lambda_{min}\{P(u)\} \geq C.$$

Since $\{a_j(\cdot)\}_j$ is β-Lipschitz we have

$$\begin{aligned} &|\mathbb{E}(H_{t-i}(\frac{t}{N},t)^2) - \mathbb{E}(H_{t-i,N}(t)^2)|_1 \leq \frac{K}{N^\beta} \\ &|B_{t,N}(i) - B(\frac{t}{N},i)|_1 \leq \frac{K}{N^\beta}, \text{ for } i = 0, \ldots, p-1. \end{aligned}$$

Therefore $\|P_{t,N} - P(\frac{t}{N})\| \leq |P_{t,N} - P(\frac{t}{N})|_1 \leq \frac{K}{N^\beta}$ which leads to

$$\lambda_{min}\{P_{t,N}\} = \lambda_{min}\{P(\frac{t}{N}) + [P_{t,N} - P(\frac{t}{N})]\} \geq \lambda_{min}\{P(\frac{t}{N})\} - \|P_{t,N} - P(\frac{t}{N})\| \geq C - \frac{K}{N^\beta}$$

and therefore to

$$\lambda_{min}\big(\mathbb{E}\{\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T|\mathcal{F}_{t-k,N}\}\big) > C - \frac{K}{N^\beta}. \tag{106}$$

Thus for N large enough we have (100). (99) follows in the same way. □

Lemma A.2 *Suppose that Assumption 2.1 holds with $r = 4$. Then, for $k \geq p+1$, we have for N large enough*

$$\lambda_{min}\big(\mathbb{E}\{\frac{\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T}{|\mathcal{X}_{t,N}|_1^2}|\mathcal{F}_{t-k}\}\big) > \frac{C}{|\mathcal{X}_{t-k,N}|_1^4} \tag{107}$$

and

$$\lambda_{min}\big(\mathbb{E}\{\frac{\mathcal{X}_t(u)\mathcal{X}_t(u)^T}{|\mathcal{X}_t(u)|_1^2}\}\big) > C, \tag{108}$$

where C is a constant independent of t, N and u.

PROOF. We now prove (107). By definition

$$\lambda_{min}\{\mathbb{E}_{t-k}(\frac{\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T}{|\mathcal{X}_{t,N}|_1^2})\} = \inf_{|\underline{x}|=1} \mathbb{E}_{t-k}\big(\frac{\underline{x}^T\mathcal{X}_{t,N}}{|\mathcal{X}_{t,N}|_1}\big)^2.$$

Since

$$(\underline{x}^T\mathcal{X}_{t,N})^2 = \frac{\underline{x}^T\mathcal{X}_{t,N}}{|\mathcal{X}_{t,N}|_1}\underline{x}^T\mathcal{X}_{t,N}|\mathcal{X}_{t,N}|_1$$

and $\sup_{|\underline{x}|=1} |\underline{x}^T\mathcal{X}_{t,N}|_1 \leq |\mathcal{X}_{t,N}|_1$, we obtain by using Cauchy's inequality and (29)

$$\begin{aligned}
\mathbb{E}_{t-k}(\underline{x}^T\mathcal{X}_{t,N})^2 &\leq \big\{\mathbb{E}_{t-k}(\frac{\underline{x}^T\mathcal{X}_{t,N}}{|\mathcal{X}_{t,N}|_1})^2\big\}^{1/2}\{\mathbb{E}_{t-k}(\underline{x}^T\mathcal{X}_{t,N}|\mathcal{X}_{t,N}|_1)^2\}^{1/2} \\
&\leq \big\{\mathbb{E}_{t-k}(\frac{\underline{x}^T\mathcal{X}_{t,N}}{|\mathcal{X}_{t,N}|_1})^2\big\}^{1/2}\{\mathbb{E}_{t-k}(|\mathcal{X}_{t,N}|_1^4)\}^{1/2} \\
&\leq \big\{\mathbb{E}_{t-k}(\frac{\underline{x}^T\mathcal{X}_{t,N}}{|\mathcal{X}_{t,N}|_1})^2\big\}^{1/2}\{K|\mathcal{X}_{t-k,N}|_1^4\}^{1/2}.
\end{aligned}$$

Therefore by using the above and Lemma A.1 for large N we obtain

$$\inf_{|\underline{x}|=1} \mathbb{E}_{t-k}\big(\frac{\underline{x}^T\mathcal{X}_{t,N}}{|\mathcal{X}_{t,N}|_1}\big)^2 \geq \inf_{|\underline{x}|=1} \frac{\big[\mathbb{E}_{t-k}(\underline{x}^T\mathcal{X}_{t,N})^2\big]^2}{\mathbb{E}_{t-k}(|\mathcal{X}_{t,N}|_1^4)} \geq \frac{C}{|\mathcal{X}_{t-k,N}|_1^4},$$

where C is a positive constant, thus giving (107). To prove (108) we use (99) to get

$$\mathbb{E}(\mathcal{X}_t(u)\mathcal{X}_t(u)^T) = \mathbb{E}_{-\infty}(\mathcal{X}_t(u)\mathcal{X}_t(u)^T) > C,$$

and using the arguments above we have

$$\lambda_{min}\big(\mathbb{E}\big\{\frac{\mathcal{X}_t(u)\mathcal{X}_t(u)^T}{|\mathcal{X}_t(u)|_1^2}\big\}\big) > \frac{C}{\mathbb{E}(|\mathcal{X}_{t-k}(u)|_1^4)}.$$

