A note on uniform convergence of an ARCH(∞) estimator

Suhasini Subba Rao
Department of Statistics
Texas A&M University
College Station, Texas, TX-77843, USA

December 12, 2006

Abstract

We consider parameter estimation for a class of ARCH(∞) models, which do not necessarily have a parametric form. The estimation is based on a normalised least squares approach, where the normalisation is the weighted sum of past observations. The number of parameters estimated depends on the sample size and increases as the sample size grows. Using maximal inequalities for martingales and mixingales we derive a uniform rate of convergence for the parameter estimator. We show that the rate of convergence depends both on the number of parameters estimated and the rate that the ARCH(∞) parameters tend to zero.

1 Introduction

Consider the process $\{X_t\}_t$ satisfying the ARCH(∞) representation

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

where $\{Z_t\}_t$ are independent, identically distributed random variables with $\mathbb{E}(Z_t) = 0$ and $\mathbb{E}(Z_t^2) = 1$. Giraitis, Kokoskza, and Leipus (2000) have shown if $\sum_{j=1}^{\infty} a_j < 1$, then (1) has a unique stationary causal solution that can be written as a Volterra series expansion, thus by using Stout (1974), Theorem 3.5.8, $\{X_t\}$ is an ergodic process. The ARCH(∞) process was considered in Robinson (1991) and more recently in Giraitis et al. (2000), Giraitis and Robinson (2001) and Koulikov (2003). A classical example of a process which satisfies (1) are the class of GARCH(p, q) models. For a review of recent advances in ARCH modelling see, for example, Fan and Yao (2003) and Giraitis, Leipus, and Surgailis (2005).

Suppose we observe $\{X_t : t = 0, 1, \ldots, n\}$, our object is to estimate the parameters $\{a_j\}$. In the estimator we propose, we suppose the number of parameters to be estimated depends on n, and let the number grow as $n \to \infty$. The estimator is based on the following AR(∞) representation of (1)

$$X_t^2 = a_0 + \sum_{k=1}^{\infty} a_j X_{t-k}^2 + (Z_t^2 - 1)\sigma_t^2, \tag{2}$$

where $\{(Z_t^2 - 1)\sigma_t^2\}_t$ are martingale differences. To estimate the parameters, we first use a self-weighting normalisation of (2). That is we divide (2) throughout by $g_t = g + \sum_{j=1}^{\infty} \ell(j) X_{t-j}^2$ and use

$$\frac{X_t^2}{g_t} = \frac{a_0}{g_t} + \sum_{j=1}^{\infty} a_j \frac{X_{t-j}^2}{g_t} + \frac{(Z_t^2 - 1)\sigma_t^2}{g_t} \tag{3}$$

to estimate the parameters, where g is a known constant and $\ell(j) \to 0$, as $j \to \infty$. We observe that by dividing by g_t we are attempting to mimic the unknown conditional variance σ_t^2. A similar normalisation was used in Fryzlewicz et al. (2006) to estimate the parameters of time-varying ARCH(p) processes, where $p < \infty$ and in Ling (2006) to estimate the parameters of an ARMA-IGARCH model. We note that one could use the AR representation (2) and the Yule-Walker method described in An, Chen, and Hannan (1982) to estimate the parameters, however such an estimator would require the existence of eight moments of the marginal distribution of $\{X_t\}$, which for the ARCH process places a severe restriction on the parameters $\{a_j\}$. In other words if $\{Z_t\}$ were Gaussian we require that $\sum_{j=1}^{\infty} a_j < 1/\{\mathbb{E}(Z_t^8)\}^{1/4}$, which means the parameters would have be close to zero for the condition to be satisfied. It is worth mentioning that Bose and Mukherjee (2003) used a two-stage least squares method to estimate the parameters of a ARCH(p) processes, where $p < \infty$. Further, Robinson and Zaffaroni (2006) consider a particular class of ARCH(∞) processes, whose parameters can be represented as $a_j = \psi_j(\zeta)$, where ζ is a finite set of unknown parameters which are estimated using the Quasi-Maximum likelihood, and $\{\psi_j(\cdot)\}$ are known functions (for example a GARCH(p, q) process with known p, q satisfies these conditions).

Suppose we restrict the parameters to be estimated to $\underline{a}_{p_n} = (a_0, \ldots, a_{p_n})^T$, then the least squares estimator of $\underline{a}_{p_n}$ using (3) is $\underline{\hat{a}}_{p_n}$, where

$$\underline{\hat{a}}_{p_n} = \hat{R}_{n,p_n}^{-1} \underline{\hat{r}}_{n,p_n} \tag{4}$$

and

$$\begin{aligned} \underline{\hat{r}}_{n,p_n} &= \frac{1}{n} \sum_{t=p_n}^{n} \frac{X_t^2}{g + \sum_{k=1}^{t} \ell(k) X_{t-k}^2} \underline{X}_{p_n,t-1} \\ \hat{R}_{n,p_n} &= \frac{1}{n} \sum_{t=p_n}^{n} \frac{1}{(g + \sum_{k=1}^{t} \ell(k) X_{t-k}^2)^2} \underline{X}_{p_n,t-1} \underline{X}_{p_n,t-1}^T, \end{aligned} \tag{5}$$

with $\underline{X}_{p_n,t-1} = (1, X_{t-1}^2, \ldots, X_{t-p_n}^2)^T$ and g is known. Suitable values of g have been suggested in Fryzlewicz et al. (2006). For example, if the estimator is to be insensitive to the magnitude of a_0 (as in the quasi-maximum likelihood estimator of the ARCH parameters) it is suggested in Fryzlewicz et al. (2006) to let g be the unconditional variance of the process or an estimator of it.

To study the sampling properties of estimators which involve at some stage using the ARCH(∞) estimator $\underline{\hat{a}}_{p_n}$, it maybe necessary to obtain a uniform rate of convergence for $\underline{\hat{a}}_{p_n}$. For example, given $\underline{\hat{a}}_{p_n}$ and $\{X_t; t = 1, \ldots, n\}$ we can use $\hat{Z}_t$ as an estimator of the residuals Z_t where

$$\hat{Z}_t = \frac{X_t}{\sqrt{\hat{a}_{p_n,0} + \sum_{j=1}^{p_n} \hat{a}_{p_n,j} X_{t-j}^2}}, \quad \text{with} \quad \underline{\hat{a}}_{p_n} = (\hat{a}_{p_n,0}, \ldots, \hat{a}_{p_n,p_n})^T.$$

We would require a uniform rate of convergence for the estimator $\underline{\hat{a}}_{p_n}$ to determine how close $\hat{Z}_t$ is to the true Z_t.

The aim of this paper is to obtain a uniform, almost sure rate of convergence for the vector $\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}$. In Section 3 we partition the estimator $\underline{\hat{a}}_{p_n}$ into a term which is the sum of martingale differences and an additional term which arises because only a finite number of the ARCH(∞) parameters are estimated. We evaluate uniform bounds for each of these terms.

Because of the rather complicated nature of the summands in (5), in particular the random weighting and that $\{X_t\}$ is a nonlinear process, the methods developed in An et al. (1982) cannot be directly applied to our situation. Instead, in the Appendix we show that the elements in the sum (5) are L_2-Near Epoch Dependent (L_2-NED). We use this property to write the sums in (5) as doubly infinite sums of martingale differences, which we use to obtain a maximal inequality for $\hat{R}_{n,p_n}$ and $\underline{\hat{r}}_{n,p_n}$. This is used to prove the results in the main section. The methods and results in the Appendix may also be of independent interest.

2 Assumptions and notation

We denote $tr(A)$, $\lambda_{min}(A)$ and $\lambda_{max}(A)$ as the trace, the smallest and largest eigenvalues of the matrix A respectively. Let

$$R_m = \mathbb{E}\left(\frac{\underline{X}_{m,0}\underline{X}_{m,0}^T}{(g + \sum_{k=1}^{\infty} \ell(k) X_{-k}^2)^2}\right). \tag{6}$$

We assume throughout, that $\ell(j) = Kj^{-(3+\delta)}$ for some $\delta > 0$ and finite K. We note that we require this as a minimum rate of convergence in order to obtain the required L_2-NED rate of convergence (see Theorem A.1 in the Appendix). Using a slower rate of $\ell(j)$ would mean we would not obtain the desired L_2-NED rate, whereas a faster rate for $\ell(j)$ would require a faster rate of convergence for the parameters $\{a_j\}$ (see Assumption 2.1(ii) below).