By Lemma 4.1 $\sup_u \mathbb{E}(|\mathcal{X}_t(u)|_1^4) < \infty$, which leads to (108). □

Corollary A.2 *Suppose Assumption 2.1 holds with $r = 4$. Let $F(u)$ be defined as in (10). Then there exists C and λ_1 such that for all $\lambda \in [0, \lambda_1]$ and $u \in (0, 1]$ we have*

$$\lambda_{max}\{I - \lambda F(u)\} \leq (1 - \lambda C). \tag{109}$$

There exists a $0 < \delta \leq C$ and K such that for all k

$$\|\{I - \lambda F(u)\}^k\| \leq K(1 - \lambda\delta)^k. \tag{110}$$

Furthermore

$$\sum_{k=0}^{t-1} \lambda\{I - \lambda F(u)\}^{2k} \to \frac{1}{2}F(u)^{-1} \tag{111}$$

where $\lambda \to 0$ and $\lambda t \to \infty$ as $t \to \infty$.

PROOF. (109) follows directly from (108). Furthermore, since $(I - \lambda F(u))$ is symmetric matrix, we have $\|\{I - \lambda F(u)\}^k\| \leq \|(I - \lambda F(u))\|^k \leq (1 - \lambda\delta)^k$, that is (110).

We now prove (111). We have

$$\lambda \sum_{k=0}^{t-1}\{I - \lambda F(u)\}^{2k} = \lambda\big\{I - (I - \lambda F(u))^2\big\}^{-1}\big\{I - (I - \lambda F(u))^{2t}\big\}.$$

Since $\lambda_{min}\{F(u)\} > C$, for some $C > 0$, we have $|\{I-\lambda F(u)\}^{2t}|_1 \leq \|\{I-\lambda F(u)\}^{2t}\||I_{p+1}| \to 0$ as $\lambda \to 0$, $\lambda t \to \infty$ and $t \to \infty$. Furthermore, $\lambda\big\{I-(I-\lambda F(u))^2\big\}^{-1} \to \frac{1}{2}F(u)^{-1}$. Together they give (111). □

Lemma A.3 *Suppose that Assumption 2.1 holds with $r = 4$. For a sufficiently large N and for every $R > 0$ there exists a $s_0 > p+1$ and C_1 such that, for all $s \geq s_0$ and $t = 1, \ldots, N-s$ we have*

$$\mathbb{E}\Big\{\lambda_{min}\Big(\sum_{k=t}^{t+s-1} \frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{|\mathcal{X}_{k,N}|_1^2}\Big|\mathcal{F}_t\Big)\Big\} \geq C_1 I(\|\mathcal{X}_{t,N}\|_1 \leq R),$$

where I denotes the identity function.

PROOF. The result can be proved using the methods given in Moulines et al. (2005), Lemma 19. We outline the proof. By using $\lambda_{min}(A) \geq \lambda_{min}(B) - \|A - B\|$ (see Moulines et al. (2005), Lemma 19) we have

$$\mathbb{E}_t\Big\{\lambda_{min}\Big(\sum_{k=t}^{t+s-1}\frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{|\mathcal{X}_{k,N}|_1^2}\Big)\Big\} \geq$$
$$\sum_{k=t}^{t+s-1}\lambda_{min}\{\mathbb{E}_t(\frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{|\mathcal{X}_{k,N}|_1^2})\} - \mathbb{E}_t\|\sum_{k=t}^{t+s-1}\Big\{\frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{|\mathcal{X}_{k,N}|_1^2} - \mathbb{E}_t(\frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{|\mathcal{X}_{k,N}|_1^2})\Big\}\|.$$

We now evaluate an upper bound for the second term above. Let $\Delta_{k,N} = \frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{\|\mathcal{X}_{k,N}\|_1^2} - \mathbb{E}_t(\frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{\|\mathcal{X}_{k,N}\|_1^2})$, then

$$\mathbb{E}_t\|\sum_{k=t}^{t+s-1}\Delta_{k,N}\|^2 \leq \mathbb{E}_t|\sum_{k=t}^{t+s-1}\Delta_{k,N}|^2 \leq \sum_{i,j=1}^{p+1}\sum_{k_1,k_2=t}^{t+s-1}\mathbb{E}_t((\Delta_{k_1,N})_{i,j}(\Delta_{k_2,N})_{i,j}). \quad (112)$$

We now bound $\mathbb{E}_t((\Delta_{k_1,N})_{i,j}(\Delta_{k_2,N})_{i,j})$. For $k_1 = k_2 = k$ we obtain with (29)

$$\mathbb{E}_t\big\{(\Delta_{k,N})_{i,j}\big\}^2 \leq 2\mathbb{E}_t|\mathcal{X}_{k,N}|_1^4 \leq K|\mathcal{X}_{t,N}|_1^4. \quad (113)$$