Assumption 2.1 *Suppose the process $\{X_t\}_t$ satisfies the ARCH(∞) representation in (1), and have* *(i)* $\sum_{j=1}^{\infty} a_j < 1$.

(ii) *For some monotonically decreasing sequence* $\{c_j\}$ *with* $c_j \to 0$ *as* $j \to \infty$, *we have* $\sum_{j=1}^{\infty} \frac{a_j}{\ell(j) c_j} = K \sum_{j=1}^{\infty} j^{3+\delta} a_j c_j^{-1} < \infty$.

(iii) $\mathbb{E}(Z_t^4) < \infty$.

(iv) *The process* $\{X_t\}_t$ *is strongly mixing with rate* $-a$ *where* $a > 1$.

(v) *Let* $\{R_m\}$ *be defined as in (6). Then the smallest eigenvalues matrices* $\{R_m\}$ *is uniformly bounded away from zero, that is for some* $c > 0$, $\inf_m \lambda_{\min}(R_m) \geq c$.

Assumptions 2.1(i,iii) are standard in ARCH estimation. In fact Assumption 2.1(i) implies that $\mathbb{E}(X_t^2) < \infty$. It can be shown that Assumption 2.1(iv) is true for stationary ARCH processes and certain GARCH processes. Assumption 2.1(ii), which is only required in Lemma 3.1, imposes a condition on the rate of decay of the parameter $\{a_j\}$. It basically means that $\frac{a_j}{\ell(j)} \leq K j^{-(1+\eta)}$ (thus $a_j \leq K j^{-(4+\delta+\eta)}$), for some $\eta > 0$. We note that the Yule-Walker method described in An et al. (1982) would require $\sum_{j=1}^{\infty} a_j \leq 1/\{\mathbb{E}(Z_t^8)\}^{1/4}$ and $\sum_{j=1}^{\infty} j a_j < \infty$. Comparing these assumptions with those in Assumption 2.1(ii) we see that $\sum_{j=1}^{\infty} j a_j < \infty$ is weaker than $\sum_{j=1}^{\infty} a_j/(\ell(j) c_j) < \infty$, since the former implies $a_j = O(j^{-2-\eta})$ and the latter implies $a_j = o(j^{-4-\eta})$, for some $\eta > 0$. However the Yule-Walker method would require the additional assumption $\sum_{j=1}^{\infty} a_j \leq 1/\{\mathbb{E}(Z_t^8)\}^{1/4}$, which places a restriction on the magnitude of the coefficients, whereas Assumption 2.1(ii) is a restriction on the rate they converge to zero. Furthermore, we show in Remark 3.1 if $\mathbb{E}(X_t^4) < \infty$, then Assumption 2.1(ii) can be relaxed to $\sum_{j=1}^{\infty} j a_j < \infty$. In other words, by using the normalisation in the estimation we can obtain the same results as the Yule-Walker method with the same parameter restrictions but weaker moment assumptions.

In order to derive the results below we need to define the the vector and matrix $\underline{\tilde{r}}_{n,p_n}$ and $\tilde{R}_{n,p_n}$, respectively. Let

$$\begin{aligned} \underline{\tilde{r}}_{n,p_n} &= \frac{1}{n} \sum_{t=p_n}^{n} \frac{X_t^2}{(g + \sum_{k=1}^{\infty} \ell(k) X_{t-k}^2)^2} \underline{X}_{p_n,t-1} \\ \text{and} \quad \tilde{R}_{n,p_n} &= \frac{1}{n} \sum_{t=p_n}^{n} \frac{1}{(g + \sum_{k=1}^{\infty} \ell(k) X_{t-k}^2)^2} \underline{X}_{p_n,t-1} \underline{X}_{p_n,t-1}^T. \end{aligned} \qquad (7)$$

We note that $\underline{\tilde{r}}_{n,p_n}$ and $\tilde{R}_{n,p_n}$ are similar to $\underline{\hat{r}}_{n,p_n}$ and $\hat{R}_{n,p_n}$ except that the normalisation goes to the infinite past. Furthermore we will require the following definitions

$$\mathcal{H}_n(i,j) = \frac{1}{n} \sum_{t=1}^{n} \left\{ \frac{X_{t-i}^2 X_{t-j}^2}{(g + \sum_{k=1}^{t} \ell(k) X_{t-k}^2)^2} - \frac{X_{t-i}^2 X_{t-j}^2}{(g + \sum_{k=1}^{\infty} \ell(k) X_{t-k}^2)^2} \right\} \qquad (8)$$

$$c_t(i,j) = \frac{X_{t-i}^2 X_{t-j}^2}{(g + \sum_{k=1}^{\infty} \ell(k) X_{t-k}^2)^2} \qquad (9)$$

Let $\|\cdot\|_r$ denote the ℓ_r-norm of a vector of matrix (where $\|\cdot\|_\infty$ denotes the sup norm) and assume throughout that K is a finite constant.

3 A uniform rate of convergence for the parameter estimators

In order to derive a bound for $\|\hat{\underline{a}}_{p_n} - \underline{a}_{p_n}\|_1$ (which immediately gives $\|\hat{\underline{a}}_{p_n} - \underline{a}_{p_n}\|_\infty$) we partition $(\hat{\underline{a}}_{p_n} - \underline{a}_{p_n})$ into a term with zero mean and an additional term due to our estimating only a subset of the coefficients $\{a_j\}$. By using the definition of $\hat{\underline{a}}_{p_n}$ given in (4) we have

$$\hat{\underline{a}}_{p_n} - \underline{a}_{p_n} = \hat{R}_{n,p_n}^{-1}\left(\hat{\underline{r}}_{n,p_n} - \hat{R}_{n,p_n}\underline{a}_{p_n}\right) = R_{p_n}^{-1}\left(\tilde{\underline{r}}_{n,p_n} - \tilde{R}_{n,p_n}\underline{a}_{p_n}\right) + \mathcal{C}_{p_n} \tag{10}$$

where

$$\mathcal{C}_{p_n} = \hat{R}_{n,p_n}^{-1}\left(\{\hat{\underline{r}}_{n,p_n} - \tilde{\underline{r}}_{n,p_n}\} + \{\hat{R}_{n,p_n} - \tilde{R}_{n,p_n}\}\underline{a}_{p_n}\right) + (\hat{R}_{n,p_n}^{-1} - R_{p_n}^{-1})\left(\hat{\underline{r}}_{n,p_n} - \hat{R}_{n,p_n}\underline{a}_{p_n}\right).$$

In the following section we find bounds for the terms on the right hand side of (10). More precisely, we obtain bounds for $\hat{\underline{r}}_{n,p_n} - \hat{R}_{n,p_n}\underline{a}_{p_n}$, $\hat{R}_{n,p_n} - R_{p_n}$ and $\hat{\underline{r}}_{n,p_n} - \tilde{\underline{r}}_{n,p_n}$.

3.1 A rate for $(\tilde{\underline{r}}_{n,p_n} - \tilde{R}_{n,p_n}\underline{a}_{p_n})$

Let us consider the first term in (10). It is straightforward to show that

$$\tilde{\underline{r}}_{n,p_n} - \tilde{R}_{n,p_n}\underline{a}_{p_n} = \frac{1}{n}\sum_{t=p_n}^{n}\frac{X_t^2 - \sum_{j=1}^{p_n} a_j X_{t-j}^2}{(g + \sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2}\underline{X}_{p_n,t-1}. \tag{11}$$

It follows, if the process were an ARCH(p) process of finite order, where $p < p_n$, then $\sum_{j=0}^{p_n} a_j X_{t-j}^2 = (Z_t^2 - 1)\sigma_t^2$ and $(\tilde{\underline{r}}_{n,p_n} - \tilde{R}_{n,p_n}\underline{a}_{p_n})$ would be the sum of martingale differences. However for the general ARCH(∞) model this is not the case. We can consider the parameters $\{a_j : j > p_n\}$, omitted in the estimation as the contribution to the bias of the estimator. Therefore a larger p_n would lead to a smaller bias. However a larger p_n means that more parameters are being estimated leading to a larger variance. We note that this is similar to the bias-variance trade off problem which often arises in nonparametric statistics. We now make the discussion above precise. Since $X_t^2 - \sum_{j=1}^{p_n} a_j X_{t-j}^2 = \sum_{j=p_n+1}^{\infty} a_j X_{t-j}^2 + (Z_t^2 - 1)\sigma_t^2$, and substituting this into (11) we obtain

$$\tilde{\underline{r}}_{n,p_n} - \tilde{R}_{n,p_n}\underline{a}_{p_n} = \mathcal{A}_{n,p_n} + \mathcal{B}_{n,p_n} \tag{12}$$

where

$$\mathcal{A}_{n,p_n} = \frac{1}{n}\sum_{t=p_n}^{\infty}\frac{(Z_t^2 - 1)\sigma_t^2}{(g + \sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2}\underline{X}_{p_n,t-1} \tag{13}$$

$$\text{and}\quad \mathcal{B}_{n,p_n} = \frac{1}{n}\sum_{t=p_n}^{n}\frac{\sum_{j=p_n+1}^{\infty} a_j X_{t-j}^2}{(g + \sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2}\underline{X}_{p_n,t-1}. \tag{14}$$

We first study the term $\mathcal{B}_{n,p_n}$.

Lemma 3.1 *Suppose Assumption 2.1(i,ii) holds and let $\mathcal{B}_{n,p_n}$ be defined as in (14), then we have*

$$\|\mathcal{B}_{n,p_n}\|_\infty \le c_{p_n} g^{-1}\left(\frac{1}{n}\sum_{t=1}^{n} Y_t\right) \qquad \textit{where} \qquad Y_t = \sum_{j=1}^{\infty} \frac{a_j}{\ell(j)c_j} X_{t-j}^2. \tag{15}$$

Furthermore, almost surely we have $\frac{1}{n}\sum_{t=1}^{n} Y_t \to \mathbb{E}(X_0^2)\sum_{j=1}^{\infty}\frac{a_j}{\ell(j)c_j}$ as $n \to \infty$.

PROOF. By using the normalisation we are able to reduce some of the high moment assumptions that would normally arise for quadratic forms. We note that for $1 \le i \le p_n$, we obtain the bound $\|\underline{X}_{p_n,t}\|_\infty/(g + \sum_{k=1}^{\infty}\ell(k)X_{t-k}^2) \le \frac{1}{\ell(p_n)}$. Using this we have

$$\|\mathcal{B}_{n,p_n}\|_\infty \le \frac{1}{ng}\sum_{t=p_n}^{n}\frac{1}{\ell(p_n)}\sum_{j=p_n}^{\infty} a_j X_{t-j}^2 = \frac{1}{ng}\sum_{t=p_n}^{n} c_{p_n}\sum_{j=1}^{\infty}\frac{a_{j+p_n}}{\ell(p_n)c_{p_n}} X_{t-j-p_n}^2.$$

Under Assumption 2.1(ii) we have $\sum_{j=1}^{\infty}\frac{a_j}{\ell(j)c_j} < \infty$ and we note that monotonically $\ell(j)c_j \to 0$ as $j \to \infty$, this gives

$$\|\mathcal{B}_{n,p_n}\|_\infty \le \frac{c_{p_n}}{ng}\sum_{t=p_n}^{n}\sum_{j=1}^{\infty}\frac{a_{j+p_n}}{\ell(j+p_n)c_{j+p_n}} X_{t-j-p_n}^2 \le \frac{c_{p_n}}{ng}\sum_{t=p_n}^{n}\sum_{j=1}^{\infty}\frac{a_j}{\ell(j)c_j} X_{t-j}^2.$$

thus we obtain (15). Since $\{X_t\}_t$ is an ergodic process (see Giraitis et al. (2000)) and $\sum_{j=1}^{\infty}\frac{a_j}{\ell(j)c_j} < \infty$, by using Stout (1974), Theorem 3.4.5, we have $\{Y_t\}_t$ is also an ergodic process. Finally since $\mathbb{E}(X_t^2) < \infty$, we have $\frac{1}{n}\sum_{t=1}^{n} Y_t \to \mathbb{E}(X_0^2)\sum_{j=1}^{\infty}\frac{a_j}{\ell(j)c_j}$. □

It is clear to see from Lemma 3.1, that $\|\mathcal{B}_{n,p_n}\|_\infty \stackrel{\text{a.s.}}{\to} 0$ as $n \to \infty$.

Remark 3.1 *If we are able to assume that the fourth moment of the ARCH process exists, we can relax the rate of decay required on the ARCH coefficients $\{a_j\}$. To be precise, let us suppose there exists a decreasing sequence $\{d_n\}$, where $d_n \to 0$ such that $\sum_{j=1}^{\infty}\frac{a_j}{d_j} < \infty$ and $n\,d_n \to \infty$. Since $\{X_t\}_t$ is an ergodic sequence and $\sum_{j=1}^{\infty}\frac{a_j}{d_j} < \infty$, then $\{\sum_{j=1}^{\infty}\frac{a_j}{c_j}X_{t-j}^2X_{t-i}^2\}_t$ is an ergodic sequence. Now because the fourth moment of X_t exists, by using similar a derivation to those used to prove Lemma 3.1 we have*

$$\|\mathcal{B}_{n,p_n}\|_1 \le d_{p_n}\sum_{i=1}^{p_n}\frac{1}{n}\sum_{t=1}^{n}\sum_{j=1}^{\infty}\frac{a_j}{d_j}X_{t-j}^2X_{t-i}^2 \stackrel{a.s.}{\to} d_{p_n}\sum_{i=1}^{p_n}\sum_{j=1}^{\infty}\frac{a_j}{d_j}\mathbb{E}(X_{t-j}^2X_{t-i}^2).$$

Therefore since $d_{p_n} \to 0$ as $p_n \to 0$, the right hand side of the above converges to zero if $d_n n \to 0$ as $n \to \infty$. We note that since we require $\sum_{j=1}^{\infty}\frac{a_j}{d_j} < \infty$ and $d_n n \to 0$, this means that a_j should converge at the rate $a_j \le K\frac{1}{(j(\log j)^{1+\kappa})^2}$. We note under these conditions the rates in Theorem 3.1 change. In particular, the c_{p_n} in (29) is replaced with d_{p_n}.

It is worth noting that the Yule-Walker method proposed in An et al. (1982) is for ARMA processes, thus the parameters in the AR representation decay exponentially and $\sum_{j=1}^{\infty} j a_j < \infty$.

The condition $\sum_{j=1}^{\infty} ja_j < \infty$ is almost equivalent to the condition given here where $\sum_{j=1}^{\infty} \frac{a_j}{d_j} < \infty$ and $d_n n \to 0$ as $n \to \infty$. Therefore under the same rate of convergence of the parameters, by using the normalised least squares estimator we only require the existence of $\mathbb{E}(X_t^4)$, rather than $\mathbb{E}(X_t^8) < \infty$ which is required in the Yule-Walker method.

Having obtained a bound for the bias, we now consider the remaining term $\mathcal{A}_{n,p_n}$ in (38). The analysis involves using some maximal inequalities. To exploit them, we first observe that the summands $\hat{R}_{n,p_n}$, $\tilde{R}_{n,p_n}$, $\underline{\hat{r}}_{n,p_n}$ and $\underline{\tilde{r}}_{n,p_n}$ are defined in the interval $[p_n, n]$. Typically maximal inequalities are derived for sums which have a fixed lower sum bound. Hence we replace the lower limit of the summand from p_n to $\max(i,j)$, since $\max(i,j) < p_n$ we have

$$\frac{1}{n}\sum_{t=p_n}^{n} \frac{X_{t-i}^2 X_{t-j}^2}{(g+\sum_{k=1}^{t}\ell(k)X_{t-k}^2)^2} \leq \frac{1}{n}\sum_{t=\max(i,j)}^{n} \frac{X_{t-i}^2 X_{t-j}^2}{(g+\sum_{k=1}^{t}\ell(k)X_{t-k}^2)^2}. \tag{16}$$

By definition of $\mathcal{A}_{n,p_n}$ and using the above it is straightforward to see $\|\mathcal{A}_{n,p_n}\|_\infty \leq n^{-1}\sup_{1\leq r\leq p_n}\mathcal{S}_{r,n}$, where

$$\mathcal{S}_{r,n} = \sum_{t=r}^{n} \frac{(Z_t^2-1)\sigma_t^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2} X_{t-r}^2. \tag{17}$$

For $\eta > 0$, define

$$q_{\eta,n} = \frac{(\log n(\log\log n)^{(1+\eta)})^{1/2}}{\ell(p_n)n^{1/2}}. \tag{18}$$

We now bound $\|\mathcal{A}_{n,p_n}\|_\infty$.