Now let $k_1 \neq k_2$. Let $\phi(\underline{x}) = \underline{x}^T\underline{x}/|\underline{x}|_1^2$, where where $\underline{x} = (1, x_1, \ldots, x_p)$. Since $\phi \in Lip(1)$, by using (28) we have for $k_1 < k_2$:

$$\begin{aligned} |\mathbb{E}_{k_1}((\Delta_{k_2,N})_{i,j})| &= |\mathbb{E}_{k_1}(\frac{(\mathcal{X}_{k_2,N}\mathcal{X}_{k_2,N}^T)_{i,j}}{|\mathcal{X}_{k_2,N}|_1^2}) - \mathbb{E}_t(\frac{(\mathcal{X}_{k_2,N}\mathcal{X}_{k_2,N}^T)_{i,j}}{|\mathcal{X}_{k_2,N}|_1^2})| \\ &\leq K\rho^{k_2-k_1}(\mathbb{E}_t(|\mathcal{X}_{k_1,N}|_1) + |\mathcal{X}_{k_1,N}|_1). \quad (114) \end{aligned}$$

By using (29), (113) and (114) we have

$$\begin{aligned} |\mathbb{E}_t\big\{(\Delta_{k_1,N})_{i,j}(\Delta_{k_2,N})_{i,j}\big\}| &\leq |\mathbb{E}_t\big\{\mathbb{E}_{k_1}(\Delta_{k_1,N})_{i,j}(\Delta_{k_2,N})_{i,j}\big\}| \\ &\leq |\mathbb{E}_t\big\{(\Delta_{k_1,N})_{i,j}\mathbb{E}_{k_1}(\Delta_{k_2,N})_{i,j}\big\}| \\ &\leq \{\mathbb{E}_t(\Delta_{k_1,N})_{i,j}^2\}^{1/2}\big\{\mathbb{E}_t(\mathbb{E}_{k_1}(\Delta_{k_2,N})_{i,j})^2\big\}^{1/2} \\ &\leq K\rho^{k_2-k_1}|\mathcal{X}_{t,N}|_1^3 \leq K\rho^{k_2-k_1}|\mathcal{X}_{t,N}|_1^4. \quad (115) \end{aligned}$$

Substituting (115) into (112) we obtain

$$\mathbb{E}_t\Big\|\sum_{k=t}^{t+s-1}\Delta_{k,N}\Big\|^2 = Ks|\mathcal{X}_{t,N}|_1^4, \quad (116)$$

where K is a constant independent of s. Therefore by using (107) and (116) we have for N sufficiently large

$$\mathbb{E}_t\Big\{\lambda_{min}\Big(\sum_{k=t}^{t+s-1}\frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{|\mathcal{X}_{k,N}|_1^2}\Big)\Big\} \geq \frac{Cs}{2}|\mathcal{X}_{t,N}|_1^{-4} - K\big\{s|\mathcal{X}_{t,N}|_1^4\big\}^{1/2}.$$

Therefore for any $R > 0$,

$$\mathbb{E}_t\Big\{\lambda_{min}\Big(\sum_{k=t}^{t+s-1}\frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{|\mathcal{X}_{k,N}|_1^2}\Big)\Big\} \geq \{\frac{Cs}{2R^4} - Ks^{1/2}R^2\}I(|\mathcal{X}_{t,N}|_1 \leq R)$$
$$= \frac{s^{1/2}}{R^4}\big(Cs^{1/2} - KR^2\big)I(|\mathcal{X}_{t,N}|_1 \leq R).$$

Now choose s_0 and corresponding C_1 such that

$$C_1 = \frac{s_0^{1/2}}{R^4}\big(Cs_0^{1/2} - KR^2\big) > 0.$$

Then it is clear if $s > s_0$ we have

$$\mathbb{E}_t\Big\{\lambda_{min}\Big(\sum_{k=t}^{t+s-1}\frac{\mathcal{X}_{k,N}\mathcal{X}_{k,N}^T}{|\mathcal{X}_{k,N}|_1^2}\Big)\Big\} \geq \frac{s^{1/2}}{R^4}\big(Cs^{1/2} - KR^2\big)I(|\mathcal{X}_{t,N}|_1 \leq R) \geq C_1 I(|\mathcal{X}_{t,N}|_1 \leq R),$$

thus we have the result. □

PROOF of Theorem 4.1 We prove the result by applying Theorem A.1 to $\{A_l = F_{l,N}, \mathcal{F}_l = \sigma(Z_l, Z_{l-1}, \ldots) : p+1 \leq l \leq N\}$ and verifying the conditions in Theorem A.1, Theorem 4.1 then immediately follows. Let $\phi_l := |\mathcal{X}_{l,N}|_1$, $V_l := |\mathcal{X}_{l,N}|_1$ and $A_l := F_{l,N} = \mathcal{X}_{l,N}\mathcal{X}_{l,N}^T/|\mathcal{X}_{l,N}|_1^2$. Let $k \geq p+1$ and $1 \leq s \leq N-k$. Then by using (27) we have $\mathbb{E}(V_{t+s}|\mathcal{F}_t) \leq K(1 + \rho^s|\mathcal{X}_{t,N}|_1)$. Therefore for any R_1 and $s > 0$ we have

$$\mathbb{E}(V_{t+s}\big|\mathcal{F}_t) \leq \{K\rho^s + \frac{K}{R_1}I(|\mathcal{X}_{t,N}|_1 > R_1)\}|\mathcal{X}_{t,N}|_1 + KI(|\mathcal{X}_{t,N}|_1 \leq R_1).$$