Lemma 3.2 *Suppose Assumption 2.1(i,iii) holds. Let $\mathcal{S}_{r,n}$ and $q_{\eta,n}$ be defined as in (17) and (18). Then for any $\eta > 0$ we have almost surely*

$$\|\mathcal{A}_{n,p_n}\|_\infty = O\left(p_n^{1/2} q_{\eta,n} + n^{-1}p_n\right). \tag{19}$$

PROOF. We recall that $\|\mathcal{A}_{n,p_n}\|_\infty \leq \sup_{1\leq r\leq p_n}\mathcal{S}_{r,n}$. By using Lemma A.2 we have a bound for $\mathbb{E}(\sup_{r\leq t\leq n}\mathcal{S}_{r,t}^2)$. Using this result and applying Lemma A.1 we obtain the result. □

By using (15) and (19) we almost surely obtain

$$\left\|\underline{\tilde{r}}_{n,p_n} - \tilde{R}_{n,p_n}\underline{a}_{p_n}\right\|_\infty = O\left(p_n^{1/2} q_{\eta,n} + c_{p_n}\right). \tag{20}$$

This gives a bound for the first part of (10). Below we study and bound the second term $\mathcal{C}_{p_n}$.

3.2 A rate for $\mathcal{C}_{p_n}$

Recalling the definition of $\mathcal{C}_{p_n}$ in (10), we now bound $\hat{R}_{n,p_n} - \tilde{R}_{n,p_n}$, $\underline{\hat{r}}_{n,p_n} - \underline{\tilde{r}}_{n,p_n}$ and $\hat{R}_{n,p_n}$.

Since the elements $(\hat{R}_{n,p_n} - \tilde{R}_{n,p_n})_{i+1,j+1} = \mathcal{H}_n(i,j)$ and $(\underline{\hat{r}}_{n,p_n} - \underline{\tilde{r}}_{n,p_n})_j = \mathcal{H}_n(0,j)$ (where $\mathcal{H}_n(i,j)$ is defined in (8)), in the following lemma we obtain a bound for $\mathcal{H}_n(i,j)$.

Lemma 3.3 *Suppose Assumption 2.1(i,ii) holds. Let $\mathcal{H}_{i,j}(n)$ be defined as in (8). If $1 \leq i,j \leq p_n$, then we have*

$$|\mathcal{H}_n(i,j)| \leq K\frac{p_n^{5+3\delta/2}}{n}V_\delta \quad \textit{and} \quad |\mathcal{H}_n(0,j)| \leq K\frac{p_n^{2+\delta}}{gn}W_\delta V_\delta \tag{21}$$

where K is a finite constant and

$$V_\delta = \left(\sum_{t=1}^{\infty}\frac{1}{t^{1+\delta/2}}X_t^2\right) \quad W_\delta = \left(\sum_{j=1}^{\infty}\frac{1}{j^{1+\delta/2}}X_{-j}^2\right).$$

PROOF. We first observe that the elements in the sum of $\mathcal{H}_n(i,j)$ can be bounded by

$$\left|\frac{X_{t-i}^2X_{t-j}^2}{(g+\sum_{k=1}^{t}\ell(k)X_{t-k}^2)^2} - \frac{X_{t-i}^2X_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2}\right| \leq \frac{2X_{t-i}^2X_{t-j}^2}{(g+\sum_{k=1}^{t}\ell(k)X_{t-k}^2)^2}\sum_{k=t+1}^{\infty}\ell(k)X_{t-k}^2. \tag{22}$$

Furthermore if $1 \leq i,j \leq p_n$, then the above can be further bounded by $\sum_{k=t+1}^{\infty}\ell(k)X_{t-k}^2 \leq \frac{2\sum_{k=t+1}^{\infty}\ell(k)X_{t-k}}{\ell(p_n)^2}$. Now by using this and that $\{\ell(k)\}$ is a monotonically decreasing sequence ($\ell(k) = Kk^{-(3+\delta)}$) we have

$$\begin{aligned}
|\mathcal{H}_n(i,j)| &\leq \frac{K}{n}\sum_{t=p_n}^{n}\sum_{k=t+1}^{\infty}\frac{\ell(k)}{\ell(p_n)^2}X_{t-k}^2 \leq \frac{K}{n\ell(p_n)^2}\sum_{j=1}^{n-p_n}\sum_{k=1}^{\infty}\ell(p_n+j+k)X_{-j-k}^2 \\
&\leq \frac{Kp_n^{2(3+\delta)}}{n}\sum_{j=1}^{n-p_n}\frac{1}{(p_n+j)^{2+\delta/2}}\sum_{k=1}^{\infty}\frac{1}{k^{(1+\delta/2)}}X_{-k}^2 \leq \frac{Kp_n^{2(3+\delta)-(1+\delta/2)}}{n}\sum_{k=1}^{\infty}k^{-(1+\delta/2)}X_{-k}^2
\end{aligned}$$

which gives the first inequality (21). To prove the second inequality in (21) we use similar arguments to above. By using (22) we have

$$\left|\frac{X_t^2X_{t-j}^2}{(g+\sum_{k=1}^{t}\ell(k)X_{t-k}^2)^2} - \frac{X_t^2X_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2}\right| \leq \frac{2X_t^2}{g\ell(p_n)}\sum_{k=t+1}^{\infty}\ell(k)X_{t-k}^2.$$

This leads to

$$\begin{aligned}
|\mathcal{H}_n(0,j)| &\leq \frac{Kg^{-1}}{n\ell(p_n)}\sum_{t=p_n}^{n}X_t^2\sum_{k=t+1}^{\infty}\ell(k)X_{t-k}^2 = \frac{Kg^{-1}}{n\ell(p_n)}\sum_{t=p_n}^{n}X_t^2\sum_{j=1}^{\infty}\ell(t+j)X_{-j}^2 \\
&\leq \frac{Kg^{-1}p_n^{3+\delta}}{n}\sum_{t=p_n}^{n}\frac{1}{t^{2+\delta/2}}X_t^2\sum_{j=1}^{\infty}\frac{1}{j^{1+\delta/2}}X_{-j}^2 \leq \frac{Kg^{-1}p_n^{2+\delta}}{n}\left(\sum_{t=1}^{\infty}\frac{1}{t^{1+\delta/2}}X_t^2\right)\left(\sum_{j=1}^{\infty}\frac{1}{j^{1+\delta/2}}X_{-j}^2\right).
\end{aligned}$$

and the desired result. □

We notice that the bound for $\mathcal{H}_n(i,j)$ does not depend on i or j when $1 \leq i,j \leq p_n$, this allows us to obtain uniform bounds for $\hat{R}_{n,p_n} - \tilde{R}_{n,p_n}$ and $\underline{\hat{r}}_{n,p_n} - \underline{\tilde{r}}_{n,p_n}$.

Finally, we require a uniform rate of convergence for the elements of $\hat{R}_{n,p_n}$. We will also use this result to bound $(\hat{R}_{n,p_n}^{-1} - R_{p_n}^{-1})$. Let

$$\mathcal{S}_n(i,j) = \sum_{t=\max(i,j)}^{n} c_t(i,j) = \sum_{t=\max(i,j)}^{n} \frac{X_{t-i}^2 X_{t-j}^2}{(g + \sum_{k=1}^{\infty} \ell(k) X_{t-k}^2)^2} \tag{23}$$

Lemma 3.4 *Suppose Assumption 2.1(i,iv) holds. Let $c_n(i,j)$ be defined as in (28). Then almost surely for large enough n we have*

$$\|\tilde{R}_{n,p_n} - R_{p_n}\|_\infty \leq \sup_{1\leq i,j\leq p_n} |\frac{1}{n} \sum_{k=\max(i,j)}^{n} \{c_n(i,j) - \mathbb{E}[c_n(i,j)]\}| = O(p_n q_{\eta,n}). \tag{24}$$

PROOF. We first note that by using Lemma A.5 we have

$$\left\{ \mathbb{E}\left(\max_{1\leq t\leq n} \frac{1}{n} \sum_{k=\max(i,j)}^{n} \{c_n(i,j) - \mathbb{E}[c_n(i,j)]\} \right)^2 \right\}^{1/2} \leq \frac{2K}{\ell(\max(i,j))^2} n^{1/2}. \tag{25}$$

To obtain an almost sure rate of convergence for $S_n(i,j)$ we use the result in Lemma A.1 together with (25). As a consequence of this we immediately obtain the result. □

Combining Lemmas 3.3 and 3.4 we have $\|\hat{R}_{n,p_n} - R_{p_n}\|_1 \leq \|\tilde{R}_{n,p_n} - \hat{R}_{n,p_n}\|_1 + \|\tilde{R}_{n,p_n} - R_{p_n}\|_1 \leq p_n^2(p_n q_{\eta,n} + n^{-1}p^{5+3\delta/2})$. The corollary below follows from Lemma 3.4.

Corollary 3.1 *Suppose Assumption 2.1(i,iv,v) holds. Let $\hat{R}_{n,p_n}$ and $p_{\eta,n}$ be defined as in (7) and (18) respectively. Then if $p_n^2(q_{\eta,n} + n^{-1}p^{5+3\delta/2}) \to 0$ we have almost surely*

$$\left\| \left(\hat{R}_{n,p_n}^{-1} - R_{p_n}^{-1} \right) \right\|_1 = O\left(p_n^2 (p_n q_{\eta,n} + n^{-1} p_n^{5+3\delta/2}) \right). \tag{26}$$

PROOF. We mention that the techniques used in the proof are similar to those in An et al. (1982), Theorem 4. We observe that we can write $\hat{R}_{n,p_n}^{-1} - R_{p_n}^{-1} = \hat{R}_{n,p_n}^{-1}\left(\hat{R}_{n,p_n} - R_{p_n}\right) R_{p_n}^{-1}$. We will now bound $\hat{R}_{n,p_n}^{-1}$. By using (24) and (21) we have $\lambda_{max}(|\hat{R}_{n,p_n} - R_{p_n}|) \leq tr((\hat{R}_{n,p_n} - R_{p_n})^2) = O(p_n^2(p_n q_{\eta,n} + n^{-1}p_n^{5+3\delta/2}))$. Furthermore, since $\hat{R}_{n,p_n} = R_{p_n} + (\hat{R}_{n,p_n} - R_{p_n})$, it is clear that $\lambda_{min}(\hat{R}_{n,p_n}) \geq \lambda_{min}(R_{p_n}) - \lambda_{max}(\hat{R}_{n,p_n} - R_{p_n})$. Using this and under Assumption 2.1(vi) we have that for a large enough n almost surely, the smallest eigenvalue of $\hat{R}_{n,p_n}$ is bounded away from zero. Therefore

$$\left\| \hat{R}_{n,p_n}^{-1} \left(\hat{R}_{n,p_n} - R_{p_n} \right) R_{p_n}^{-1} \right\|_1 = O(p_n^2(p_n q_{\eta,n} + n^{-1}p^{5+3\delta/2})), \tag{27}$$

and thus we obtain (26). □

Rewriting $\mathcal{C}_{p_n}$, using (20), Lemma 3.4 and Corollary 3.1, if $p_n^2(p_n q_{\eta,n} + n^{-1}p^{5+3\delta/2}) \to 0$ as $n \to \infty$, then

$$\begin{aligned}
\|\mathcal{C}_{p_n}\|_1 &= \left\|R_{p_n}^{-1}\left(\{\underline{\hat{r}}_{n,p_n} - \underline{\tilde{r}}_{n,p_n}\} + \{\hat{R}_{n,p_n} - \tilde{R}_{n,p_n}\}\underline{a}_{p_n}\right)\right\|_1 + \left\|(\hat{R}_{n,p_n}^{-1} - R_{p_n}^{-1})\left(\underline{\hat{r}}_{n,p_n} - \hat{R}_{n,p_n}\underline{a}_{p_n}\right)\right\|_1 + \\
&\quad \left\|(R_{p_n}^{-1} - \hat{R}_{n,p_n}^{-1})\{\underline{\hat{r}}_{n,p_n} - \underline{\tilde{r}}_{n,p_n}\}\right\|_1 + \left\|(R_{p_n}^{-1} - \hat{R}_{n,p_n}^{1})\{\hat{R}_{n,p_n} - \tilde{R}_{n,p_n}\}\underline{a}_{p_n}\right\|_1 \\
&\leq O\left(p_n\left(\frac{p_n^{2+\delta}}{gn} + \frac{p_n^{5+3\delta/2}}{n}\right)\right) + o(p_n^{1/2}q_{\eta,n} + c_{p_n}). \qquad (28)
\end{aligned}$$

3.3 The main result: A uniform rate for $(\underline{\hat{a}}_{p_n} - \underline{a}_{p_n})$

By using (20), Corollary 3.1 and Lemma 3.3 we have the necessary ingredients to bound $(\underline{\hat{a}}_{p_n} - \underline{a}_{p_n})$.

Theorem 3.1 *Suppose Assumption 2.1 holds. Let $q_{\eta,n}$ be defined as in (18). If $p_n^2(p_n q_{\eta,n} + n^{-1}p_n^{5+3\delta/2}) \to 0$, then we have almost surely*

$$\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_1 = O\left(n^{-1}p_n^{6+3\delta/2} + p_n^{3/2}q_{\eta,n} + p_n c_{p_n}\right). \qquad (29)$$

PROOF. We use (10) to prove the result. We first observe that

$$\left\|R_{p_n}^{-1}(\underline{\tilde{r}}_{n,p_n} - \tilde{R}_{n,p_n}\underline{a}_{p_n})\right\|_1 = O\left(p_n(p_n^{1/2}q_{\eta,n} + c_{p_n})\right). \qquad (30)$$

By using (28) we observe if $p_n^2(p_n q_{\eta,n} + n^{-1}p^{6+3\delta/2}) \to 0$ as $n \to \infty$, then we have $\|\mathcal{C}_{p_n}\|_1 = O(\frac{p_n^{2+\delta}}{gn} + \frac{p_n^{6+3\delta/2}}{n}) + o(p_n^{1/2}q_{\eta,n} + c_{p_n})$. Substituting this and (30) into (10), we obtain the required rate. □

Since $\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_\infty \leq \|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_1$, the rate for $\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_1$ immediately gives a bound for $\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_\infty$, thus a uniform rate of convergence for the parameter estimators. Furthermore, it is likely that $\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_\infty = O(\frac{1}{p_n}\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_1)$ (though we have not proved this).

Remark 3.2 *By using (29) we see that if $p_n^2(p_n q_{\eta,n} + n^{-1}p_n^{5+3\delta/2}) \to 0$ and $nc_n \to 0$ as $n \to \infty$, then $\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_1 \to 0$ almost surely. We observe that the rate of convergence is mainly influenced by the choice of p_n and the rate that $a_j \to \infty$ as $j \to \infty$.*

Example: The GARCH process

$\{X_t\}$ is called a GARCH process if it satisfies the representation

$$X_t = \sigma_t Z_t \qquad \sigma_t^2 = \alpha_0 + \sum_{i=1}^{p}\alpha_i X_{t-i}^2 + \sum_{j=1}^{q}\beta_j\sigma_{t-j}^2.$$

If $\sum_{i=1}^{p} \alpha_i + \sum_{j=1}^{p} \beta_j < 1$, then the GARCH process belongs to the ARCH(∞) class, where the parameters decay geometrically (that is for all j, $a_j < K\rho^j$, for some $0 < \rho < 1$). It can be shown that under these conditions, $\mathbb{E}(X_t^2) < \infty$ and the process is strongly mixing with a geometric rate (c.f. Bousamma (1998) and Basrak, Davis, and Mikosch (2002)). Therefore the GARCH process satisfies Assumption 2.1, with $c_j = j^{4+\delta}\rho^j$, where c_j is defined in Assumption 2.1(iv).