Thus we have

$$\mathbb{E}(V_{t+s}\big|\mathcal{F}_t) \leq (K\rho^s + \frac{K}{R_1})I(|\mathcal{X}_{t,N}|_1 > R_1)\}|\mathcal{X}_{t,N}|_1 + K(\rho^s+1)I(|\mathcal{X}_{t,N}|_1 \leq R_1)|\mathcal{X}_{t,N}|_1.$$

By choosing an appropriate R_1 we can find a s_1 such that for all $s \geq s_1$ we have $K\rho^s + K/R_1 < 1$ and thus condition (a) is satisfied. Condition (b) directly follows from Lemma A.3. Let $\mathbb{I}_{p+1}$ be a $(p+1)\times(p+1)$ matrix where $(\mathbb{I}_{p+1})_{ij} = 1$ for $1 \leq i,j \leq (p+1)$. Since $\|F_{l,N}\| = \frac{\mathcal{X}_{l,N}^T\mathcal{X}_{l,N}}{|\mathcal{X}_{l,N}|_1^2} = 1$, for $\lambda < 1$, we have $\lambda\|F_{l,N}\| < 1$, hence condition (c) is satisfied. Finally by using the above for any $q \geq 1$ and $t \in \{N-p, \ldots, t\}$ we have

$$\sum_{l=t}^{r+s_1}\mathbb{E}(\|F_{l,N}\|^q|\mathcal{F}_j) \leq k_1(p+1)^{2q}.$$

Therefore condition (d) is satisfied and Theorem 4.1 follows from Theorem A.1. □

A.3 The lower order terms in the pertubation expansion

In this section we will prove the auxillary results required in Section 4.2, where we showed that the second order terms in the pertubation expansion were a lower order than the principle terms. The analysis of $J_{t_0,N}^{x,2}$ is based on partitioning it into two terms

$$J_{t_0,N}^{x,2} = A_{t_0,N}^x + B_{t_0,N}^x,$$

similar to the partition in (54).

$A^x_{t_0,N}$ is the weighted sum of $\{F_{k,N} - F_k(u)\}$ whereas $B^x_{t_0,N}$ is the weighted sum of the differences between the stationary approximation $F_k(u_0)$ and $\mathbb{E}(F_0(u_0))$, that is of $\bar{F}_k(u_0) = F_k(u_0) - F(u_0)$. In this section we evaluate bounds for these two terms. We use the following lemma.

Lemma A.4 *Suppose Assumption 2.1 holds with some $r \geq 1$. Let $\mathcal{M}_{t,N}$ and $\mathcal{M}_t(u)$ be defined as in (30) and let $D_{t,N}$ be defined as in (46). Then we have*

$$\frac{\mathcal{X}_{t,N}\mathcal{X}_{t,N}^T}{|\mathcal{X}_{t,N}|_1^2} = \frac{\mathcal{X}_t(u)\mathcal{X}_t(u)^T}{|\mathcal{X}_t(u)|_1^2} + R_{t,N}(u) \tag{117}$$

$$\text{and } \mathcal{M}_{t,N} = \mathcal{M}_t(u) + (1 + Z_t^2)R_{t,N}(u)$$

where

$$|R_{t,N}(u)| \leq \big(|\frac{t}{N} - u|^\beta + (\frac{p}{N})^\beta + \frac{1}{N^\beta}\big)D_{t,N}$$

Further, for $q \leq q_0$ there exists a constant K independent of t, N and u such that

$$(\mathbb{E}|R_{t,N}(u)|_q)^{1/q} \leq K(|\frac{t}{N} - u|^\beta + (\frac{p+1}{N})^\beta). \tag{118}$$

PROOF. The proof uses (22) and the method given in Dahlhaus and Subba Rao (2006), Lemma A.4. □

We now give a bound for a general $A^x_{t,N}$.

Lemma A.5 *Suppose Assumption 2.1 holds with $r = q_0$ and let $\{G_{t,N}\}$ be a random process which satisfies $\sup_{t,N} \|G_{t,N}\|^E_{q_0} < \infty$. Let $F_{t,N}$, $F_t(u)$ and $F(u)$ be defined as in (31) and (10), respectively. Then if $|\frac{t_0}{N} - u_0| < 1/N$ and $q \leq q_0$ we have*

$$\mathbb{E}|\sum_{k=0}^{t_0-p-1}(I - \lambda F(u_0))^k\big(F_{t_0-k-1,N} - F_{t_0-k-1}(u_0)\big)G_{t_0-k,N}|^{q/2})^{2/q} \leq \frac{K}{(\delta N\lambda)^\beta}\sup_{t,N}(\mathbb{E}|G_{t,N}|^q)^{1/q}, \tag{119}$$

where K is a finite constant.

PROOF. By using (110) and (117) we get the result. □

We now consider the second term $B^x_{t_0,N}$. A crude way to bound $\mathbb{E}|B^x_{t_0,N}|^q$ is to use Minkowski's inequality and directly take the norm into the summand. However, with this method we do not obtain the desired lower order. To overcome this problem we use Burkhölder type inequalities.

We now state a generalisation of Proposition B.3, in Moulines et al. (2005). The following result can be proved by adapting the proof of Proposition B.3, in Moulines et al. (2005).

Lemma A.6 *Suppose $\{M_t\}$ and $\{F_t\}$ are random matrices and F is a positive definite, deterministic matrix, with $\lambda_{min}(F) > \delta$, for some $\delta > 0$. Let $\mathcal{F}_t = \sigma(F_t, M_t, F_{t-1}, M_{t-1}, \ldots)$.*

We assume for some $q \geq 2$

(i) *$\{F_t\}$ are identically distributed with mean zero.*

(ii) *$\mathbb{E}(M_t|\mathcal{F}_{t-1}) = 0$.*

(iii) *$\mathbb{E}|\mathbb{E}(F_t|\mathcal{F}_k)|^{2q})^{1/2q} \leq K\rho^{t-k}$*

(iv) *$\mathbb{E}|\mathbb{E}(M_tF_s|\mathcal{F}_k) - \mathbb{E}(M_tF_s)|^q)^{1/q} \leq K\rho^{s-k}$ if $k \leq s \leq t$.*

(v) *$\sup_t(\mathbb{E}|M_t|^{2q})^{1/(2q)} < \infty$ and $(\mathbb{E}|F_t|^{2q})^{1/(2q)} < \infty$.*

Then we have

$$\left(\mathbb{E}|\sum_{k=0}^{t-p-1}\sum_{i=0}^{t-p-k-2}(I-\lambda F)^{k+i}F_{t-k-1}M_{t-k-i-1}|^q\right)^{1/q} \leq \frac{K}{\delta\lambda}, \tag{120}$$

where K is a finite constant.

PROOF. We note since $\lambda_{min}\{F\} \geq \delta > 0$, then

$$\|(I-\lambda F)^k\| \leq K(1-\lambda\delta)^k. \tag{121}$$

Change of variables in (120) gives

$$\sum_{k=0}^{t-p-1}\sum_{i=0}^{t-p-k-2}(I-\lambda F)^{k+i}F_{t-k-1}M_{t-k-i-1} = \sum_{k=p}^{t-1}\sum_{i=p+1}^{k}(I-\lambda F)^{t-i-1}F_kM_i$$
$$= \sum_{k=p}^{t-1}(I-\lambda F)^{t-k-1}F_k\sum_{i=p+1}^{k}(I-\lambda F)^{k-i}M_i$$

(using the convention $\sum_{i=1}^{0} = 0$). Let $\phi_k = (I-\lambda F)^{t-k-1}F_k\sum_{i=1}^{k}(I-\lambda F)^{k-i}M_i$. Then we have

$$\left(\mathbb{E}|\sum_{k=p}^{t-1}\sum_{i=p+1}^{k}(I-\lambda F)^{t-i-1}F_kM_i|^q\right)^{1/q} \leq K\left(\mathbb{E}|\sum_{k=p}^{t-1}\mathbb{E}(\phi_k)| + \{\mathbb{E}|\sum_{k=p}^{t-1}\{\phi_k - \mathbb{E}(\phi_k)|^q\}^{1/q}\right) \tag{122}$$

We first consider $|\sum_{k=p}^{t-1}\mathbb{E}(\phi_k)|$. Since $\mathbb{E}(F_kM_i) = \mathbb{E}(M_i\mathbb{E}(F_k|\mathcal{F}_i))$, under the stated assumptions, and using the norm inequality, $|Ax| \leq \|A\|\,|x|$, we obtain

$$|\mathbb{E}((I-\lambda F)^{t-i-1}F_kM_i)| \leq K(1-\lambda\delta)^{t-i-1}\mathbb{E}\big(|M_i|\,|\mathbb{E}(F_k|\mathcal{F}_i)|\big) \leq K\rho^{k-i}\mathbb{E}|M_i| \leq K\rho^{k-i}.$$

Substituting the above into $\mathbb{E}(\phi_k)$ gives the bound

$$|\mathbb{E}(\phi_k)| \leq K(1-\lambda\delta)^{t-k-1}\sum_{i=p+1}^{k}\rho^{k-i} \leq KS(\rho)(1-\lambda\delta)^{t-k-1},$$

where $S(\rho) = 1/(1-\rho)$. Thus

$$|\sum_{k=p}^{t-1} \mathbb{E}(\phi_k)| \leq K \sum_{k=p}^{t-1} (1-\lambda\delta)^{t-k-1} S(\rho) = O(\lambda^{-1}). \tag{123}$$

We now consider the second term on the right hand side of (122), $(\mathbb{E}|\sum_{k=p}^{t-1}\{\phi_k - \mathbb{E}(\phi_k)\}|^q)^{1/q}$. Let $\bar{\phi_k} = \{\phi_k - \mathbb{E}(\phi_k)\}$. By using the generalised Burkhölder inequality (see Dedecker and Doukhan (2003)) we have

$$(\mathbb{E}|\sum_{k=p}^{t-1} \bar{\phi_k}|^q)^{1/q} \leq \left\{2q \sum_{k=p}^{t-1} (\mathbb{E}|\bar{\phi}_k|_q)^{1/q} \sum_{j=k}^{t-1} (\mathbb{E}|\mathbb{E}(\bar{\phi}_j|\mathcal{F}_k)|^q)^{1/q}\right\}^{1/2}. \tag{124}$$