Suppose we observe $\{X_t : t = 1, \ldots, n\}$. Using the ARCH($\infty$) representation of the GARCH process, let $\underline{a}_{p_n}$ denote the first p_n parameters of the ARCH(∞) process. Let $\underline{\hat{a}}_{p_n}$ be the estimator defined as in (4). Then by using Theorem 3.1 almost surely we have

$$\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_1 = O\left(n^{-1}p_n^{6+3\delta/2} + p_n^{3/2}q_{\eta,n} + p_n^{5+\delta}\rho^{p_n}\right).$$

Now suppose $\zeta > 0$ and we let $p_n = (\log n)^{1+\zeta}$, then $\rho^{p_n} = n^{-\gamma(\log n)^{1+\zeta}}$ and $\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_1 = O\left((\log n)^{3(1+\zeta)/2}q_{\eta,n}\right)$. Hence $\|\underline{\hat{a}}_{p_n} - \underline{a}_{p_n}\|_1 \overset{\text{a.s.}}{\to} 0$ as $n \to \infty$

Acknowledgment

The author wishes to thank Jan Johannes for several interesting discussions. Furthermore, the author is extremely grateful to an anonymous referee for making several suggestions, which greatly improved the paper. This research was supported by the DFG (DA 187/12-3). Part of this research was conducted while the author was at the Universität Heidelberg.

A Appendix

In order to prove almost sure convergence of the martingales $\mathcal{S}_{r,n}$ and the sample moments $S_n(i,j)$ considered in (17) and (23) we will make repeated use of the lemma below, which is a straightforward generalisation of Móricz (1976), Theorem 6, to a uniform rate over multiple sums. The lemma below gives an almost sure rate of convergence for general sums of random variables. Let

$$S_n(r_1, \ldots, r_m) = \sum_{k=1}^{n} \xi_k(r_1, \ldots, r_m), \tag{31}$$

where $\{\xi_k(r_1, \ldots, r_m)\}_t$ is an arbitrary random process indexed by $(r_1, \ldots, r_m)$.

Lemma A.1 *Let $S_n(r_1, \ldots, r_m)$ be defined as in (31). Suppose there exists a monotonically increasing sequence $\{G(n)\}_n$, such that*

$$\sup_{1 \le r_1, \ldots, r_m \le p_n} \mathbb{E}\left(\sup_{1 \le t \le n} |S_t(r_1, \ldots, r_m)|^2\right) \le G(n).$$

Assume further that there exists a $K < \infty$ such that for all sequences $\{n_j\}$ which satisfy $2^j \le n_j < 2^{j+1}$, we have $\sup_j |G(n_j)/G(n_{j-1})| < K$. Then we have almost surely

$$\sup_{1 \le r_1, \ldots, r_m \le p_n} |S_n(r_1, \ldots, r_m)| = O\left(p_n^{m/2}(G(n)\log n(\log\log n)^{1+\eta})^{1/2}\right) \tag{32}$$

for any $\eta > 0$.

PROOF. Let $\{\phi(n)\}$ be an arbitrary positive sequence. Before proving the result, we note that $P(\sup_{1\leq r_1,\ldots,r_m\leq p_n}\sup_{1\leq t\leq n}|S_t(r_1,\ldots,r_m)| > \phi(n)) \leq \sum_{1\leq r_1,\ldots,r_m\leq p_n} P(\sup_{1\leq t\leq n}|S_t(r_1,\ldots,r_m)| > \phi(n))$, therefore by using the Chebyshev inequality we have

$$P\left(\sup_{1\leq r_1,\ldots,r_m\leq p_n}\sup_{1\leq t\leq n}|S_t(r_1,\ldots,r_m)| > \phi(n)\right) \leq \frac{p_n^m G(n)}{\phi(n)^2}. \tag{33}$$

We now specify the sequence $\{\phi(n)\}$ such that we can obtain a tight bound for the rate of convergence. Let

$$\tilde{\phi}(n) \quad = \quad \left(p_n^m G(n)\log n(\log\log n)^{1+\eta}\right)^{1/2}$$

and $M_n = \sup_{1\leq r_1,\ldots,r_m\leq p_n}\sup_{1\leq t\leq n}|S_t(r_1,\ldots,r_m)|$. By using (33) we have

$$P\left(M_n > \tilde{\phi}(n)\right) \quad \leq \quad \frac{1}{\log n(\log\log n)^{1+\eta}}.$$

Let $\{n_j\}$ be an arbitrary subsequence such that $2^j \leq n_j < 2^{j+1}$. By appealing to the Borel-Cantelli Lemma we can show for a large enough n_j we almost surely have

$$|M_{n_j}| \leq \tilde{\phi}(n_j).$$

Suppose $n_{j-1} < n \leq n_j$, then by the definition $\{M_{n_j}\}$, $|\sup_{1\leq r_1,\ldots,r_m\leq p_n} S_n(r_1,\ldots,r_m)| \leq M_{n_j}$ and $\tilde{\phi}(n_{j-1}) \leq \tilde{\phi}(n)$. Then we have almost surely

$$\frac{\sup_{1\leq r_1,\ldots,r_m\leq p_n}|S_n(r_1,\ldots,r_m)|}{\tilde{\phi}(n)} \leq \frac{M_{n_j}}{\tilde{\phi}(n_{j-1})} \leq \frac{\tilde{\phi}(n_j)}{\tilde{\phi}(n_{j-1})} \leq K,$$

which proves (32). □

We use the lemma above and the following maximal inequality to obtain a rate of convergence for $\mathcal{S}_{r,n}$.

Lemma A.2 *Suppose Assumption 2.1(i,iii) holds. Let $\mathcal{S}_{r,n}$ be defined as in (17). Then we have*

$$\left[\mathbb{E}(\sup_{r\leq t\leq n}|\mathcal{S}_{r,t}|)^2\right]^{1/2} \quad \leq \quad \frac{n^{1/2}}{\ell(p_n)}(\mathbb{E}|Z_0^2-1|^2)^{1/2}\left\{\mathbb{E}|\frac{\sigma_0^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{-k}^2)}|^2\right\}^{1/2}. \tag{34}$$

PROOF. Recalling that $S_{r,t}$ is the sum of martingale differences, $X_{t-r}^2/(g+\sum_{k=1}^{\infty}\ell(k)X_{-k}^2) \leq \ell(p_n)^{-1}$ and by using Doob's and Burkhölder's inequalities, we have

$$\mathbb{E}\left(\max_{r\leq t\leq n}\mathcal{S}_{r,t}^2\right) \quad \leq \quad 4\mathbb{E}\left(|\mathcal{S}_{n,t}|^2\right) \leq 8(\mathbb{E}|Z_t^2-1|^2)^{1/2}(\mathbb{E}|\frac{\sigma_0^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{-k}^2)^2}X_{t-r}^2|^2)^{1/2}n^{1/2}$$

thus giving us the required result. □

Unfortunately, since $\mathcal{S}_n(i,j)$ (defined in (23)) is not the sum of martingale differences we cannot directly apply the maximal inequalities (a combination of Doob's and Burkhölder's inequality) to obtain $\mathbb{E}(\max_{1\leq t\leq n}|\mathcal{S}_t(i,j)|^2)$. Instead we show that $\{c_t(i,j)\}_t$ is L_2-NED with respect to the mixing process $\{X_t\}_t$. The advantage of Near Epoch Dependence is that a process which is NED can be expressed as an infinite sum of martingale differences (usually called a mixingale, c.f. (Davidson, 1994), Theorem 17.5). This allows us to apply the maximal inequalities to the infinite sum of martingales. To show that $\{c_t(i,j)\}_t$ is L_2-NED we need to obtain a rate of decay for $\mathbb{E}|c_t(i,j) - \mathbb{E}(c_t(i,j)|\mathcal{F}^t_{t-r})|^2$. To do this we first require an appropriate bound for $c_t(i,j)$. Let

$$g_{X,t} = g + \sum_{k=1}^{\infty}\ell(k)X_{t-k}^2 \text{ and } g_{Y,t} = g + \sum_{k=1}^{\infty}\ell(k)Y_{t-k}^2. \tag{35}$$