We treat the two summands, on the right hand side of (124), separately. By using Hölder's inequality, the norm inequality and (121) we have

$$\begin{aligned}
(\mathbb{E}|\bar{\phi}_k|^q)^{1/q} &= \left(\mathbb{E}|(I-\lambda F)^{t-k-1} \sum_{i=p+1}^{k} (I-\lambda F)^{k-i}\{M_i F_k - \mathbb{E}(M_i F_k)\}|^q\right)^{1/q} \\
&\leq 2K(1-\lambda\delta)^{t-k-1} (\mathbb{E}|\sum_{i=p+1}^{k} (1-\lambda\delta)^{k-i} M_i F_k|^q)^{1/q} \\
&\leq 2K(1-\lambda\delta)^{t-k-1} (\mathbb{E}|F_k|^{2q})^{1/(2q)} (\mathbb{E}|\sum_{i=p+1}^{k} (1-\lambda\delta)^{k-i} M_i|^{2q})^{1/(2q)}. \quad (125)
\end{aligned}$$

By applying Burkhölder's inequality to $(\mathbb{E}|\sum_{i=p+1}^{k}(1-\lambda\delta)^{k-i}M_i|^{2q})^{1/(2q)}$, we have

$$(\mathbb{E}|\sum_{i=p+1}^{k} (1-\lambda\delta)^{k-i} M_i|^{2q})^{1/(2q)} \leq \left\{2q \sum_{i=p+1}^{k} (1-\lambda\delta)^{2(k-i)} ((\mathbb{E}|M_i|^{2q})^{1/q})\right\}^{1/2}.$$

Substituting the above into (125) gives

$$\begin{aligned}
(\mathbb{E}|\bar{\phi}_k|^q)^{1/q} &\leq 2K(1-\lambda\delta)^{t-k-1} \sup_k (\mathbb{E}|F_k|^{2q})^{1/(2q)} \sup_i (\mathbb{E}|M_i|^{2q})^{1/2q} \{2q \sum_{i=0}^{k-p-1} (1-\lambda\delta)^{2i}\}^{1/2} \\
&\leq \frac{K}{\lambda^{1/2}} (1-\lambda\delta)^{t-k-1}. \quad (126)
\end{aligned}$$

We now consider $\sum_{j=k}^{t-1}(\mathbb{E}|\mathbb{E}(\bar{\phi}_j|\mathcal{F}_k)|^q)^{1/q}$. We treat the cases $M_i \in \mathcal{F}_k$ and $M_i \notin \mathcal{F}_k$

separately, and obtain the following partition

$$\begin{aligned}
\sum_{j=k}^{t-1} \|\mathbb{E}(\bar{\phi}_j|\mathcal{F}_k)\|_q^E &\leq \sum_{j=k}^{t-1}\left(\mathbb{E}|(I-\lambda F)^{t-j-1}\sum_{i=p+1}^{j}(I-\lambda F)^{j-i}\mathbb{E}\{[M_iF_j-\mathbb{E}(M_iF_j)]|\mathcal{F}_k\}|^q\right)^{1/q} \\
&\leq \sum_{j=k}^{t-1}\left(\mathbb{E}|(I-\lambda F)^{t-j-1}\sum_{i=p+1}^{k}(I-\lambda F)^{j-i}\mathbb{E}\{[M_iF_j-\mathbb{E}(M_iF_j)]|\mathcal{F}_k\}|^q\right)^{1/q} + \\
&\quad + \sum_{j=k}^{t-1}\left(\mathbb{E}|(I-\lambda F)^{t-j-1}\sum_{i=k+1}^{j}(I-\lambda F)^{j-i}\mathbb{E}\{[M_iF_j-\mathbb{E}(M_iF_j)]|\mathcal{F}_k\}|^q\right)^{1/q} \\
&= A_k + B_k. \qquad (127)
\end{aligned}$$

Since for all random variables A, $|\mathbb{E}(A)| \leq (\mathbb{E}|\mathbb{E}(A|\mathcal{F}_k|^q)^{1/q}$ (if $q \geq 1$) and $\mathbb{E}(M_iF_j|\mathcal{F}_k) = M_i\mathbb{E}(F_j|\mathcal{F}_k)$ (for $i \leq k$), we have

$$\begin{aligned}
A_k &\leq 2\sum_{j=k}^{t-1}\|(I-\lambda F)^{t-j-1}\mathbb{E}\{\sum_{i=p+1}^{k}(I-\lambda F)^{j-i}M_iF_j|\mathcal{F}_k\}\|_q^E \\
&\leq 2K\sum_{j=k}^{t-1}(1-\lambda\delta)^{t-j-1}\left(\mathbb{E}|\mathbb{E}(F_j|\mathcal{F}_k)\sum_{i=p+1}^{k}(I-\lambda F)^{j-i}M_i|^q\right)^{1/q}.
\end{aligned}$$

Now by using the Hölder inequality and $(\mathbb{E}|\mathbb{E}(F_j|\mathcal{F}_k)|^{2q})^{1/(2q)} \leq K\rho^{j-k}$ we have

$$\begin{aligned}
A_k &\leq 2\sum_{j=k}^{t-1}(1-\lambda\delta)^{t-j-1}(\mathbb{E}|\mathbb{E}(F_j|\mathcal{F}_k)|^{2q})^{1/2q}\,(\mathbb{E}|\sum_{i=p+1}^{k}(I-\lambda F)^{j-i}M_i|^{2q})^{1/2q} \\
&\leq 2\sum_{j=k}^{t-1}K\rho^{j-k}\{2q\sum_{i=p+1}^{k}(1-\lambda\delta)^{j-i}(\mathbb{E}|M_i|^{2q})^{1/q})\}^{1/2} \leq \frac{K}{\lambda^{1/2}}. \qquad (128)
\end{aligned}$$

We now bound B_k:

$$B_k \leq K\sum_{j=k}^{t-1}(1-\lambda\delta)^{t-j-1}\sum_{i=k+1}^{j}(1-\lambda\delta)^{j-i}(\mathbb{E}|\mathbb{E}\{[M_iF_j-\mathbb{E}(M_iF_j)]|\mathcal{F}_k\}|^q)^{1/q}. \quad (129)$$

We will evaluate two different bounds for $\|\mathbb{E}\{(M_iF_j-\mathbb{E}(M_iF_j))|\mathcal{F}_k\}\|_q^E$. We use the minimum of these bounds to evaluate B_k. Since $i > k$, $(\mathbb{E}|\mathbb{E}(A|\mathcal{F}_k)\|_q^E \leq (\mathbb{E}|\mathbb{E}(A|\mathcal{F}_i)|^q)^{1/q}$, under the assumption $(\mathbb{E}|\mathbb{E}(F_j|\mathcal{F}_i)|^{2q})^{1/2q} \leq K\rho^{j-i}$, we have

$$\begin{aligned}
(\mathbb{E}|\mathbb{E}\{[M_iF_j-\mathbb{E}(M_iF_j)]|\mathcal{F}_k\}|^q)^{1/q} &\leq (\mathbb{E}|\mathbb{E}\{[M_iF_j-\mathbb{E}(M_iF_j)]|\mathcal{F}_i\}|^q)^{1/q} \\
&\leq 2(\mathbb{E}|\mathbb{E}\{M_iF_j|\mathcal{F}_i\}|^q)^{1/q} \leq 2(\mathbb{E}|M_i\mathbb{E}\{F_j|\mathcal{F}_i\}|^q)^{1/q} \\
&\leq 2(\mathbb{E}|M_i|^{2q})^{1/2q}|\mathbb{E}\{F_j|\mathcal{F}_i\}|^{2q})^{1/(2q)} \leq K\sup_i(\mathbb{E}|M_i|^{2q})^{1/(2q)}\rho^{j-i}.
\end{aligned}$$

Using again the stated assumption (iv), we get the alternative bound

$$(\mathbb{E}|\mathbb{E}\{[M_iF_j-\mathbb{E}(M_iF_j)]|\mathcal{F}_k\}|^q)^{1/q} \leq K\rho^{i-k}.$$

Therefore we have

$$(\mathbb{E}|\mathbb{E}\{(M_iF_j - \mathbb{E}(M_iF_j))|\mathcal{F}_k\}|^q)^{1/q} \leq K\min(\rho^{j-i}, \rho^{i-k})$$

leading to

$$\begin{aligned} B_k &\leq K\sum_{j=k}^{t-1}(1-\lambda\delta)^{t-j-1}\sum_{i=k+1}^{j}(1-\lambda\delta)^{j-i}\min(\rho^{j-i}, \rho^{i-k}) & (130)\\ &\leq \sum_{j=k}^{t-1}\sum_{i=k+1}^{j}\min(\rho^{j-i}, \rho^{i-k}) \leq K. & (131)\end{aligned}$$

Substituting (128) and (130) into (127) gives

$$\sum_{j=k}^{t-p-1}(\mathbb{E}|\mathbb{E}(\bar{\phi}_j|\mathcal{F}_k)|^q)^{1/q} \leq A_k + B_k \leq K + \frac{K}{\lambda^{1/2}}, \tag{132}$$

which, by substituting (126) and (132) into (124), leads to

$$\mathbb{E}|\sum_{k=p}^{t-1}\bar{\phi}_k|^q)^{1/q} \leq \{2q\sum_{k=0}^{t-1}(1-\lambda\delta)^{t-k-1}\frac{K}{\lambda^{1/2}}\big(K + \frac{K}{\lambda^{1/2}}\big)\}^{1/2} \leq \frac{K}{\lambda}. \tag{133}$$

Finally by using (122), (123) and (133) we get the assertion. □

We now apply the lemma above to the tvARCH online recursive algorithm.