Lemma A.3 *Let $\{X_t\}$ and $\{Y_t\}$ be sequences, $\{\ell(k)\}_k$ be a decreasing sequence and $Z_t(X,Y) = 2\sum_{k=1}^{\infty}\ell(k)(X_{t-k}^2 - Y_{t-k}^2)$. Then we have*

$$\mathbb{E}\left|\frac{X_{t-i}^2X_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2} - \frac{Y_{t-i}^2Y_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)Y_{t-k}^2)^2}\right| \leq \frac{(1+g^{-1})}{\ell(\max(i,j))^2}\min\{1, Z_t(X,Y)\}. \tag{36}$$

PROOF. Recalling (35) we have

$$\begin{aligned}\left|\frac{X_{t-i}^2X_{t-j}^2}{g_{X,t}^2} - \frac{Y_{t-i}^2Y_{t-j}^2}{g_{Y,t}^2}\right| &= \left|\frac{X_{t-i}^2}{g_{X,t}}\left(\frac{X_{t-j}^2}{g_{X,t}} - \frac{Y_{t-j}^2}{g_{Y,t}}\right) + \frac{Y_{t-j}^2}{g_{Y,t}}\left(\frac{X_{t-i}^2}{g_{X,t}} - \frac{Y_{t-i}^2}{g_{Y,t}}\right)\right| \\ &\leq \frac{(1+g^{-1})}{\ell(\max(i,j))^2}|g_{X,t} - g_{Y,t}|. \end{aligned}\tag{37}$$

We also observe that

$$\left|\frac{X_{t-i}^2X_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2} - \frac{Y_{t-i}^2Y_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)Y_{t-k}^2)^2}\right| \leq \frac{2}{\ell(\max(i,j))^2}. \tag{38}$$

Combining (37) and (38) we obtain (36). □

It can easily be shown that $\mathbb{E}\left(\frac{X_{t-i}^2X_{t-j}^2}{g_{X,t}^2} - \frac{Y_{t-i}^2Y_{t-j}^2}{g_{Y,t}^2}\right)^2$, can be bounded by a random variable which involves the fourth moment of X_t. However, by using the lemma above, which shows that the difference $\frac{X_{t-i}^2X_{t-j}^2}{g_{X,t}^2} - \frac{Y_{t-i}^2Y_{t-j}^2}{g_{Y,t}^2}$ will always be less than one, we are able to reduce the number of moments required.

Lemma A.4 *Suppose $\{X_t\}$ and $\{Y_t\}$ are stationary random sequences with $\mathbb{E}(X_t^2) < \infty$ and $\mathbb{E}(Y_t^2) < \infty$, and $\{\ell(j)\}$ a decreasing sequence then we have*

$$\mathbb{E}\left(\frac{X_{t-i}^2X_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2} - \frac{Y_{t-i}^2Y_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)Y_{t-k}^2)^2}\right)^2 \leq \left(\frac{2}{\ell(\max(i,j))^2}\right)^2\mathbb{E}\left(\sum_{k=1}^{\infty}\ell(k)|X_{t-k}^2 - Y_{t-k}^2|\right).$$

PROOF. We use the definition of $Z_t(X_t, Y_t)$ in Lemma A.3 and recall the definition of $g_{X,t}$ given in (35). Let P denote the probability measure on $\{X_t\}$ and $\{Y_t\}$. Partitioning the integral, using Lemma A.3 and noting $Z_t(X, Y) = 2\sum_{k=1}^{\infty} \ell(k)(X_{t-k}^2 - Y_{t-k}^2)$ we have

$$\begin{aligned}
&\mathbb{E}\left(\frac{X_{t-i}^2 X_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2} - \frac{Y_{t-i}^2 Y_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)Y_{t-k}^2)^2}\right)^2 \\
=& \int_{|Z_t(X_t,Y_t)|\leq 1}\left(\frac{X_{t-i}^2 X_{t-j}^2}{g_{X,t}^2} - \frac{Y_{t-i}^2 Y_{t-j}^2}{g_{Y,t}^2}\right)^2 dP + \int_{|Z_t(X_t,Y_t)|>1}\left(\frac{X_{t-i}^2 X_{t-j}^2}{g_{X,t}^2} - \frac{Y_{t-i}^2 Y_{t-j}^2}{g_{Y,t}^2}\right)^2 dP \\
\leq& \left(\frac{2}{\ell(\max(i,j))^2}\right)^2 \left\{\int_{|Z_t(X,Y)|\leq 1} Z_t(X,Y)^2 dP + \int_{|Z_t(X,Y)|>1} 1^2 dP\right\}.
\end{aligned}$$

Since $Z_t(X, Y)^2 \leq Z_t(X, Y)$ when $Z_t(X, Y)^2 < 1$ and by definition the second integral on the right hand side of the above inequality is bounded above by $1 \leq Z_t(X, Y)$, then we have

$$\mathbb{E}\left(\frac{X_{t-i}^2 X_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2} - \frac{Y_{t-i}^2 Y_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)Y_{t-k}^2)^2}\right)^2 \leq \left(\frac{2}{\ell(\max(i,j))^2}\right)^2 \mathbb{E}\left(\sum_{k=1}^{\infty}\ell(k)|X_{t-k}^2 - Y_{t-k}^2|\right).$$

Thus giving the required result. □

Below we use Lemmas A.3 and A.4 to show that $\{c_t(i, j)\}_t$ is a L_2-NED process with respect to the mixing process $\{X_t\}$. To show L_2-NED of $\{c_t(i, j)\}_t$ with respect to $\{X_t\}_t$, we need to obtain a rate of decay for $c_t(i, j)$ when conditioned on the mixing process. More precisely we require a rate for $\mathbb{E}[c_t(i, j) - \mathbb{E}(c_t(i, j)|X_t, \ldots, X_{t-r})]^2$, which we derive in the following lemma.

Theorem A.1 *Suppose Assumption 2.1(i,iv) holds. Let* $\mathcal{F}_{t-r}^t = \sigma(X_t, \ldots, X_{t-r})$. *and* $\{c_t(i, j)\}_t$ *be defined as in (28). Then we have*

$$\left(\mathbb{E}|c_t(i,j) - \mathbb{E}(c_t(i,j)|\mathcal{F}_{t-r}^t)|^2\right)^{1/2} \leq \frac{2\sqrt{2}}{\ell(\max(i,j))^2}(\mathbb{E}|X_0^2|)^{1/2}\left(\sum_{k=r}^{\infty}\ell(k)\right)^{1/2}. \tag{39}$$

Additionally if Assumption 2.1(ii,v) holds, then $\{c_t(i, j)\}$ *is* L_2*-NED with respect to the mixing process* $\{X_t\}$ *of rate* $-(1 + \delta/2)$, *with constant* $2\sqrt{2}\|X_0^2\|_1^{1/2}/\ell(\max(i, j))^2$.

PROOF. Define

$$H_{t-r}^t = \frac{\mathbb{E}(X_{t-i}^2|\mathcal{F}_{t-r}^t)\mathbb{E}(X_{t-j}^2|\mathcal{F}_{t-r}^t))}{(g+\sum_{k=r}^{\infty}\ell(k)\mathbb{E}(X_{t-k}^2|\mathcal{F}_{t-r}^t))^2}, \tag{40}$$

then it is clear that $H_{t-r}^t \in \mathcal{F}_{t-r}^t$. We recall that $\mathbb{E}(c_t(i, j)|\mathcal{F}_{t-r}^t)$ is the best estimator of $c_t(i, j)$ in $\mathcal{F}_{t-r}^t$ under the ℓ_2-norm, hence any g which is $\mathcal{F}_{t-r}^t$-measurable $\mathbb{E}\{c_t(i, j) - \mathbb{E}(c_t(i, j)|\mathcal{F}_{t-r}^t)\}^2 \leq \mathbb{E}\{c_t(i, j) - g\}^2$. We now exploit this idea, where we use the inequality in Lemma A.4, setting

$X_t \equiv X_t$ and and $Y_t \equiv \mathbb{E}(X_{t-i}^2|\mathcal{F}_{t-r}^t)$. Noting that for $k \leq r$, $Y_{t-k} = \mathbb{E}(X_{t-k}^2|\mathcal{F}_{t-r}^t) = X_{t-k}^2$, altogether this gives

$$\begin{aligned}
(\mathbb{E}|c_t(i,j) - \mathbb{E}(c_t(i,j)|\mathcal{F}_{t-r}^t)|^2)^{1/2} &\leq \Big(\mathbb{E}\Big|\frac{X_{t-i}^2 X_{t-j}^2}{(g+\sum_{k=1}^{\infty}\ell(k)X_{t-k}^2)^2} - H_{t-r}^t\Big|^2\Big)^{1/2} \\
&\leq \frac{2}{\ell(\max(i,j))^2}\left\{\mathbb{E}\left[\sum_{k=r}^{\infty}\ell(k)\left\{X_{t-k}^2 + \mathbb{E}(X_{t-k}^2|\mathcal{F}_{t-r}^t)\right\}\right]\right\}^{1/2} \\
&\leq \frac{2\sqrt{2}}{\ell(\max(i,j))^2}(\mathbb{E}|X_0^2|)^{1/2}\left(\sum_{k=r}^{\infty}\ell(k)\right)^{1/2}
\end{aligned}$$

thus giving us (39). Finally, using the above and $\ell(k) = K\,k^{-(3+\delta)}$ we observe that $(\sum_{k=r}^{\infty}\ell(k))^{1/2} \leq Kr^{-(1+\delta/2)}$, thus $c_t(i,j)$ is L_2-NED of rate $-(1+\delta/2)$ with respect to the process $\{X_t\}$. □

We will use the result above to show that $\{c_t(i,j)\}$ can be written as the infinite sum of martingale difference, which requires the following corollary.

Corollary A.1 *Suppose Assumption 2.1(i,iv) holds. Let $c_t(i,j)$ be defined as in (28) and $\mathcal{F}_t = \sigma(X_t, X_{t-1},\ldots)$. Then we have*

$$\left(\mathbb{E}\Big|\mathbb{E}(c_t(i,j) - \mathbb{E}\{c_t(i,j)\}|\mathcal{F}_{t-m})\Big|^2\right)^{1/2} \leq \frac{K}{\ell(\min(i,j))^2}\rho(m) \tag{41}$$

where $\rho(m) = \left(\frac{m}{2}\right)^{-\min\{(2+\delta)/2,(1+a)/2\}}$.

PROOF. To prove this result we use the same arguments given in Gallant (1987) or Davidson (1994), Chapter 17, hence we omit the details. □

A useful implication of Corollary A.1 is that $\{c_t(i,j)\}$ can be written as the sum of martingale differences and satisfies the representation

$$c_t(i,j) - \mathbb{E}(c_t(i,j)) = \sum_{m=0}^{\infty} V_{t,m}(i,j) \tag{42}$$

where $V_{t,m}(i,j) = \mathbb{E}(c_t(i,j)|\mathcal{F}_{t-m}) - \mathbb{E}(c_t(i,j)|\mathcal{F}_{t-m-1})$ with $\mathcal{F}_t = \sigma(X_t, X_{t-1},\ldots)$. It is clear that $\{V_{t,m}(i,j)\}_m$ is a sequence of martingale differences. By using this representation we can rewrite $S_t(i,j)$ as

$$S_t(i,j) - \mathbb{E}(S_t(i,j)) = \tilde{S}_t(i,j) = \sum_{k=1}^{t}\sum_{m=0}^{\infty} V_{k,m}(i,j). \tag{43}$$

Lemma A.5 *Suppose Assumption 2.1(i,iv) holds. Let $\tilde{S}_t(i,j)$ be defined as in (43). Then we have*

$$(\mathbb{E}|\max_{1\leq t\leq n}\tilde{S}_t(i,j)|^2)^{1/2} \leq \frac{\tilde{K}}{\ell(\max(i,j))^2}n^{1/2} \tag{44}$$

for some $\tilde{K} > 0$.

PROOF. By using (43) and interchanging the summands we have

$$(\mathbb{E}|\sup_{1\leq t\leq n}\tilde{S}_t(i,j)|^2)^{1/2} \;\leq\; \sum_{m=0}^{\infty}(\mathbb{E}|\sup_{1\leq t\leq n}\sum_{k=\max(i,j)}^{t}V_{k,m}(i,j)|^2)^{1/2}.$$

We note for m fixed, $\{V_{k,m}(i,j)\}_k$ is a sequence of martingale differences. Therefore by using Doob's inequality we have

$$(\mathbb{E}|\sup_{1\leq t\leq n}\tilde{S}_t(i,j)|^2)^{1/2} \;\leq\; \sum_{m=0}^{\infty}\left(\sum_{k=\max(i,j)}^{n}\mathrm{var}\{V_{k,m}(i,j)\}\right)^{1/2}. \quad (45)$$

We now bound $\mathrm{var}(V_{k,m}(i,j))$. By using (41) we have

$$\begin{aligned}\mathrm{var}\{V_{k,m}(i,j)\} &\leq \Big(\big(\mathbb{E}\big|\mathbb{E}(c_k(i,j)\big|\mathcal{F}_{k-m}) - \mathbb{E}(c_k(i,j))|^2\big)^{1/2} + (\mathbb{E}\big|\mathbb{E}(c_k(i,j)|\mathcal{F}_{k-m-1}) - \mathbb{E}(c_k(i,j))\big|^2)^{1/2}\Big)^2 \\ &\leq \frac{4K}{\ell(\max(i,j))^4}\left(\frac{m}{2}\right)^{-\min\{(2+\delta),(1+a)\}}. \quad (46)\end{aligned}$$

Substituting (46) into (45) gives us the required result. □

References

An, H.-Z., Chen, Z.-G., & Hannan, E. (1982). Autocorrelation, autoregression and autoregressive approximation. *Ann. Statist.*, *10*, 926-936.

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

Bose, A., & Mukherjee, K. (2003). Estimating the ARCH parameters by solving linear equations. *Journal of Time Series Analysis*, *24*, 127-136.

Bousamma, F. (1998). *Ergodicité, mélange et estimation dans les modèles GARCH.* Unpublished doctoral dissertation, Paris 7.

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

Fan, J., & Yao, Q. (2003). *Nonlinear time series.* New York: Springer-Verlag.

Fryzlewicz, P., Sapatinas, T., & Subba Rao, S. (2006). Normalised least-squares estimation in time-varying ARCH models. *Under revision.*

Gallant, R. (1987). *Nonlinear statistical models.* New York: Wiley.

Giraitis, L., Kokoskza, P., & Leipus, R. (2000). Stationary ARCH models: Dependence structure and central limit theorem. *Econometric Theory*, *16*, 3-22.

Giraitis, L., Leipus, R., & Surgailis, D. (2005). Recent advances in ARCH modelling. In A. Kirman & G. Teyssiere (Eds.), *Long Memory in Economics* (p. 3-39). Berlin.

Giraitis, L., & Robinson, P. (2001). Whittle estimation of ARCH models. *Econometric Theory*, *17*, 608-631.

Koulikov, D. (2003). Long memory ARCH(∞) models: specification and quasi-maximum likelihood estimation. *Preprint (see* `http://www.cls.dk/caf/wp/wp-163.pdf`*)*.

Ling, S. (2006). Self-weighted and local quasi-maximum likelihood estimation for ARMA-GARCH/IGARCH models. *To appear in Journal of Econometrics.*

Móricz, F. (1976). Moment inequalities and the strong law of large numbers. *Z. Wahrsch. verw. Gebiete*, *35*, 298-314.

Robinson, P. (1991). Testing for strong serial correlation and dynamic conditional heteroskedasity in multiple regression. *Journal of Econometrics*, *47*, 67-78.

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

Stout, W. (1974). *Almost Sure Convergence.* New York: Academic Press.