Lemma A.7 *Suppose Assumption 2.1 holds for $r > 4$. Let $\{F_t(u)\}$, $\{\mathcal{M}_{t,N}\}$ and $\{\mathcal{M}_t(u)\}$ be defined as in (30) and $\bar{F}_t(u) = F_t(u) - \mathbb{E}(F_t(u))$. Then, we have*

$$\left(\mathbb{E}\sum_{k=0}^{t_0-p-1}\sum_{i=0}^{t_0-p-k-2}\{I-\lambda F(u)\}^{k+i}\bar{F}_{t_0-k-1}(u)\mathcal{M}_{t_0-k-i-1,N}|^{r/2}\right)^{2/r} \leq \frac{K}{\lambda} \tag{134}$$

and

$$\left(\mathbb{E}|\sum_{k=0}^{t_0-p-1}\sum_{i=0}^{t_0-p-k-2}\{I-\lambda F(u)\}^{k+i}\bar{F}_{t_0-k-1}(u)\mathcal{M}_{t_0-k-i-1}(u)|^{r/2}\right)^{2/r} \leq \frac{K}{\lambda}. \tag{135}$$

PROOF. We prove (134), the proof of (135) is the same. We will prove the result by verifying the conditions in Lemma A.6, then (134) immediately follows. By using (108) we have that $\lambda_{min}\{F(u)\} \geq \delta$, for some $\delta > 0$. Let $M_t := \mathcal{M}_{t,N}$, $F_t := \bar{F}_t(u)$, $F := F(u)$ and $\mathcal{F}_t = \sigma(Z_t, Z_{t-1}, \ldots)$. It is clear from the definition that the series $\{\bar{F}_t(u)\}_t$ has zero mean and are identically distributed, also $\mathbb{E}(\mathcal{M}_t(u)|\mathcal{F}_{t-1}) = 0_{p+1\times p+1}$. By using (25) we have

$$\big(\mathbb{E}|\mathbb{E}(\bar{F}_t(u)|\mathcal{F}_k)|^r\big)^{1/r} = (\mathbb{E}|\mathbb{E}(F_t(u)|\mathcal{F}_k) - \mathbb{E}(F_t(u))|^r)^{1/r} \leq K\rho^{t-k},$$

thus condition (iii) is satisfied. Since $\mathcal{M}_{t,N} = (Z_t^2 - 1)\sigma_{t,N}^2 \mathcal{X}_{t-1,N}/|\mathcal{X}_{t-1,N}|_1^2$, and $F_t \leq \mathbb{I}_{p+1}$ and $\sigma_{t,N}^2 \mathcal{X}_{t-1,N}/|\mathcal{X}_{t-1,N}|_1^2 \leq \mathbb{I}_{p+1}$, by using Corollary A.1 and $\sup_{t,N}(\mathbb{E}|\mathcal{X}_{t,N}|^{r/2})^{2/r} < \infty$, we can show that condition (iv) is satisfied.

Moreover, $F_{t,N}$ is a bounded random matrix, hence all its moments exist. Finally, since $\sup_{t,N} |\mathcal{M}_{t,N}|_r^r \leq K(Z_t^2 + 1)^r$ and $\mathbb{E}(Z_0^{2r}) < \infty$ we have for all $k \leq s \leq t \leq N$, that $\sup_{t,N}(\mathbb{E}|\mathcal{M}_{t,N}|^r) < \infty$, leading to condition (v). Thus all the conditions of Lemma A.6 are satisfied and we obtain (134). □

References

Aguech, R., Moulines, E., & Priouret, P. (2000). On a perturbation approach for the analysis of stochastic tracking algorithm. *SIAM Journal on Control and Optimization*, *39*, 872-99.

Chen, X., & White, H. (1998). Nonparametric learning with feedback. *Journal of Economic Theory*, *82*, 190-222.

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

Gill, R., & Levit, B. (1995). Applications of the van Trees inequality: a Bayesian Cramér-Rao bound. *Bernoulli*, *1/2*, 59-79.

Guo, L. (1994). Stability of recursive stochastic tracking algorithms. *SIAM J. on Control and Optimisation*, *32*, 1195-1225.

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

Kushner, H., & Yin, G. (2003). *Stochastic approximation and recursive algorithms and applications.* Heidelberg: Springer.

Ljung, L., & Söderström, T. (1983). *Theory and practice of recursive identification.* Cambridge, MA: MIT Press.

Mikosch, T., & Stărică, C. (2000). Is it really long memory we see in financial returns. In P. Embrechts (Ed.), *Extremes and integrated risk management* (p. 439-459). London: Risk Books.

Mikosch, T., & Stărică, C. (2004). Non-stationarities in financial time sries, the long-range dependence, and the IGARCH effects. *The Review of Econometrics and Statistics*, *86*, 378-390.

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

Solo, V. (1981). The second order properties of a time series recursion. *Ann. Statist.*, *9*, 307-317.

Solo, V. (1989). The limiting behavior of the LMS. *IEEE Trans. Acoust. Speach and Signal Proceess*, *37*, 1909-1922.

Subba Rao, S. (2004). On some nonstationary, nonlinear random processes and their stationary approximations. *Preprint*.

White, H. (1996). Parametric statistical estimation using artifical neural networks. In P. Smolensky, M. Mozer, & D. Rumelhart (Eds.), *Mathematical perspectives on neural networks* (p. 719-775). Hilldale, NJ: L. Erlbaum Associates.